

**TOWARDS A THEORETICALLY MOTIVATED MODEL FOR THE  
LEXICOGRAPHIC REPRESENTATION OF KINYARWANDA PHRASEMES IN  
GENERAL KINYARWANDA DICTIONARIES**

**Emmanuel Nikuze**



**Thesis presented in fulfilment of the requirements for the degree of Master of  
Philosophy (Lexicography) at the Stellenbosch University**

**Promoter: Prof. Rufus H. Gouws**

**April 2014**

## **DECLARATION**

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

April 2014

Copyright © 2014 Stellenbosch University

All rights reserved

## ABSTRACT

The decision to investigate the lexicographic representation of Kinyarwanda phrasemes has been motivated by two problems. Firstly, the compilers of Kinyarwanda general dictionaries do not take cognisance of the wide-ranging nature of Kinyarwanda phrasemes. As a result, the confusion of different types of phrasemes entails inadequate treatment of phraseology in Kinyarwanda dictionaries. Furthermore, some types of phrasemes are mistakenly left out or under-represented in Kinyarwanda dictionaries due to the fact that the multifaceted nature of Kinyarwanda phrasemes is not recognised. Secondly, the representation of phrasemes in current Kinyarwanda dictionaries is unsatisfactory as far as data presentation, distribution and accessibility is concerned. Therefore, not only is the multifaceted nature of phrasemes not reflected in the representation of phrasemes in Kinyarwanda dictionaries but Kinyarwanda dictionaries also fall short as far as the application of adequate metalexicographic principles and tools is concerned. It is anticipatively argued that the use of innovative tools and suggestions offered by the contemporary theoretical lexicography coupled with insights from the linguistic classification of phrasemes will enable a user-friendly presentation of phrasemes in general dictionaries of Kinyarwanda. This thesis attempts to bring solutions to these shortcomings by formulating theoretically motivated proposals for lexicographic representation of phraseological items in Kinyarwanda general dictionaries.

To achieve this goal and in order to formulate a model for lexicographic representation of phrasemes this study uses a positivist approach and utilises two theories of lexicography, namely the functional theory of lexicography mainly developed by Bergenholz and Tarp and the general theory of lexicography outlined by Wiegand. The theory of lexicographic functions provided this study with decisive heuristics such as the concept of user situation, lexicographic function and user profile that have been employed to devise a function-motivated model of lexicographic representation of different types of phrasemes. Insights from the theory of dictionary structures were used to determine how various dictionary structures could be used to position phrasemes in the dictionary so that easy access is secured. These theories helped formulate policies for inclusion and treatment of different classes of phrasemes taking into account the function that the dictionary is purported to fulfil and the role different structures of the dictionary play in user-friendly distribution of lexicographic data. In addition, this research relied on various sources of data and used a qualitative approach to data analysis and argument construction.

In Chapter 3 the concept of phraseme in the framework of Meaning-Text linguistics is defined and various pitfalls in lexicographic representation of phrasemes in existing dictionaries of Kinyarwanda are addressed. In Chapter 4, proposals have been made regarding both straight lemmatisation and sublemmatisation of Kinyarwanda idioms in the central list whereby the proposed models have been evaluated taking into account different lexicographic functions. In addition, suggestions have been made as to how to use both the central list section and the outer texts, especially the back matter and the middle matter to present proverbs in general language dictionaries of Kinyarwanda. Further, a model for lexicographic presentation of collocations has been formulated in Chapter 5, taking into account four lexicographic functions: text production, text translation, text reception and cognitive-oriented dictionary. In this chapter it is shown what the representation of collocations could look like when satisfying each of the above-mentioned lexicographic functions. Finally two models for the lexicographic representation of pragmatemes or pragmatic phrasemes are proposed in Chapter 6. In this chapter, both the internal-article treatment of pragmatemes in the central list and the presentation of

pragmatemes in the outer texts are discussed and evaluated. Chapter 7 outlines the summary of major research findings along with related recommendations on the lexicographic treatment of different types of phrasemes and assesses whether the initial research questions have been properly answered, while at the same time paving the way for further research and more comprehensive discussions.

## OPSOMMING

Die besluit om die leksikografiese verteenwoordiging van Kinyarwanda fraseme te ondersoek, word deur twee probleme gemotiveer. Eerstens word die uiteenlopende aard van Kinyarwanda fraseme nie deur die samestellers van algemene woordeboeke vir Kinyarwanda in ag geneem nie. Die verwarring van die verskillende tipes fraseme lei tot onvoldoende hantering van fraseologie in Kinyarwanda woordeboeke. Voorts word sekere tipes fraseme eenvoudig uit Kinyarwanda woordeboeke weggelaat of swak verteenwoordig, huis omdat die veelsydige aard van Kinyarwanda fraseme misgekyk word. Tweedens is die verteenwoordiging van fraseme in die bestaande Kinyarwanda woordeboeke onbevredigend wat betref die aanbod, verspreiding en toeganklikheid van data. Daarom word die veelsydige aard van fraseme nie net swak verteenwoordig in Kinyarwanda woordeboeke nie, maar sodanige woordeboeke skiet ook ver te kort wat die toepassing van voldoende metaleksikografiese beginsels en toepassings betref. Daar word vooruitskouend 'n saak voor uitgemaak dat die gebruik van innoverende instrumente en voorstelle vanuit die kontemporêre teoretiese leksikografie tesame met insigte oor die linguistiese klassifikasie van fraseme, 'n gebruikersvriendelike fraseemaanbod in Kinyarwanda algemene woordeboeke sal bewerkstellig. Hierdie tesis probeer om oplossings vir die tekortkominge te bied deur middel van teoreties gemotiveerde voorstelle vir die leksikografiese verteenwoordiging van fraseologiese items in Kinyarwanda algemene woordeboeke.

Om hierdie doel te verwesenlik en om 'n model vir die leksikografiese verteenwoordiging van fraseme te ontwikkel, word 'n positivistiese benadering gevolg en word twee leskikografieteorieë, naamlik die funksionele teorie, hoofsaaklik ontwikkel deur Bergenholz en Tarp, en die algemene leksikografieteorie soos uiteengesit deur Wiegand, toegepas. Die teorie van leksikografiese funksies het hierdie studie van bepalende, deurslaggewende riglyne voorsien, soos die konsep van die gebruiker se situasie, die leksikografiese funksie, en die gebruikersprofiel wat gebruik is om 'n funksie-gemotiveerde model van leksikografiese verteenwoordiging vir die verskillende tipes fraseme op te stel. Daarbenewens is insigte uit die woordeboekstrukturteorie aangewend om te bepaal hoe verskillende woordeboekstrukture gebruik kan word om fraseme sodanig in woordeboeke te posisioneer dat vinnige en maklike toegang verseker word.

Hierdie teorieë het bygedra tot beleidsformulering vir die insluiting en hantering van verskillende klasse fraseme met inagnome van die veronderstelde woordeboekfunksie en die rol wat verskillende woordeboekstrukture in die gebruikersvriendelike verspreiding van leksikografiese data speel. Die navorsing het ook op verskeie databronne staatgemaak en 'n kwalitatiewe benadering tot data-analise en argumentkonstruksie is gebruik.

In hoofstuk 3 is die konsep *fraseem* binne die raamwerk van die linguistiese teorie van *Meaning-Text* (Betekenis-Teks) gedefinieer en verskeie slaggate in die leksikografiese verteenwoordiging van fraseme in bestaande Kinyarwanda woordeboeke word ondersoek. In hoofstuk 4 word voorstelle ten opsigte van die gewone lemmatisering en sublemmatisering van Kinyarwanda idiome in die sentrale lys bespreek en die voorgestelde modelle word geëvalueer met inagnome van die verskillende leksikografiese funksies. Daarbenewens word voorstelle gemaak oor hoe om sowel die sentrale lys as die buitetekste, veral die agter- en middelteks, te gebruik om spreekwoorde in Kinyarwanda woordeboeke vir algemene taal aan te bied. 'n Model vir die leksikografiese voorstelling van kollokasies word in hoofstuk 5 geformuleer. Vier leksikografiese funksies word in ag geneem, naamlik teksproduksie, teksvertaling, teksresepsie en kognitief-georiënteerde woordeboeke. Daar word aangetoon hoe kollokasies verteenwoordig kan word sodat elk van die bogenoemde leksikografiese funksies in ag geneem word.

Twee modelle vir die leksikografiese verteenwoordiging van pragmateme of pragmatiese fraseme word in hoofstuk 6 voorgestel. Sowel die interne-artikelhantering van pragmateme as die hantering van pragmateme in die buitetekste word bespreek en evaluateer. Ten slotte word die belangrikste navorsingsresultate saam met ander voorstelle ten opsigte van die leksikografiese hantering van verskillende tipes fraseme in hoofstuk 7 uitgelig. Daar word ook gekyk of die navorsingsvrae behoorlik beantwoord is, terwyl die weg gebaan word vir verdere navorsing en meer omvattende bespreking.

## ACKNOWLEDGEMENTS

This thesis would not have been possible without the support, advice and encouragement from various people and institutions. I would like to express my sincere gratitude and appreciation for the assistance that was rendered to me throughout my research.

First and foremost, my sincere thanks to Prof Rufus H. Gouws for his diligent guidance and patient supervision throughout the course of this research. His suggestions, encouragement and numerous suggestions enabled me to complete this work. He aroused a genuine interest in theoretical lexicography in me. I am deeply indebted to him. In addition to research supervision, Prof Gouws generously provided funds in order to remunerate one of the two editors who proofread this work. I appreciate his academic guidance and financial assistance from the bottom of my heart.

I also want to express my gratitude for the financial support I received from the Institute of Scientific and Technological Research (IRST), Rwanda that sponsored a portion of the tuition fees. I also extend my appreciation to my former and current superiors, Dr Jean Baptiste Nduwayezu, Director General of the IRST, Academician Dr Cyprien Niyomugabo, Chairman of Rwanda Academy of Language and Culture (RALC) and Dr James Vuningoma, Executive Secretary of RALC for their support, understanding and encouragement.

To my former colleague, Mr Simon Bizimana, (now retired), my gratitude for kindly sharing with me during five years we worked together at the IRST, his wealth of experience not only in linguistics but also in lexicography of Kinyarwanda. He was always ready to respond to my questions. Furthermore I am greatly indebted to Honorable Prof. Laurent Nkusi (Senate of the Republic of Rwanda) and Prof. Evariste Ntakirutimana (University of Rwanda), my undergraduate professors for having nurtured my research interest in lexicology and phraseology and sharpened my research thinking.

The final manuscript of this thesis was proofread by Mrs Jana Walters, a professional translator-editor (Republic of South Africa) and Dr Ruth Brink (United States of American). More particularly, Dr Brink kindly edited this thesis without any remuneration. I would like to express my sincere thanks, gratitude and appreciation to them for their high-quality editing work and their valuable comments and suggestions on the language, structure and content of this thesis. I am also grateful to my colleagues, more particularly Mr Clement Ndibwami, Mr Claude Ngirabanzi, Mr Modeste Nsanzabaganwa and Dr Jacques Nzabonimpa for advice and encouragement I received from them throughout the way.

To the administrative staff at Stellenbosch University, my sincere gratitude for their cooperation in particular, Ms Winifred Fourie who has always kindly responded to my inquiries related to my status as distance student.

Finally, from the bottom of my heart, I want to convey gratitude to my beloved wife Grace and our son Davies, for their support, love and encouragement and for the sacrifices made, both morally and financially in order for me to complete this project. Their readiness and availability in times of need will forever be acknowledged.

**DEDICATION**

To

My beloved wife Grace  
and  
our son Davies

I dedicate this thesis.

## **ABBREVIATIONS**

ConceptR: Conceptual Representation

DSyntR: Deep Syntactic Representation

DSemR: Deep Semantic Representation

ECL: Explanatory and Combinatorial Lexicology

INRS: Institut National de Recherche Scientifique

IRST: Institut de Recherche Scientifique et Technologique (Institute for Scientific and Technological Research)

LGP: Language for General Purposes

LSP: Language for Specific Purposes

MTM: Meaning-Text Model

MTT: Meaning-Text Theory

PartSp: Part-of-Speech

PhonR: Phonetic Representation

SemR: Semantic Representation

SIT: Situation (of communication)

SyntR: Syntactic Representation

## TABLE OF CONTENTS

<b>DECLARATION.....</b>	<b>II</b>
<b>ABSTRACT .....</b>	<b>III</b>
<b>OPSOMMING .....</b>	<b>V</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>VII</b>
<b>DEDICATION .....</b>	<b>IX</b>
<b>ABBREVIATIONS .....</b>	<b>X</b>
<b>TABLE OF CONTENTS .....</b>	<b>XI</b>
<b>CHAPTER 1: GENERAL INTRODUCTION.....</b>	<b>1</b>
1.1 INTRODUCTION .....	1
1.2 STATEMENT OF THE RESEARCH PROBLEM .....	1
1.3 PRELIMINARY STUDY .....	3
1.4 OBJECTIVE AND SCOPE OF THE STUDY .....	5
1.5 HYPOTHESES.....	7
1.6 THEORETICAL APPROACH .....	7
1.7 METHODOLOGICAL APPROACH .....	9
1.8 BACKGROUND TO KINYARWANDA.....	12
1.8.1 <i>Sociolinguistic profile of Kinyarwanda</i> .....	12
1.8.2 <i>Classification and designation of the language</i> .....	15
1.8.3 <i>Overview of linguistic research on Kinyarwanda</i> .....	16
1.9 OVERVIEW OF THE THESIS.....	17
1.10 CONCLUDING REMARKS .....	18
<b>CHAPTER 2: THEORETICAL FRAMEWORK.....</b>	<b>19</b>
2.1 INTRODUCTION .....	19
2.2 THE FUNCTIONAL THEORY OF LEXICOGRAPHY .....	19
2.2.1 <i>Background and postulates of the theory of lexicographic functions</i> .....	19
2.2.2 <i>Key concepts of the theory of lexicographic functions</i> .....	23
2.2.2.1 Types of users.....	23
2.2.2.2 User situations.....	25
2.2.2.3 User needs.....	26
2.2.2.4 Genuine purpose of a dictionary.....	27
2.2.2.5 Lexicographic functions .....	28
2.3 THE STRUCTURAL THEORY OF LEXICOGRAPHY .....	29
2.3.1 <i>Theoretical foundations</i> .....	29
2.3.2 <i>Access structure and data presentation in a dictionary</i> .....	30
2.3.3 <i>Frame structure</i> .....	32
2.3.3.1 Central list and access structure.....	32
2.3.3.2 Outer texts and the access structure .....	33
2.3.4 <i>The Macrostructure</i> .....	36
2.3.4.1 Alphabetical macrostructure .....	36
2.3.4.2 Thematic macrostructure .....	38
2.3.5 <i>The microstructure</i> .....	40
2.3.5.1 Microstructure and data distribution structure.....	40
2.3.5.2 Microstructure and data architecture.....	41
2.3.5.3 Types of microstructures and addressing structures .....	43
2.3.4 <i>The mediostructure</i> .....	44
2.4 CONCLUSION.....	46

<b>CHAPTER 3: PHASEMES IN KINYARWANDA LINGUISTICS AND LEXICOGRAPHY: THE EXTENT OF THE PROBLEM .....</b>	<b>47</b>
3.1 INTRODUCTION .....	47
3.2 THE CONCEPT OF PHASEME IN THE MEANING-TEXT THEORY .....	47
3.2.1 <i>The Meaning-Text Theory</i> .....	47
3.2.2 <i>Definition of the concept of phaseme</i> .....	48
3.3 PHASEMES AND KINYARWANDA STUDIES: THE EXTENT OF THE PROBLEM .....	52
3.3.1 <i>Kinyarwanda phrasemes in linguistic studies</i> .....	52
3.3.2 <i>Phrasemes in the metalexicography and lexicography of Kinyarwanda</i> .....	56
3.3.2.1 Phrasemes in the metalexicography of Kinyarwanda.....	56
3.3.2.2 The lexicographic representation of phrasemes in current Kinyarwanda dictionaries .....	60
3.3.2.3 Semantic description of phrasemes in Kinyarwanda dictionaries .....	74
3.4. CONCLUSION.....	76
<b>CHAPTER 4: A MODEL FOR LEXICOGRAPHIC REPRESENTATION OF KINYARWANDA IDIOMS AND PROVERBS .....</b>	<b>78</b>
4.1 INTRODUCTION .....	78
4.2 THE CONCEPTS OF IDIOM AND PROVERB.....	78
4.2.1 <i>Definition of idiom</i> .....	78
4.2.2 <i>Difference between idioms and proverbs</i> .....	80
4.3 LEXICOGRAPHIC REPRESENTATION OF IDIOMS IN KINYARWANDA DICTIONARIES: A TRANSFORMATIVE ACCOUNT .....	82
4.3.1 <i>Working assumptions</i> .....	82
4.3.2 <i>Idioms and lexicographic functions</i> .....	83
4.3.2.1 Idioms in a dictionary with a text reception function .....	83
4.3.2.2 Idioms in a dictionary with a text production function .....	85
4.3.2.3 Idioms in a dictionary with a cognitive function .....	87
4.3.3 <i>Towards the lemmatisation of idioms in Kinyarwanda dictionaries</i> .....	87
4.3.3.1 Sublemmatisation of idioms .....	88
4.3.3.2 Straight lemmatisation of idioms in a text production dictionary .....	98
4.3.3.3 Challenges and special issues in the straight lemmatisation of idioms .....	103
4.4 LEXICOGRAPHIC TREATMENT OF KINYARWANDA PROVERBS .....	110
4.4.1 <i>Proverbs as lexicographic data category</i> .....	111
4.4.2 <i>Lexicographic treatment of proverbs in general dictionaries</i> .....	112
4.4.2.1 Lemmatizing proverbs as macrostructural primary level treatment units.....	112
4.4.2.2 Treatment of proverbs within the microstructure .....	113
4.4.2.3 Treatment of proverbs in the outer texts.....	117
4.5 CONCLUSION.....	121
<b>CHAPTER 5: A FUNCTION-DRIVEN MODEL FOR THE TREATMENT OF COLLOCATIONS IN KINYARWANDA LEXICOGRAPHY .....</b>	<b>123</b>
5.1 INTRODUCTION .....	123
5.2 THE CONCEPT OF COLLOCATION .....	123
5.2.1 <i>Two approaches to the concept of collocations</i> .....	123
5.2.2. <i>On the definition of the concept of collocation in the Meaning-Text approach</i> .....	125
5.2.3 <i>Types of collocations in Kinyarwanda</i> .....	127
5.2.3.1 Decodability .....	127
5.2.3.2 Predictability .....	128
5.2.3.3 Binary and dissymmetric aspects .....	129
5.3 COLLOCATIONS IN A TEXT PRODUCTION AND TRANSLATION-ORIENTED DICTIONARY .....	131
5.3.1 <i>The role of collocations in text production and translation</i> .....	131
5.3.2 <i>An onomasiological approach to the lexicographic inclusion of Kinyarwanda collocations</i> .....	132
5.3.2.1 Theoretical foundations for an onomasiological presentation of collocations .....	132
5.3.2.2 A comparison of semasiological and onomasiological presentations of collocations .....	137
5.3.2.3 Implementation of the proposed onomasiological model .....	139
5.4 KINYARWANDA COLLOCATIONS IN A DICTIONARY DESIGNED FOR TEXT RECEPTION.....	146
5.5 KINYARWANDA COLLOCATIONS AND KNOWLEDGE-RELATED NEEDS.....	149

5.6 CONCLUSION.....	151
<b>CHAPTER 6: THE LEXICOGRAPHIC REPRESENTATION OF KINYARWANDA PRAGMATEMES .....</b>	<b>153</b>
6.1 INTRODUCTION .....	153
6.2 CHARACTERISATION OF PRAGMATEMES AND THE IMPORTANCE OF PRAGMATEMES IN COMMUNICATION .....	153
<i>6.2.1 Definition of pragmateme.....</i>	153
<i>6.2.2 Pragmateme and situation of communication (SIT).....</i>	155
<i>6.2.3 Pragmateme and conceptual representation (ConceptR).....</i>	156
<i>6.2.4 Pragmatemes vs idioms and collocations.....</i>	156
<i>6.2.5 Importance of pragmatemes in communication .....</i>	157
6.3 A MODEL FOR THE LEXICOGRAPHIC REPRESENTATION OF PRAGMATEMES IN KINYARWANDA LEXICOGRAPHY .....	159
<i>6.3.1 Preliminary considerations .....</i>	159
<i>6.3.2 Pragmatemes as part of the central list.....</i>	160
<i>6.3.2.1 How should the guiding element under which a pragmateme is listed be selected?.....</i>	160
<i>6.3.3.2 Which type of microstructure is appropriate for article-internal inclusion of pragmatemes?.....</i>	162
<i>6.3.2.3 What should the comment on pragmatics contain? .....</i>	164
<i>6.3.3 A thematic treatment of pragmatemes in outer texts .....</i>	169
<i>6.3.3.1 Rationale for the inclusion of pragmatemes in outer texts .....</i>	169
<i>6.3.3.2 Why the onomasiological presentation? .....</i>	170
<i>6.3.3.3 Trial of thematic arrangement of selected pragmatemes with “warning” as speech act.....</i>	173
<i>6.3.3.4 Observations on the inclusion of pragmatemes in outer texts and access structure related issues.....</i>	182
6.4 CONCLUSION.....	184
<b>CHAPTER 7: GENERAL CONCLUSION .....</b>	<b>186</b>
7.1 INTRODUCTION .....	186
7.2 SUMMARY OF RESEARCH FINDINGS .....	186
7.3 RESEARCH HYPOTHESES, QUESTIONS AND CONTRIBUTION OF THE RESEARCH RESULTS.....	190
7.4 CONCLUDING REMARKS AND WAY FORWARD .....	193
<b>BIBLIOGRAPHY .....</b>	<b>195</b>

## CHAPTER 1: GENERAL INTRODUCTION

### 1.1 Introduction

This thesis deals with the representation of Kinyarwanda phrasemes in the context of general language lexicography of Kinyarwanda. Its ultimate goal is to devise a theoretically motivated model for lexicographic representation and treatment of Kinyarwanda phrasemes with a view to improve their accessibility in both existing and future dictionaries of Kinyarwanda. It is expected that the results of this research will provide Kinyarwanda lexicographers with critical information on which to base their decisions regarding the handling of phraseology in the dictionary. The recommendations drawn from this research will help to improve the treatment of phrasemes in Kinyarwanda dictionaries. The aim of this chapter is to explain the rationale, the purpose and the approaches used in this research and to outline the content thereof.

To this end, the rest of this introductory chapter is organised as follows: In Section 1.2 the research problem is stated. Section 1.3 contains a preliminary study on the area of research and the motivation and feasibility of the study is exposed. In Section 1.4 the objectives and scope of this research is provided and in Section 1.5 the hypotheses of research is outlined. In Section 1.6 the theoretical approaches used as theoretical framework in this study are explained and Section 1.7 highlights the methodological approaches adopted in this thesis. A background to Kinyarwanda is provided in Section 1.8 and the sociolinguistic profile of Kinyarwanda is explained. Furthermore, Section 1.8 contains remarks on the classification and designation of the language and an overview of linguistic research on Kinyarwanda. In Section 1.9 the content of the thesis is outlined chapter by chapter while the last section (1.10) provides concluding remarks.

### 1.2 Statement of the research problem

This study investigates the representation of various types of Kinyarwanda phrasemes in general dictionaries of Kinyarwanda. In order to help the potential users to fully benefit from the use of a dictionary, the presentation and treatment of dictionary data in the dictionary must be based on sound theoretical principles of lexicography. Without the use of an appropriate theoretical model, the long-standing problem of the representation and accessibility of phrasemes cannot be solved. In this regard, Kinyarwanda dictionaries leave something to be desired as far as lexicographic representation and access to phrasemes are concerned. This situation is due to the linguistic perspective of the

lexicography of Kinyarwanda and the non-application of suitable principles and tools emanating from theoretical lexicography in dealing with phrasemes.

It is observed that in some dictionaries of Kinyarwanda, idioms are treated as co-textual entries while they should be lemmatised as primary or secondary level treatment units. In this way they are treated as illustrative materials. Because of this access to idioms is complicated. They do not receive direct treatment as a distinct data category and the same holds for collocations in general. Another delicate issue relates to the selection of the guiding element under which an idiom should be recorded if the lexicographer opts for sublemmatising it as a secondary level treatment unit. In the case of collocations, their lexicographic status rules out a lemmatic treatment in a general language dictionary, but because they belong to phraseological phenomena they should be systematically addressed in a dictionary. However, the problem is to know whether they should be treated equally as free illustrative sentences or not. Furthermore, as far as their lexicographic inclusion is concerned a problem regarding the guiding element also arises. It is important to know whether a collocation should be recorded in the article of the constituent representing either its base or collocate. Thus, proposals need to be made regarding the way in which collocations should be included and treated in general language dictionaries.

It is also important to formulate an explicit inclusion model for pragmatemes and proverbs, taking into account the structure of these phraseology units but also and more importantly the functions of dictionaries, the users' needs and the access structures of the dictionary.

The main issue addressed in this research is how to present various types of phrasemes in a dictionary for the user to access them in an unproblematic way and to successfully retrieve the needed information in accordance with different situations of dictionary consultation. Therefore, the questions this research seeks to address can be stated as follows: *1) What adequate lexicographic principles can lexicographers rely on when formulating informed policies for representation and description of various types of Kinyarwanda phrasemes, in order to ease their accessibility in Kinyarwanda dictionaries and ensure an unproblematic retrieval of information? 2) How could those insights from theoretical lexicography and particularly various structural parts of a dictionary be used in order to accommodate phrasemes in a manner which is compatible with the requirements of each type of phraseme and ensuring easy access to phraseological data in accordance with the specific situation of use or dictionary function? 3) Could insights in linguistic approaches to phraseology*

**contribute to lexicographic representation of Kinyarwanda phrasemes?** These relevant questions are worth investigating.

### 1.3 Preliminary study

It is common knowledge that phraseological phenomena are ubiquitous in most if not all languages. It is also acknowledged that phrasemes constitute a good portion of the lexical stock of any language and that phrasemes enable the expression of many senses and concepts that cannot be expressed in any other way. Hausmann has pertinently said that all is idiomatic in the language (Hausmann 1997). Therefore, the need for recording phrasemes in lexicographic work is understood; what needs emphasis is how to do it in accordance with the genuine purpose of the dictionary. In other words, for a dictionary to meet both communicational and cognitive needs of the targeted users, it must display a sound presentation of phrasemes.

This should also be true in Kinyarwanda where phrasemes occur abundantly, not solely in spoken texts but also in written texts. As per available literature, the importance of phrasemes in Kinyarwanda cannot be doubted. On the one hand, considerable work has been done in the field of ethnographic collection and folk etymology resulting in many compilations of Kinyarwanda phrasemes (Bigirumwami 1967; Kayigana 1971, 2007; Overschelde 1972; Mulihano & Mineduc 1980; Mulihano & Minesupres 1986; MNR 2005). On the other hand, lexicographers have endeavoured to represent phrasemes in the general purpose dictionaries of Kinyarwanda (Schumacher 1956; Jacob 1983, 1985, 1987; IRST 2002; Coupez, Kamanzi, Bizimana, Sematama, Rwabukumba & Ntazinda *et al.* 2005; Bizimana & Kayumba *et al.* 2010, 2011). However, more can still be done to ensure the appropriate presentation and treatment of phrasemes in a user-friendly way.

Although many scholars have recognised the challenge of lexicographic representation of phrasemes, the study of recent developments in metalexicography and the emphasis put on the user perspective prompted the author hereof to question the current lexicographic representation of Kinyarwanda phrasemes and to investigate the application of metalexicographic insights in order to mitigate these issues. For example, a critical review of the representation of phrasemes in the most important Kinyarwanda dictionaries reveals many problems regarding the concept of phrasemes itself, the lexicographic representation, and the accessibility of phrasemes in dictionaries. These problems are partially due to the lack of knowledge of the linguistic status of phrasemes. All dictionaries of Kinyarwanda reflect a homogeneous idea of the concept of phrasemes whereas it could be seen as a

generic term covering different categories of phraseological items with different requirements as to their lexicographic representation with respect to the various situations of use. Very few lexical semantic studies on phrasemes have been conducted (Coupez 1985; Bizimana 2002; Munyakazi 1992; Nkusi 1986, 1986a; Nikuze 2006, 2008). However, a major breakthrough was made by Nikuze (2006, 2008) who classified Kinyarwanda phrasemes that were previously globally referred to as one class (*inshóberamáhaánga*, idioms), into different categories. Although the results of these typological studies have not yet been reproduced in any published dictionary of Kinyarwanda, such lexical and semantic studies of phrasemes are uncontestedly important to lexicographic analysis of phrasemes as they help define various types of phrasemes and determine the nature of data to be provided in their treatment. Thus, these studies have bridged the gap of knowledge about the wide-ranging nature of Kinyarwanda phraseology and have provided solid arguments for considering different identified classes of phrasemes as lexicographic data categories.

Nevertheless, such knowledge is nothing in lexicography when dictionary compilers do not make use of fresh insights of metalexicography and do not foresee the specific needs and situations of the target group that play a decisive role in the successful transfer of lexicographic data. The lexical semantic knowledge alone reveals nothing of the way the user gains access to data presented in a given dictionary. The way dictionary users cost effectively access the data presented in a dictionary and how this positively affects the retrieval of the needed information is the main concern in dictionary compilation. A dictionary with complicated access will be of no help to the user who engages in punctual consultation even if it contains valuable data. In fact, insights from theoretical lexicography, e.g. lexicographic function(s) and structural components of dictionaries play a decisive role in ensuring rapid access to phraseological data and successful retrieval of information from a given dictionary. Therefore, apart from the lexical semantic peculiarities of each type of phraseme, the projected model will seek to account for various situations in which users experience dictionary consultation needs and viable venues of phrasemes across various structural components in a given dictionary.

In this regard, interesting suggestions for making the presentation and treatment of phrasemes in a dictionary more user-friendly have been made. One of these suggestions is the proposal by Gouws for the sublemmatisation of multiword expressions as part of the dictionary macrostructure through niching or nesting procedures (Gouws 2003, 2007, 2010), or the use of non-integrated microstructures, or the back matter of a dictionary as possible venue for fixed expressions (Gouws 2006). Gouws has defended the lexical status of multiword expressions, particularly of idioms. The strict lemmatisation of

phraseological units was applied by Chimhundu (2001) in a dictionary of Shona as a viable lexicographic inclusion approach and the challenges posed by such an approach were outlined in and discussed by Chabata (2002) and Mpofu (2002, 2007). Other researchers such as Mabika-Mbokou (2002, 2006) and Louw (2006) have addressed the issue of lexicographic representation of various types of phraseological items in dictionaries. Outside the African languages, theoretical discussions on lexicographic representation of phrasemes have also attracted the attention of various researchers working in both linguistics and metalexicography, e.g. Benson 1990; Hausmann 1991, 1997; Hausmann and Blumenthal 2006; Heid 2004; Patsala 2006; Polguère 2000; and Vrbinc and Vrbinc 2011, just to mention a few. Most of these studies have resulted in innovative proposals towards a transformational view of phrasemes in lexicography.

Unfortunately, not only is the wide-ranging nature of phrasemes not reflected in the representation of phrasemes in Kinyarwanda dictionaries but these dictionaries also leave room for improvement as far as the application of metalexicographic principles is concerned. Moreover, despite the lack of an appropriate approach to the lexicographic representation of phraseological expressions in all LGP dictionaries of Kinyarwanda, so far no scientific investigation has been carried out to address this issue. Given the abovementioned shortcomings such a study is desperately needed in order to improve the quality of current dictionaries and create better ones.

This research will seek to adapt and apply those theoretical suggestions to the lexicography of Kinyarwanda and will devise lexicographic proposals that could help dictionary compilers to solve most theoretical and practical problems facing them when treating phrasemes in reference tools. Therefore, this thesis will investigate different aspects of the lexicographic representation of phrasemes and will propose different ways in which dictionary structures and lexicographic functions can be utilised to render the lexicographic representation of different types of Kinyarwanda phrasemes more user-friendly.

## **1.4 Objective and scope of the study**

With the view of finding a solution to the abovementioned shortcomings, the overarching objective of this research is to formulate a theory-led and lexicographically-driven model that could be used to improve the representation and treatment of phrasemes in the existing Kinyarwanda dictionaries as well as in future dictionaries. For this goal to be attained, this study seeks to achieve the following specific objectives:

- To review the current lexicographic representation of phrasemes in Kinyarwanda lexicography and to highlight and address access-related problems caused by inappropriate practices in the lexicographic representation of Kinyarwanda phrasemes.
- To present and discuss lexicographic theories underpinning sound lexicographic practices and to highlight the importance of theoretical insights and lexicographic tools to the conceptualisation and compilation of a dictionary.
- To formulate and propose different models for lexicographic representation of Kinyarwanda phrasemes (collocations, idioms, pragmatemes and proverbs) taking into account various situations of use and the efficient use of different structural components of a given dictionary.

This thesis is about theoretical lexicography with a special focus on the lexicographic treatment of phrasemes. Firstly, it relies heavily on theoretical considerations from the theory of lexicographic functions and the theory of dictionary structures, also known as a general theory of lexicography, in order to formulate a model for the lexicographic representation of Kinyarwanda phrasemes to assist the dictionary user to quickly access phraseological data. Given the role of the user perspective in the present study, it is necessary to grasp as accurate as possible related concepts such as lexicographic functions, user profile, user situation and user needs and how these theoretical considerations related to these aspects influence the representation of phraseological data in a dictionary. The lexicographic representation of the lexicon is a vast domain encompassing not only phrasemes but also mono and sublexical units. The present study is not concerned with the lexicographic modelling of all classes of lexical units. It focuses exclusively on phrasemes due to multiple challenges facing lexicographers when processing them during the compilation of dictionaries and those facing users when endeavouring to gain access to phraseological data. This class of phrasemes as envisaged in this research includes four subcategories, i.e. idioms, collocations, proverbs and pragmatemes; it does not refer to compounds formed by composition procedures that, in the lexicography of Kinyarwanda, do not pose problems as they are written as monolexical units and lemmatised as primary level treatment units. The dictionary contemplated here is a general dictionary because it is not only the type of Kinyarwanda dictionary in which the representation of phrasemes is highly problematic but also the one which is most developed in Kinyarwanda lexicography. Although the metalexicographic insights formulated in this research may be extended to dictionaries with phrasemes as subject matter, phrasemes are above all envisaged in the framework of a general monolingual, hybrid or bilingual dictionary.

However, although this thesis utilises a classification of phrasemes developed within the framework of the Meaning-Text Theory, its main goal is not confined to a linguistic analysis of phrasemes. It is above all a lexicographic study and goes beyond linguistics by invoking relevant theories of lexicography. It considers the lexicography as a science in its own right that could where applicable, utilise insights from other disciplines, including linguistics, as any other science would do. The linguistic information will be used in a way that it sheds light on the on-going lexicographic investigation and does not undermine the independent status of lexicography. Thus, I will not expound too much on the description of the Meaning-Text Theory although I utilise a typology of phrasemes developed in this framework. Therefore, the scope of this thesis will revolve on the lexicographic inclusion and treatment of phrasemes with respect to their different types and according to relevant lexicographic functions and dictionary structures.

## **1.5 Hypotheses**

It has been argued that current practices of lexicographic representation of phrasemes in Kinyarwanda dictionaries are user-unfriendly in that they make it difficult to access the phraseological data presented. In addition, the importance of some types of phrasemes is overlooked, resulting in them being mistakenly left out or under-represented in the compilation of Kinyarwanda dictionaries. Faced with such problems, it is anticipatively argued that theoretical lexicography offers innovative tools and suggestions that if applied could enable a user-friendly presentation of phrasemes in general dictionaries of Kinyarwanda.

Secondly, while recognising the primacy of metalexicography in providing the guiding principles for a user-driven representation and treatment of phrasemes in Kinyarwanda dictionaries, this study claims that the linguistic typology of phrasemes plays a certain role in identifying and defining the corresponding lexicographic data categories or data types and that the lexicographic representation models to be designed will be helpful as long as they take cognisance of the wide-ranging nature of phrasemes. It therefore assumes a relative complementarity between LGP lexicography and linguistics.

## **1.6 Theoretical approach**

The title of this thesis "*Towards a Theoretically Motivated Model for the Lexicographic Representation of Kinyarwanda Phrasemes in General Kinyarwanda Dictionaries*" reflects the importance of lexicographic theories in achieving its objective. As far as lexicography is concerned, it is important to remember that the history of lexicography has witnessed the development of two

competing theoretical conceptions of lexicography (Heid 2004: 730). The first conception regards lexicography as a sub-discipline of applied linguistics dealing with the art of dictionary compilation, using principles developed in the framework of lexicology and semantics, not its own scientific principles, e.g. Mel'čuk, Clas and Polguère (1995). This conception is deeply entrenched in Western lexicography that views lexicography as applied lexicology, an art without a theory.

The second conception and the most compelling one, considers lexicography as an independent discipline, not a sub-field nor a field of applied linguistics. As a scientific discipline and practice of dictionary compilation, lexicography has its own subject matter, the dictionary, and its own methodology or process leading to dictionary production through five major phases of dictionary compilation: the preparation phase (dictionary conceptualisation and planning), the phase of acquisition of the material, the phase of preparation of the material, the phase of processing or exploitation of the material and the phase of the preparation of the manuscript for publication (Mihindou 2006: 23; Wiegand 1998:151).

During this process lexicographic research draws attention to a set of theoretical considerations including dictionary users, structures, content, user situations, genuine purpose and functions of a dictionary that determine to what extent a given dictionary will meet the reference needs of its target users and to ensure a cost-effective retrieval of the required communication or cognitive information. In this perspective, the reflections focus on the way dictionary data are distributed and positioned in the word book; the policy of lemma selection and presentation, macrostructural, microstructural and mediostructural organisation, briefly how data are presented for whom, with which needs, and in which situation of dictionary consultation. This theoretical shift was stimulated and convincingly asserted by the works of prominent scholars in lexicography including Zgusta 1971;Wiegand 1984, 1998; Bergenholz 2003; 2012; Bergenholz and Gouws 2010; Bergenholz and Tarp 1995, 2003, 2004; Gouws and Prinsloo 2005; Nielsen 1999; Nielsen and Mourier 2007; Nielsen and Tarp 2009; Tarp 2000, 2002, 2004, 2004a, 2005, 2008. These scholars consider dictionary compilation as a scientific practice that is guided by insights and principles derived from theoretical lexicography, and where relevant, insights derived from other sciences, including linguistics (Soami 2010: 9). In addition, they argue in support of the independent status of lexicography which is regarded as an independent two-fold discipline (theoretical and practical lexicography), not a sub-domain of applied linguistics. Among the most influential theories of lexicography developed in this conception are the functional theory of lexicography and Wiegand's general theory of lexicography.

Most of the concepts and metalexicographic principles used in this thesis stem from the functional theory of lexicography developed by Bergenholz and Tarp, and the structural component of Wiegand's general theory of lexicography (Wiegand 1984, 1996) in which Wiegand elaborates on different structures that are negotiated in dictionary conceptualisation and realisation.

Although these theories emphasise the independent status of lexicography and conceive the dictionary as utility tool, they oppose one another with respect to preferred methods and the ultimate outcomes of lexicographic research. The functional approach to lexicography presents itself as a transformative theory of which the point of departure is not an existing dictionary, but the users in their diverse types, situations of use and needs on the one side, and the dictionary function on the other side. Instead of confining the investigation to the observation of existing dictionaries (contemplative approach), research in the transformative approach provides guidelines in consideration of which new dictionaries can be compiled and existing dictionaries revised by taking into account a specific group of users and various situations in which dictionary consultation needs arise. Of course, the contemplative approach is also useful in that it allows identifying where existing dictionaries should be improved. Therefore, these approaches are not mutually exclusive; rather, they can be used in a complementary way as is shown in this research.

Keeping with the foregoing, both of the above-mentioned perspectives have been useful in this research. Concepts from the functional theory of lexicography and the theory of lexicographic texts have been accounted for in the proposed models to ensure that the data distribution and access structures of the proposed models enable easy and function-led access to phrasemes in the dictionary. In addition, the trans-textual approach to lexicography developed by Gouws (2007a) has been utilised in this research to determine an appropriate venue for different types of phrasemes and the adequate inclusion strategy for each one. A presentation of the main concepts of these theories is provided in Chapter 3. Occasionally attention is drawn to other pertinent ideas by scholars working from different theoretical backgrounds of linguistics, such as the Meaning-Text Theory and Cognitive Linguistics.

## 1.7 Methodological approach

For research results to be replicable and to assure objectivity, precision and validity, the methodology employed in collection and analysis of data used in the research must be set out in a detailed way (Bhattacherjee 2012: 5). To this end, this section outlines various methodological strategies used in data

collection and analysis for the present study. It explains the types of data involved and how they were collected and used in the overall argumentation of the thesis.

As far as types of data and data collection strategies are concerned, this study draws mainly on secondary sources of data. In the context of social science research, secondary sources of data<sup>1</sup> consist of scholarly studies that are realised by recognised experts in the relevant fields by using primary data collected for the purpose of the study and secondary sources heretofore available on the topic in question. The data are obtained through library or documentary research and provide analytical tools and a basis for the researcher's argumentation. Not only have the data constituted the basis of this study but they have also been used in support of the overall argumentation of the thesis.

The study deals with the lexicographic representation of phrasemes in Kinyarwanda dictionaries. In view of the purpose of the study, the first step of data collection has been the identification of dictionaries which would constitute the basis of investigation. Four general dictionaries were selected as corpus of the study based on their importance in terms of language representativeness and their use and acceptability within the speech community. All these dictionaries have been realised and published by the language section of the Institute of Scientific and Technological Research of Rwanda (IRST as French acronym) and display the following functional characteristics:

- *Inkoranya y'ikinyarwaanda mu kinyarwaanda or DIMO (First edition 2002, Second edition 2011)*: a monolingual polyfunctional (reception and production) dictionary designed for native speakers with a high level of language proficiency,

---

<sup>1</sup>About that, a comment is necessary to avoid the confusion that could be created by the use of a term originating from social sciences research (secondary source of data) in an essay on metalexicography wherein the same term is used with a different meaning. According to Soami (2010: 23) (relying on Wiegand 1984), in lexicography primary, secondary and tertiary sources of data refer to different sources of materials used as the basis of dictionary compilation. The lexicographic primary source of data include computerised corpora build from both written and oral texts, raw data collected using questionnaires and interviews in user research, etc, while *secondary sources of data* relate to the use of existing dictionaries as basis in compilation of new dictionaries (Nkomo 2008: 32). Tertiary sources of data relate to works done in the field of linguistics (Soami 2010: 23).

- *Inkoranya iciriritse y'ikinyarwanda or DIEL* (2010): a monolingual polyfunctional (reception and production) dictionary designed for native speakers and learners with low and medium level of proficiency,
- *Dictionnaire rwandais-français de l'INRS* (Vol. 1 :1983, Vol. 2 :1985, Vol.3 :1987) : a bilingual monodirectional (Kin-Fr) and monofunctional dictionary for text translation into French,
- *Inkoranya y ikinyarwaanda mu kinyarwaanda nó mu gifaraansá/ Dictionnaire rwanda-rwanda et rwanda-français* or DIFO (2005): a hybrid (Kin-Kin-Fr) polyfunctional (text reception, production and translation) dictionary designed for both native and foreign language speakers with a high level of proficiency in Kinyarwanda.

In addition, the literature pertaining to phrasemes in Kinyarwanda linguistics, lexicography and metalexicography was explored and perused in order to clearly identify and define the gap of knowledge justifying the need for this research. More particularly, a contemplative study of these works was realised in Chapter 3 where I have pinpointed and discussed shortcomings in the representation of phrasemes in the above-listed dictionaries and highlighted the misconception of *phraseme* in both linguistics and lexicography of Kinyarwanda.

In keeping with the transformative view of this research, however, it is essential to show how the proposed model could be applied in improving the representation of phrasemes in existing dictionaries as well as in new dictionaries to be created. To this end, the proposed model is taken into consideration in the conceptualisation of the on-going hybrid Kinyarwanda-Kinyarwanda-English dictionary with text reception, text production and text translation as main functions.

In this research, qualitative and hermeneutic methods were used for data analysis and the construction of arguments. This study formulates a model which will be implemented in the lexicographic treatment of phrasemes using existing theories and existing knowledge. The theory-testing orientation of this research compels me to rely heavily on a deductive and qualitative approach to model building and adaptation of the new knowledge to the area of study. As the bibliography shows, the researcher has had to collect, extract and interpret data from an extensive selection of literature related to various theories used in this research. The information retrieved from data acquired through documentary research has provided theoretical insights and guidelines and supported the argumentation of the thesis in all chapters of the study.

In this research, lexicographic theories and to a certain extent theoretical linguistics, were used. Firstly, both phrasemes and phraseologisation, e.g. the process leading to the formation and operationalisation of phrasemes are linguistic phenomena. As Chabata (2009) has noted, the role of linguistic insights in making informed decisions in the lexicographic description of the language cannot be overlooked, particularly in LGP lexicography. This point is also stressed by Rundell (2012: 60-71) while discussing the relevance of linguistics insights in the process of dictionary making. In this line of thought, this study assumes a relative complementarity between linguistics and lexicography in the formulation of the lexicographic model for inclusion and treatment of phraseological units. It claims that the lexicographic presentation of Kinyarwanda phrasemes, i.e. the synthesis of phraseological data necessitates a prior linguistic categorization and characterisation (i.e. the analysis of data). To this end, this study uses the concept of *phraseme* as defined in the framework of the Meaning-Text Theory. It considers also the MTT's classification of phrasemes into mainly idioms, proverbs, collocations and pragmatemes.

Regarding the lexicographic approach to this research, I have collected various theoretical considerations relevant to the area of study from both functional theory and structural theory of lexicography. These metalexicographic insights are outlined and discussed in Chapter 2 on the theoretical framework and they are reflected in the core chapters of the thesis where they have been referred to or applied to generate the proposed models. A lexicographic model for each type of phraseme has been designed, taking into account the role and requirements of the respective types of phrasemes in participating and responding to users' needs in different situations of dictionary consultation, e.g. lexicographic functions. The proposed models for the types of phrasemes considered in this study have been applied to trial articles of a few selected lemmata and are discussed in Chapters 4, 5 and 6.

## 1.8 Background to Kinyarwanda

This section underlines both the demographic and political importance of Kinyarwanda not only as second lingua franca in the African region south of the equator, but also as national and official language of Rwanda. It also provides an overview on the linguistics and classification of Kinyarwanda.

### 1.8.1 Sociolinguistic profile of Kinyarwanda

Kinyarwanda is the national language of Rwanda spoken by more than 99.7% of Rwandans, and one of three official languages besides English and French. According to Kimenyi (Kimenyi 2004: 1527), it is

probably, after Kiswahili, the second largest linguistic cluster in the Bantu group. It is a sister dialect of Kirundi, the national language of Burundi and Giha, Shubi, Vinza and Hangaza, other dialects spoken in Tanzania. Its speakers are estimated to be more than 20 million people. According to the last census conducted in August 2012, Rwanda has around 10,537,222 people and Burundi approximately 10,216,190 inhabitants. In addition, Kinyarwanda is intelligible for speakers of Giha, a language spoken in North-West Tanzania. It is also spoken by ethnic Banyarwanda in Southern Uganda (Bafumbira) and in the Southern and Northern Kivu Provinces (Democratic Republic of Congo).

On the endocentric level, Rwanda is a monolingual nation, Kinyarwanda being the only national language of Rwanda. However, this simplicity hides dialectological complexity. On a micro-dialectological level, one can point out the existence of various regional varieties of Kinyarwanda, which differ mainly on phonological, semantic and morphological levels but are similar enough to guarantee the intercommunication between Rwandans.

The main dialect of Kinyarwanda is Central Kinyarwanda which is spoken in central, southern and eastern Rwanda and has been standardized and made the official dialect of Kinyarwanda. It is used in administration, education, media and research. It is also this dialect which is used in Kinyarwanda lexicographic activities. Among other dialects, there are also Kireera, used in North-Western Rwanda, and Kigoyi used in the far north-western region of Rwanda. “Central Kinyarwanda” is also called “Kinyanduga” because its area corresponds to the former historical Kingdom of Nduga that used to cover the most southern and central parts of the current Republic of Rwanda. Nevertheless, this last appellation (i.e. Kinyanduga) is avoided because it recalls the socio-political tensions between Bakiga and Banyenduga that characterized Rwanda in the 1980s and early 1990s, in the context of which the term Kinyanduga was derisively used to designate the dialect of the central and southern population. This politically-motivated conflict opposed the Northern region, where most of the political leaders of the time came from (using Kirera and Kigoyi, two dialects also called Gikiga with reference to Abakiga: people living in high altitude regions), and the central-southern regions that were considered as being politically disadvantaged. With these dialects being mutually intelligible, it is impossible to define precise delimitations since a speaker of a given dialect is understood when in the region where a different dialect is used. This intelligibility creates a diglossic situation.

On the exocentric level, Kinyarwanda coexists with three important foreign languages, i.e. French, English and Swahili. English and French together with Kinyarwanda are official languages of the Republic of Rwanda. Recently, there has been a shift in the roles played by these two western

languages, and when one looks at the place English is taking, especially since the 2008 government decision to drop French from education as language of instruction in favour of English, it is very likely to have more influence than French in the future (Kayigema 2010:1). However, the recent resumption of diplomatic relations between Rwanda and France may reshape this language change.

The Swahili language is used in business, particularly in towns, for radio broadcasting and is taught in secondary schools. It is also to some extent used in the army to give military orders, alternating with English and Kinyarwanda (Hitayezu 1999; Ntakirutimana 1996, 2002; Ntawigira 1997; Nyiramahirwe 1999; Sugira 2004). Brought to the country by German colonisers in the beginning of the 20<sup>th</sup> century, and maintained by Belgians until 1929, this language remained in use in commercial towns with high concentrations of Arabs and Indians (Kigali, Gisenyi, Butare, Cyangugu and Ruhengeri) and by some religious confessions (Adventists) before being revived during the second Republic since 1973 (Niyibizi 1980:116-147; Ntakirutimana 2007:12-21). The admission of Rwanda in the East African Community has more than ever stirred the Rwandans' interest in learning the Swahili language.

Despite the lack of detailed demolinguistic statistics of Luganda in Rwanda, one could also note the increasing number of people speaking this language in Rwanda (Rwanda Focus, 1<sup>st</sup> April 2009). This language is used by Ugandans living in Rwanda for business purposes and Rwandans who used to live in Uganda, or who have learnt Luganda in various circumstances. More particularly, Luganda and Kinyarwanda are code-mixed in the area of mechanics (car repair workshops), but also in bars and restaurants if not in street conversations, especially in Kigali City and Nyagatare town. In addition, many Rwandans cross over to Uganda for visiting and business purposes, which is for them an opportunity to learn some Luganda words and sentences. Furthermore, some parents take their children to Ugandan schools, seeking a better grounding in English for them now that English has become the high official language of Rwanda. At the same time, these children learn to speak Luganda. With Rwanda and Uganda signing a pact to remove travel barriers between their people, the number of Luganda users in Rwanda might considerably increase in the near future.

Generally, Kinyarwanda is still the predominant language of the country. It is widely used in every day communication, in national and international media, in business, in the army and the police. It is also the dominant language of administration, especially in local governments. In education, it has been used as instruction medium until primary six, but today it is used until primary four, while it continues to be taught as a subject.

The existence and use of these foreign languages in Rwanda alongside Kinyarwanda creates a multilingual situation wherein code-switching and lexico-semantic borrowing phenomena between Kinyarwanda and the three other languages, i.e. English, Swahili and French have developed. Apart from this, there are also problems related to the multilingual education system and language planning. Within this scenario, different lexicographic teams have started working on bilingual dictionaries in order to provide urgently needed reference tools for language learning and translation.

### **1.8.2 Classification and designation of the language**

Kinyarwanda has been known for a long time in language classifications. According to Nkusi (1995:7), it was classified in 1948 by Guthrie (1948: 40-42) as “nyarwanda” occupying the area 61 in D linguistic zone (i.e. linguistic zone D, group 6, 1<sup>st</sup> position) with rundi (D62), fuliru (D63), shubi (D64), Hangaza (D65), Ha (D66) and Hanza (D67). This alphanumeric coding system is a Guthrie innovation and its classification relies mainly on geographical distribution rather than genealogical relations. Kinyarwanda is labeled D61 in Malcolm Guthrie’s classification of African languages.

Joseph Greenberg, another prominent scholar in linguistic classification, has roughly classified Kinyarwanda in the Congo-Kordofan family, in the Niger-Congo sub-family, Benue-Congo branch, bantoïd sub-branch, Bantu group. Later in 1999, Bastin, Coupez and Mann, linguists of the Tervuren Royal Museum for Central Africa, suggested a genealogical classification of Bantu languages, using a lexicostatistic approach. Their study draws attention to the degree of lexical proximity among Bantu languages (Bastin et al. 1999). In the subsequent classification, Kinyarwanda is labelled J61 following amendments made to the Guthrie’s D group by Tervuren Museum’s researchers.

For the sake of systematization, Bantu linguists have opted for the use of bare or prefixless forms (Rwanda, Ha, Rundi, Zulu, etc.) to designate Bantu languages. Thus, Kinyarwanda is referred to as “Rwanda” Coupez (1980), if not “Ruanda” in Guthrie (1971) instead of the prefixed form (KI-NYA-RWANDA). This is the case with the names of many other African languages as seen in English texts without the class prefix (Swahili, Tswana, Ndebele) but sometimes seen with the prefix (Kiswahili, Setswana, isiNdebele). However, this way of designating languages and people has been challenged by other respected scholars. For instance, Nkusi (1995:14-15) points out serious inconsistencies caused by the use of bare forms, when wondering how one should interpret a sentence like “*Rwanda of Rwanda speak Rwanda*” meaning “In Rwanda, Rwandans speak Kinyarwanda”. He also emphasised the fact that different languages can share the same root and be distinguished by the prefix. For instance,

Mbundu refers to Umbundu, language of Ovimbundu people and Kimbundu, language of Ambundu people, two different languages. Moreover, in some societies, the prefixless form is used to form other ethnic based words: in BuGanda kingdom, people are BaGanda, the language is LuGanda. Locally, the language is referred to as “ikinyarwaanda” with the initial vowel (i-) which is compared to an article in English, and the language-specific prefix (-ki-).

Others like Schumacker (1956) have once used the designation “(u)runyarwanda” (prefix -ru- referring to ururimi: language/tongue and preceded by a corresponding initial vowel), but this appellation is no longer in use. For the sake of clarity, the form KINYARWANDA, mostly used in Rwandan linguistics by both Anglophone and Francophone scholars, will be used in this research. The initial vowel and vowel duration in “rwaanda” will not be indicated. Additionally, the language prefix -KI- is more significant. Nkusi (1995) rightly indicates that in the Rwandan cultural conception, mastering Kinyarwanda goes beyond a high level of proficiency in this language. As Kinyarwanda reflects the entire Rwandan authenticity and dignity, its knowledge implies mastery of customs and usage, knowing when to speak and when to keep quiet, knowing the place of elders and young people mastery of practices and techniques of daily life; briefly, all values of the Rwandan culture. Only experienced, wise and honest people know or master Kinyarwanda, and not necessarily scholars in linguistics and literature of Kinyarwanda (Nkusi 1995:14-15).

### **1.8.3 Overview of linguistic research on Kinyarwanda**

Kinyarwanda is one of few African languages that have been considerably researched. It has drawn the attention of many scholars in different academic circles in the world and from different theoretical perspectives of linguistics: structuralism, functionalism, distributional, generative-transformational, psycho-mechanic, sociological, relational, optimal and minimalist theories (Nkusi & Nkejabahizi 1997). The first linguistic studies of Kinyarwanda were initiated and undertaken by Western scholars during the colonial period (1885-1962). One of the pioneers of Kinyarwanda linguistic and lexicographic research is Father Felix Dufays who, in 1912, published *Wörterbuch Deutsch-Kinyarwanda* (Dufays 1912), a dictionary of 390 pages. One can also mention Father Eugène Hurel who, 18 years after his arrival in Rwanda, published a bi-dimensional bilingual dictionary Kinyarwanda-French in 1926 (Hurel 1926). Another important figure is André Coupez, a Belgian linguist who carried out linguistic and literature research, firstly at the Astrida (now Butare, Rwanda) branch of the former Institute of Scientific Research in Central Africa and the National Institute of Scientific Research of Rwanda, before joining the Tervuren based Royal Museum for Central Africa.

Coupez published the first ever-complete grammar of Kinyarwanda (1980) and various anthologies of Rwandan official literature (Coupez & Kamanzi 1970; d'Heltefelt & Coupez 1964). In 1959 he also initiated and led an ambitious lexicographic project that culminated in the publication of a comprehensive Kinyarwanda-Kinyarwanda-French dictionary in 2005. He has also trained many Rwandan researchers not only in linguistics and literature of Kinyarwanda, but also in other Bantu languages. His influence on Kinyarwanda linguistics and lexicography has impacted Kinyarwanda scholars at all levels, especially with research in morpho-syntax and lexicography at INRS, IPN, IRST and the National University of Rwanda (NUR). Today, the coupezian influence is still obvious in Kinyarwanda linguistics.

After Rwanda regained independence (1962), local researchers working in different academic settings took over from their Western counterparts. Since that time, Kinyarwanda linguistics have inspired hundreds of university dissertations, papers and conference presentations at Rwanda Universities and abroad. It is not possible to give a complete account here of all that has been done, even in a partial way. A Bibliography of Kinyarwanda has been compiled (Nkusi & Nkejabahizi 1997; d'Hertefelt & de Lame 1987), though these works are not exhaustive. However, one can acknowledge that most linguistic analysis levels if not all, have been thoroughly investigated and some of the results have contributed to a better understanding of African linguistics, philosophy and cultures. For Kinyarwanda morpho-syntax, there are Bizimana (1998), Cadiou (1985), Coupez (1980), Jouannet (1983), Kimenyi (1980, 2004), Nkusi (1995), Ntampaka (1992), Ntzinda (1980), Rwabukumba (1980) and Rwabukwandi (1981). Lexicology and semantics of Kinyarwanda are quite well catered for in Bizimana (2002) and Uwilingiyimana (2000). Furthermore, the phonology of Kinyarwanda has been described among others in Bizimana (1998, 2002) and Kimenyi (1979, 2002). Regarding lexicography, monolingual and bilingual/hybrid dictionaries of Kinyarwanda have been compiled, although linguistics got the upper hand over metalexicography, which is hardly surprising since metalexicographic research on Kinyarwanda with regard to both the contemplative and transformative approach is not accounted for in the available literature.

## 1.9 Overview of the thesis

This thesis contains seven chapters. Chapter 1 is the general introduction. It is used to outline the research problem, the justification of the study, the purpose of the study, the hypotheses, and the theories and methods used in the thesis among other things. In Chapter 2 the lexicographic framework of this thesis is presented. Chapter 3 contains a study of how phrasemes are approached in the

linguistics and lexicography of Kinyarwanda. In Chapter 4 the issue of lexicographic treatment of idioms and proverbs is addressed and relevant recommendations are formulated. Then in Chapter 5 the lexicographic treatment of collocations in Kinyarwanda dictionaries is discussed. A model for lexicographic representation of Kinyarwanda pragmatemes or pragmatic phrases is developed in Chapter 6 and major research findings along with arising recommendations on the lexicographic treatment of different types of phrasemes in Kinyarwanda dictionaries are outlined in Chapter 7.

## **1.10 Concluding remarks**

The aim of this first chapter was to give an overview of the thesis and define the problem areas for this research. It has provided the background of Kinyarwanda and theoretical insights around the main aspects of the research problems. However, although some remarks were given regarding theoretical approaches guiding the realisation of this work, there is a need to expound in detail on different principles and insights from lexicographic theories utilised in this thesis and to show how they will be employed throughout the study. To this end, the second chapter introduces various lexicographic theories applied in this research.

## CHAPTER 2: THEORETICAL FRAMEWORK

### **2.1 Introduction**

A user-friendly representation of lexicographic data in a dictionary can only be achieved when the underlying model is informed by sound and reliable lexicographic theories that make provision for the central role of the user in dictionary compilation. In the present investigation, such a theoretical basis is found in the theory of lexicographic functions and the structural theory of lexicography thanks to the emphasis placed on user-friendly distribution and presentation of dictionary data. In this chapter, key tenets and concepts of the functional theory of lexicography which are relevant to the purpose of the thesis are reviewed (section 2.2). In addition, relevant concepts of the structural theory of lexicography as outlined in Wiegand (1996) are also described, particularly those pertaining to data presentation, data distribution and access to data in the dictionary (section 2.3). In the same section, Gouws's functional trans-textual approach is discussed. The concluding note (section 2.4) recaps the main ideas and indicates how these theoretical insights will be utilised in the study.

### **2.2 The functional theory of lexicography**

#### **2.2.1 Background and postulates of the theory of lexicographic functions**

The theory of lexicographic functions was outlined three decades ago by a group of Danish metalexicographers led by Henning Bergenholz, Sven Tarp and Sandro Nielsen at the Centre for Lexicography of the Aarhus School of Business (Fuertes-Olivera & Tarp 2011). It was intended to set out theoretical understanding and optimise theoretical potentials of the user perspective in the general theory of lexicography by taking cognisance of the concept of lexicographic functions.

However, the concept of lexicographic functions is not an innovation of the theory of lexicographic functions. It is a result of an idea previously put forward by Ščerba in the 1930s and 1940s. The Russian scholar introduced the idea of lexicographic functions when making a distinction between active/encoding dictionaries (a category today covering both text-production and text-translation dictionaries) from passive/decoding dictionaries (today text-reception dictionaries) (Kromann, Riiber & Rosbach 1984: 208). The Danish metalexicographers further developed the concept of lexicographic functions in many publications and extended its scope beyond dictionaries (Bergenholz 2003, 2012; Bergenholz & Bothma 2011; Bergenholz & Gouws 2010, 2010a; Bergenholz & Tarp 2003, 2004; Nielsen 1999; Nielsen & Mourier 2007; Nielsen & Tarp 2009; Tarp 1999, 2000, 2002, 2004, 2004a, 2005, 2007, 2008, 2008a, *inter alia*). The function theory has inspired metalexicographic works of

other scholars such as Gouws and Prinsloo (Gouws 2004a, 2007a, 2009a; Gouws & Prinsloo 2000, 2005). The proliferation of publications on the role of lexicographic functions in the planning and compilation of dictionaries in the last two decades is evidence of the importance of this concept in the present-day and future lexicography that places the user at the centre of dictionary compilation and lexicographic research.

Bergenholtz and Gouws (2010) point out two ways in which Wiegand's general theory of lexicography has revolutionised modern metalexicography. Although he was not the first to protest against the linguistic hold over lexicography (Nkomo (2008) mentions Zgusta (1971) for example) and although his general theory of lexicography is itself criticised for having a linguistic background (Bergenholtz & Tarp 2003: 172) Wiegand (1984: 13-14) defended the independent status of lexicography as a scientific discipline with its own subject field (i.e. dictionaries) and not a sub-field of linguistics. One of the innovations brought by this epistemological shift was the study of dictionary structures, as a container of data, on top of the semantic bias displayed by the then prevailing scholastic conception confining the lexicographers' attention to the description of the meaning of words to the detriment of other categories of data in dictionaries (Wiegand & Hausmann 1989; Gouws & Prinsloo 2005: 3).

By claiming an independent status for lexicography, Wiegand also arrived at the idea that a dictionary is a utility product made in order to satisfy specific human needs. According to Bergenholtz and Tarp, this would have led Wiegand to realise that "all theoretical and practical considerations must be based upon a determination of these needs, i.e. what is needed to solve the set of specific problems that pop up for a specific group of users with specific characteristics in specific user situations," had Wiegand's theory not been biased by the legacy of linguistics (Bergenholtz & Tarp 2003: 172). As the general theory of lexicography, the function theory of lexicography premises on the fact that a dictionary is a utility tool that should be designed in a such way that it fulfils a specific function vis-à-vis the specific needs of a specific user in a specific social activity or extra-lexicographic situation (Bergenholtz 2003: 67). Further theoretical developments introduced the notion of lexicographic functions and dictionaries being viewed in terms of their suitability "to respond to the specific needs of specific users with specific reference skills in specific situations of dictionary use" (Bergenholtz & Gouws 2010a: 103).

In addition, the functional theory reiterated the independent status of lexicography, strongly arguing that lexicography inclusive of terminology is not a sub-discipline of linguistics, but is an independent discipline aimed at producing utility tools satisfying specific users' needs in specific usage situations. In this connection, Bergenholtz (2003: 67) states that:

Lexicography is a discipline in its own right; that means e.g. that lexicography is not a linguistic discipline and not applied lexicology either. On the other hand, specialised lexicography, with its synonym terminography, is part of the discipline lexicography. (...) The object of lexicography is dictionaries. That means the object is not the language, but those tools developed by mankind for certain purposes.

Yet again, the conception of lexicography as a discipline of applied linguistics or a sister of lexicology, though archaic, is still prevailing in Western lexicographic research, including in the MTT outlined in section 3.1.

The function theory of lexicography claims that linguistics contributes to the compilation of lexicographic works in the same way as do other disciplines. In this connection, Tarp (2004: 224) writes that “lexicography has a big interdisciplinary vocation and, as such, it treats - or ought to treat - linguistics in the same way as it treats any other discipline, i.e. it utilises the results from these disciplines whenever they are needed in order to conceive high-quality utility products that meet the specific needs of specific groups of users in specific user situations.” According to this view, the relation between lexicography and linguistics in the compilation of a general language dictionary is the same as the relation between lexicography and medicine in the compilation of a medical LSP dictionary. However, Bergenholz and Tarp (2009: 24) recognise that some specific theories proposed within a linguistic perspective can contain things which may be profitable for lexicography provided they are recuperated and integrated in the mould of the general theory of lexicography. Analysing the type of relationships between lexicography and linguistics, Tarp (200: 193) insists that:

Linguistics and lexicography must be viewed as two independent disciplines where linguists sometimes - but only sometimes – make use of lexicography for their work and where lexicographers sometimes make use of linguistics for their theoretical and practical work.

However, the extent to which lexicography should assimilate concepts borrowed from linguistics remains a topical issue in recent publications by the proponents of the functional theory of lexicography. For example, discussing the nature of interaction between lexicography and linguistics, Tarp (2008: 12) maintains that:

This necessary interaction with other disciplines does not mean that lexicography can automatically take over the arsenal of concepts, theories and methods used by these other disciplines. The fact that the object of study is delimited, and the fact that lexicography has its own independent core distinguishing it from other disciplines, mean that all these concepts, theories and methods must be subjected to critical analysis with a view to determining what should be rejected, what can be used, and how the useful factors can be adjusted and adapted to suit the particular nature of lexicography

In his contribution published in a volume devoted to the problem of classifications in linguistics and lexicography that he co-edited with Götz-Votteler and Herbst, Bergenholz (2009) argues against the uncritical re-using of linguistic concepts, terminologies and the linguistic classifications in lexicography, since lexicography is no longer viewed as a sub-discipline of linguistics. For instance, Bergenholz (2009) upholds that the concepts of collocation or part-of-speech are not lexicographically relevant (cf. Tarp 2010: 461). Then Bergenholz (2009: 275) states that “different terminological systems are needed for different aims: the only criteria by which a proposed scheme of classification can be judged is that of its effectiveness in attaining a certain fixed aim.” Likewise, in the same volume Tarp (2009b) upholds that instead of relying on concepts developed in linguistic context such as polysemy and homonymy distinctions, lexicography as an independent discipline should endeavour to develop its own theory, concepts and approaches. Nevertheless, the editors of the volume are aware of the futility of any attempt to systematically reject all linguistic terms that have come to be part of lexicographic theory due to epistemological and historical reasons (Bergenholz, Götz-Votteler & Herbst 2009). Furthermore, the independent status of lexicography is still subject to debate and many counter-arguments to the rejection of linguistics in lexicographic research and in dictionary compilation have been developed (Rundell 2012; Tono 2010).

Of great importance in the functional theory is the claim that it is not a contemplative theory, but a transformative and revolutionary theory, cf. Tarp (2004) and Bergenholz, Nielsen and Tarp (2009). It is transformative in the sense that it goes beyond the study and observation of the already existing lexicographic works by way of producing theoretical guidelines for the review of existing dictionaries and the making of future improved dictionaries. It is not a passive theory; on the contrary, it is an interactive theory as it influences the whole process of making reference works while being reshaped by the insights gained from analysis of reference works that among other things, contribute to the improvement of the theory itself. The dynamic nature of the function theory of lexicography is reiterated by Gouws (2009:3) who, as referred to by Fuertes-Olivera and Sven Tarp (2011: 142-143), claims that the “needs of dictionary users are constantly changing and these changing needs should compel lexicographers to continue developing lexicographic theory and improving the quality, relevance and efficiency of dictionaries.” Tarp (2009: 279) strongly affirms that “for research into dictionary usage to be relevant, it should not only generate knowledge of *how* dictionaries are used, but also of *who* the users are, *where*, *when* and *why* they use dictionaries, and with *which* result.” To put it differently, such research should go beyond the only contemplative observation of how users consult the already existing dictionaries to encompass other lexicographic aspects such as the *user situations*,

the types of *users*, the types of *user needs*, the functions of dictionaries and the degree of satisfaction of the user needs among other things. In fact, there is a body of literature in the sphere of functional theory of lexicography where these concepts are comprehensively discussed and that can be referred to for relevant explanations and applications of these concepts (Bergenholtz 2003; Bergenholtz & Gouws 2010a; Bergenholtz & Tarp 2003; Fuertes-Olivera & Bergenholtz 2011; Gouws & Prinsloo 2005; Nielsen 2008 and Tarp 2000, 2004, 2004a, 2008, 2008a, 2009, 2009a,b,c,d, 2010, 2010a *inter alia*). A critical description of data collection methods in dictionary usage and user needs research is provided in Tarp (2009c: 283-290). More particularly, Bergenholtz (2011) and Bergenholtz and Johnsen (2007) provide a state-of-the-art analysis of the potentials of log files as an innovative survey approach in data collection on Internet dictionaries use (cf. also Kilgarriff 2012). Therefore, the following discussion will be brief and purposefully confined to only introducing the already mentioned concepts.

### **2.2.2 Key concepts of the theory of lexicographic functions**

As Bergenholtz, Nielsen and Tarp (2009: 27) claim: “The selection and preparation of needs-adapted data can only be done in a scientific way if both the user profile and the social situation in which the needs occur are taken into account.” Thus, research in extra-lexicographic situations where user needs occur, is strongly emphasised in the function theory of lexicography if lexicographers are committed to ensure the achievement of the genuine purpose of the dictionary. This section gives an overview of key concepts of user research as envisaged by the functional theory of lexicography.

#### **2.2.2.1 Types of users**

According to Tarp (2004: 225), the functional theory of lexicography is a user-oriented theory and, in this status, it pays much attention to different characteristics of users of different dictionaries. Echoing Tarp, Nkomo (2008: 53) points out that during the stage of dictionary planning, it is necessary to identify the type of potential users of the projected dictionary and to work with that user in mind throughout the whole process of dictionary compilation. Discussing user profile, Bergenholtz and Tarp (2003: 173), Tarp (2000: 194-195) and Nielsen (2008: 180) provide a list of questions that may assist the lexicographer in detecting as reliably as possible the characteristics of the targeted group of users when planning the compilation of dictionaries. These questions can be grouped into four categories according to specific domains in which the users’ competence is required.

- About the users’ competence in the mother tongue and foreign language**

- Which language is their mother tongue?
- At what level have they mastered their mother tongue?
- At what level have they mastered a foreign language?

- **About the users' competence in general cultural and encyclopaedic knowledge**
- What is the level of their general cultural and encyclopaedic knowledge?
- **About the users' level of knowledge in the special subject field**
- At what level have they mastered the special field in question (lay person, semi-expert or expert)?
- **About the users' proficiency in the relevant LSP**
- At what level have they mastered the corresponding LSP in their mother tongue?
- At what level do they master the corresponding LSP in the foreign language?

In addition, when speaking of learners' dictionaries in foreign language settings, Tarp (2003: 225-228; 2004: 225, 229, 303) describes guiding elements that can help distinguish between various types of potential users during the conception phase of a foreign language learners' dictionary. These characteristics include:

- Proficiency level (whether the learner is a beginner, an experienced or advanced learner)
- Adult or child
- Emigrant or other type of learner
- "Joint" or separate culture
- Level of general culture
- Conscious or spontaneous learning
- Learning inside or outside the language area
- Learning whilst being exposed or not to the mother tongue
- Learning with or without contrastive methodology
- Learning related or unrelated to a specific textbook or didactic system
- Learning related or unrelated to a particular topic (economy, history, culture etc.)

The determination of the user profile based on those parameters influences the amount and types of data presented in the dictionary in question. In addition, as it can emerge from the above considerations, the characterisation of users may revolve around the level of linguistic and encyclopaedic competence in the mother tongue and in the foreign language, the age, the level of education and the exposure to the subject field in the case of an LSP dictionary. Competence in foreign language can be a useful parameter in detecting the potential users of some dictionaries (Nkomo 2008: 65). In addition, the level of cultural and encyclopaedic knowledge and the level of language

competence as well as the knowledge in specialised subject fields determine the ability of users to retrieve the needed information from dictionary data. However, although important, the sole determination of the type of user is not enough in itself. In addition to user profiling, the lexicographer should seek to know the social situations in which users experience problems that make it necessary to consult a dictionary.

### 2.2.2.2 User situations

Bergenholtz, Nielsen and Tarp (2009: 26-27) strongly believe that for research in user needs to be effective, a thorough investigation must be conducted beyond the research on dictionary use as defined by Wiegand (Gouws & Prinsloo 2005: 4; Nkomo 2008: 42), but “in an extra-lexicographic social situation which takes place before and independent to this consultation.” Tarp (2000: 195) insists on detecting not only the potential user of the dictionary, but also and most importantly, the situations in which users may have recourse to the dictionary in question. Tarp distinguishes two types of situations: the knowledge-oriented situation and the communication-oriented situation or cognitive and communicative situations (cf. Nkomo 2008: 54).

In the first case of knowledge-oriented situations, the user has recourse to the dictionary in order to obtain knowledge about a special subject or to study a foreign language (cf. Tarp 2000: 195 and Fuertes-Olivera & Tarp 2011: 145-150). Nkomo (2008: 54) specifies that such “information may be cultural and encyclopaedic, specialised in respect of specialised subject fields or disciplines, or it may be related to a specific language.” On the other side, in communication-oriented situations, the user consults the dictionary in order to resolve communication related problems such as text production, reception, correction and translation (Tarp 2000: 195 and Nkomo 2008: 54). Tarp and Bergenholtz and Bothma have recently expanded the scope of user situations beyond the usual communicative and cognitive situations towards what are called operative situations (cf. Bergenholtz & Bothma 2011: 62; Tarp 2007: 177-178; Tarp 2008a) and interpretative situations (Bergenholtz & Bothma 2011: 62).

In **operative situations** the potential user has a need for assistance in connection with physical or mental operations. There is a need for guidelines for what and how the user has to act in order to carry out a certain instruction. The most commonly used information tools in operative situations are user guides, manuals, road maps and search engines.

In **interpretative situations** the potential user has a need for assistance to interpret a non-language sign of some kind, e.g. a traffic sign. The right interpretation is normally needed to be able to act in the correct way. The most commonly used information tools in interpretative situations are information material from authorities and different organisations (Bergenholtz & Bothma 2011: 62).

The theoretical implication of this new development for the lexicography of the near future is to see the scope of lexicography go beyond the traditionally known lexicographic reference works to encompass all reference works having in common with lexicography of being “tools conceived to be consulted by specific types of users in order to satisfy specific types of information needs in specific types of social situations” (Tarp 2007: 177). However, while both cognitive and communicative situations are well documented in functional lexicography literature, the theorisation about the latter two situations is still at an embryonic level. Whether their place in the functional lexicography will be firmly ascertained and what changes this will lead to in both the theory of functions and the lexicography practice, only time will tell.

### **2.2.2.3 User needs**

Tarp (2000: 195) advocates that every dictionary should be devised in such a way that it meets the lexicographic needs that arise in a specific situation. This focus on user needs is also echoed in Gouws (2004: 265) who writes: “As practical tools dictionaries should be devised in such a way that a well-identified user can retrieve the kind of information needed for a specific purpose.” In the same vein, Tarp (2007: 172) maintains that “the only types of needs that can be met by dictionaries and other lexicographic tools are the needs for information. Users consult dictionaries in order to obtain information which they can then use for a great variety of purposes, e.g. to solve communication problems or to add to their existing knowledge. However, dictionaries themselves do not contain information, but only lexicographically selected and prepared data from which the users may or may not be able to retrieve the needed information.” The processing of the dictionary data to retrieve relevant information remains in the hands of the user who must employ his encyclopaedic and linguistic knowledge to achieve a right interpretation. However, a prior screening of the target user taking into account both linguistic and cognitive aspects, attenuates the occurrence of cognitive mismatches which could impede the success of dictionary consultation.

Some of the user needs that have been identified (cf. Tarp 2000: 195-196) include the situations in which the user looks for information about the native language or about a foreign language and where he/she wants to make a comparison between these languages. In the native language, the information needs may arise in instances of both oral/written text reception and production. Likewise, dictionary data may be sought to satisfy information needs facing the user when performing proofreading of texts or marking tasks in the mother tongue (cf. Tarp 2008: 71-74).

The user may also consult a dictionary in order to extract information about the culture and the world in general or about a special subject field. Tarp mentions the comparison between the subject field in the native and foreign culture, the search for information about the native LSP or about the foreign LSP, and a comparison between the native and foreign LSP as possible user needs that lexicographers should expect to find during a user needs assessment prior to dictionary development. These are primary needs, but Tarp points out the existence of secondary needs that are not activated until a potential user turns into an actual user who needs guidance and education about its effective utilisation.

Secondary-related needs or usage-related needs as Overgaard Ptaszynski (2009) refers to them are normally catered to in the outer-texts of dictionaries. These are for example, information about the structure and the use of dictionaries. However, this clear-cut distinction between primary needs and secondary needs so prevalent in the work of Tarp is reconsidered and seriously challenged by Overgaard Ptaszynski (2009: 246-259) who alternatively proposes a more dynamic conception of these categories. Be that as it may, a reliable determination of the needs of the target users is a pre-condition in achieving a user-oriented lexicography.

#### **2.2.2.4 Genuine purpose of a dictionary**

Gouws (2004: 265) argues that any dictionary should fulfil a specific purpose, the identification of which must be done before lexicographers undertake the compilation phase, as the latter must be steered by that specific purpose. According to Gouws (2004), this prerequisite leads to the concept of the genuine purpose of a dictionary, as earlier discussed by Wiegand (1998) and later redefined by Tarp (2000). Gouws (2004: 265) provides a more expanded definition of the genuine purpose of a dictionary in the following words:

The genuine purpose of a dictionary implies that a dictionary is produced so that the target user who uses the dictionary in a typical usage context will have an instrument to assist him in achieving a successful dictionary consultation procedure by reaching the goals that motivated the search. The genuine purpose of a dictionary should therefore be to ensure successful dictionary consultation procedures. A successful dictionary consultation procedure depends on the way in which the needed linguistic information can be retrieved.

Bergenholtz (2003: 67) and Bergenholtz and Tarp (2003: 176) put a close link between the genuine purpose and the lexicographic function(s) of a given dictionary, when saying that the “genuine purpose is made up by the totality of functions of a given dictionary and the subject field(s) that it covers.” Therefore, the requirement for the identification of the real purpose of a dictionary compels lexicographers to negotiate the functions of the dictionary.

### 2.2.2.5 Lexicographic functions

Lexicographic functions constitute the most important concept of the functional theory of lexicography. Tarp (2000: 196) defines a lexicographic function as “the endeavour and ability of the dictionary to cover the complex of needs that arise in the user in a particular user situation.” Lexicographic functions of dictionaries are always closely related to the situations in which those dictionaries assist the users. Tarp (2000: 196-197) distinguishes two types of lexicographic functions that correlate with the first two types of user situations described in sub-section 2.2.2.2. These are knowledge-oriented functions and communication-oriented functions. On one side, the knowledge-oriented functions are (Bergenholtz & Tarp 2003: 176):

- to provide general cultural and encyclopaedic information
- to provide special information about the subject field or the discipline and
- to provide information about the language (e.g. when studying a foreign language)

On the other side, communication-oriented functions include providing assistance (Bergenholtz & Tarp 2003: 176):

- in text reception in the native language
- in text production in the native language
- in text reception in a foreign language
- in text production in a foreign language
- in translation of texts from the native language into a foreign language
- in translation of texts from a foreign language into the native language
- in correction of texts in the native language
- in correction of texts in a foreign language

In the final analysis, the identification of functions is a tribute to the knowledge that lexicographers have of the intended user needs and the social situations in which the latter occur (Gouws 2004: 266). The lexicographic function of a given dictionary and the needs of the target users influence the types of data presented in the dictionary to fulfil specific functions and the way in which data are represented. They also affect the type of dictionary (Gouws 2004: 266, 267, 273) and the data distribution or the access structure of the dictionary. Note that to avoid overcomplicating this chapter, a discussion about the types of dictionaries has not been attempted in this theoretical review. It should be noted, however, that with exception of few examples (e.g. Tarp 2011), most current dictionary typologies are made without taking into account the role of lexicographic functions (Gouws 2004: 273). When terms such as

monolingual, bilingual, LSP, LGP are used in this thesis, they are to be understood in their very general senses unless otherwise indicated.

## 2.3 The structural theory of lexicography

### 2.3.1 Theoretical foundations

Chabata and Nkomo (2010: 77) recount the development of the view of a dictionary as a structured textbook to the textual theory of lexicography firstly outlined by Wiegand as part of his general theory of lexicography. Wiegand (1996: 134) distinguishes four fields of dictionary research: the research on dictionary use, the critical dictionary research, the historical dictionary research and the systematic dictionary research. Within systematic dictionary research, Wiegand developed the general theory of lexicography consisting in partial theories including the theory of lexicographic texts. Wiegand considers a dictionary as a carrier of different types of texts. He claims that “dictionaries are text carriers, the texts of which are arranged into a macrotext which has a mediostructure and a metatext” (Wiegand 1996: 134).

The conception of a dictionary as a carrier of texts outlined at the very dawn of Wiegand’s theory of lexicography (Wiegand 1984) introduced a new era of theoretical lexicography and constituted a major breakthrough in establishing the academic independence of lexicography. In a good number of scientific contributions, Wiegand gradually developed and expanded this textual theory of lexicographic texts by adding a terminological system about dictionary textual concepts that is of invaluable importance in (meta-)lexicography. According to Smit (2002: 302-303) “Wiegand formulates certain hypotheses with regard to the lexicographical layout of different types of dictionaries by analysing many different dictionaries and dictionary articles. These deliberations would enable him to draw up a very detailed framework for the structure of dictionary articles, with the aim of making future dictionaries more user-friendly.”

As far as the structure of dictionary texts is concerned, Wiegand (1996: 134) identified *inter alia* the following structures: the macrostructure, access structure, frame structure, the textual word list structure, the microstructure, the article structure, the structure of the positions, the item structure, the addressing structure, the cohesion structure, the theme-rheme-structure, the mediostructure, and the microarchitecture. Some of these structures interrelate. Some of these structures are hierarchical and in inclusive relationships. For example, the emphasis on the role jointly played by outer-texts and a central list in the achievement of the dictionary functions, has led to viewing the frame structure as

more inclusive than the macrostructure. The frame structure of a dictionary contains at least a compulsory central list (both macrostructure and microstructure) and eventual outer-texts. However, of all the structures negotiated in the textual theory of lexicographic texts, the access structure plays a decisive role as it engages and is engaged in the rest of the dictionary structure to ensure that dictionary data presentation guarantees a quick access to the required data and a successful retrieval of needed information.

Drawing mainly on insights from Wiegand (1996), Steyn (2004), Gouws (2001, 2007a), Gouws and Prinsloo (2005) and Nielsen (2008), I describe in this sub-section, various dictionary structures and how they contribute to facilitating the consultation process and to achieving the genuine purpose of dictionary consultation by means of allowing a user-friendly presentation of dictionary data. I will not give a comprehensive account of all the structures negotiated in the dictionary. I will rather focus on the most important aspects for this study, i.e. the access structure, the frame structure, the macrostructure, the microstructure and the mediostructure, and the way they are negotiated in the dictionary and finally their importance with respect to the use and the functions of a dictionary.

### **2.3.2 Access structure and data presentation in a dictionary**

The study of the access route is an area of focus in theoretical lexicography (Bergenholtz, Nielsen & Tarp 2009: 26). The access structure establishes a search or access route a dictionary user follows to reach a specific lemma sign or data category in order to retrieve the needed information during the dictionary consultation procedure (Gouws 2001: 101, 102 and Gouws & Prinsloo 2005: 64). Steyn (2004: 277-278) shows that the dictionary consultation procedure involves two major phases which affect the access structure of the dictionary in two corresponding ways. In the first phase, the user strives to reach the relevant lemma sign or the guiding element in the main textual parts of the dictionary starting from the cover pages and moving to the macrostructure. In the second phase, once the right lemma sign or guiding element has been identified, the user engages in the microstructural search to identify the specific data from different entries contained in the article. This led to the identification of two types of access structures, namely the outer-access structure and the inner-access structure. The outer-access structure establishes a search route that leads the user to the desired guiding element in both central list and outer-texts while the inner-access structure relates to the access route followed by the user to reach the required data in the microstructural entries (Gouws & Prinsloo 2005: 64).

A third dimension of the access structure has been identified (Steyn 2004: 292) as the meta-access structure that deals with the match or mismatch between the language used in data conveying and the language proficiency level of the identified target user. This aspect of the access structure compels dictionary compilers to seriously consider the linguistic profile of the potential users of the intended dictionary. The immediate application areas of the meta-access structure are the comment on semantics, specifically the formulation and selection of definitions and illustrative sentences and various texts presented as outer-texts of the dictionary. Prinsloo and Gouws (2000) expose a number of criteria that can guide the lexicographers in identifying adequate example phrases when compiling or revising a dictionary. However, as users rely not only on their LGP and LSP competence when processing dictionary data, but also on the level of their encyclopaedic knowledge or cultural and specialized subject field knowledge, the meta-access structure seems to be incomplete when the lexicographers do not give a comprehensive account of the role cognition plays in retrieving the lexicographic information. The function theory of lexicography makes provision for the meta-access structure when emphasising the role of taking cognisance of the encyclopaedic and linguistic knowledge of the target users in order to calibrate the dictionary functions to their cognitive competence (cf. Tarp 2000: 194-195 and Nielsen 2008: 180).

The access structure has been recognised as one of the important structures of a dictionary (Nielsen 1995). The role of easy access to dictionary data with respect to dictionary functions and specific user needs is well understood when considering Tarp (2010: 460) who convincingly holds that:

(...) what distinguishes lexicographical works from other kinds of texts which users may also use to satisfy their information needs is that they are consultation tools focusing on punctual needs. This implies that the core of lexicography does not only consist of the principles of data selection, elaboration and presentation, based on a typology of needs related to types of users and social situations. In order to fulfil their function as consultation tools, lexicographical works must also cater for a quick and easy access to the relevant data, in printed dictionaries by means of lexicographical structures and references, and in electronic ones by means of search engines and links. Hence, the principles of quick and easy data accessibility, equally related to specific types of user needs, is therefore part and parcel of the core principles integrating the independent theory of lexicography.

In the function theory, ensuring an easy access or search route means that dictionary data have to be presented and distributed in such a way that a specific dictionary user with specific information needs in a specific social situation undertaking a look up (search) in a user-adapted and function-driven dictionary will successfully achieve the purpose of his/her consultation and reach the desired data to fit his/her information needs in a cost-effective time. This is of paramount importance because a

dictionary as a consultation instrument is not made for reading but for punctual consultation. In other words, the user is not expected to read the dictionary from A to Z but to look up an item that suits punctual information needs in a limited period.

Therefore, when assigning place to different structural components of a dictionary during the planning and compilation phases, compilers should make sure that both outer- and inner-access structures are used to safeguard an unimpeded accessibility of data included in the dictionary as well as a successful retrieval of the needed information from the given data. The view according to which lexicographic textual structures should not be used as an end but as a means in dictionary compilation, is convincingly expressed by Gouws and Prinsloo (2005: 64) when describing dictionary textual structures as functional tools or devices by means of which compilers can produce high quality and user-friendly dictionaries so that users can achieve a successful consultation procedure. Therefore, a mere utilisation of textual structures in dictionary making should not prevail over the user and functional considerations which are more important.

The following paragraphs address the relationship between the access structure and other structural aspects negotiated in different textual parts of the dictionary. Much attention is given to areas of application of the frame structure, the macrostructure, the microstructure and the mediostructure or cross-reference system.

### **2.3.3 Frame structure**

The frame structure of a dictionary contains various structural components of the dictionary (Nielsen 1995c). As far as the data distribution and access structure of the dictionary is concerned, Gouws (2003: 34) states that the texts in a dictionary can be accommodated in three major textual venues: the front matter, the central list and the back matter. Referring to Kammerer and Wiegand (1998), Gouws (2004a: 69) further defined the frame structure as the distribution structure of a dictionary displaying a central text and the outer-texts (front, middle and/or back matter texts).

#### **2.3.3.1 Central list and access structure**

In the conception of a dictionary as a carrier of different types of texts, the central list is the favourite area of dictionary consultation and hence the part of a dictionary which plays a significant role in fulfilling the function of the dictionary in question (Gouws 2003a: 93). The central list contains the list of article stretches, often alphabetically ordered, each one containing different types of articles (Gouws & Prinsloo 2005: 62). A typical article contains one lemma sign which functions as the guiding element

and a microstructure containing various entries containing items constituting the comment on form and comment on semantics and structural indicators separating and identifying different search fields (Gouws 2003: 35). The list of ordered lemma signs contained in the central list forms the macrostructure of the dictionary. Of the texts forming the frame structure, the central list is usually the target of dictionary consultation. Therefore, dictionary compilers should make sure they optimise more than elsewhere the advantages of the access structure in the central list. As noted earlier, the access to a specific article stretch and lemma sign inside the article stretch is facilitated by the outer access structure (Gouws 2001: 103). From the identified article, the inner access structure guides the user through the internal search route to the desired data. In this research, the potential of the central list as venue for phrasemes will be investigated. Very specifically, it will be employed in the formulation of proposals for representation of idioms, collocations and pragmatemes.

### **2.3.3.2 Outer texts and the access structure**

- **Generalities on outer texts**

The foregoing sub-section has described the central list as the main and compulsory section of a dictionary (cf. Hausmann & Wiegand 1989) where the role and manifestation of the access structure within both macrostructure and microstructure become more apparent. Nevertheless, the role and extent of the access structure does not end there. The word list and linguistic bias that have dominated traditional lexicography have resulted in a narrow approach minimizing the use of outer text sections and the status of dictionaries as carriers of texts (Gouws 2004a: 67). However, modern metalexicography has gone a step further by recognising the role played by outer texts with respect to data distribution in the dictionary (Bergenholtz & Tarp 1995; Chabata & Nkomo 2010; Gouws 2001, 2003a, 2004a, 2007a; Gouws & Prinsloo 2005; Nkomo 2008; Steyn 2004). The outer texts may be either front matter texts when they appear in the front section or back matter texts if they occur in the back section of the dictionary. The front matter is the section of outer texts found in all dictionaries. Although modern-day lexicography has advocated an extensive and functional use of outer texts, it is disappointing to note with Gouws (2004a: 75, 2007a: 79) that these theoretical developments regarding outer texts have not yet been properly reproduced in practical lexicography. In the process of dictionary making, outer texts are still randomly selected without due regard to the genuine purpose and functions of the dictionary. Gouws (2007a: 79) goes on deplored the arbitrary or traditional basis guiding selection and distribution of data in outer texts and makes a plea for a function-led distribution programme for a better use of outer texts in future dictionaries.

- **A transtextual functional approach to outer texts**

In the theory of a dictionary as a carrier of text types, a typology of outer texts is formulated based on their integration in the genuine purpose of the dictionary (cf. Gouws 2009a: 69 referring to Bergenholz, Tarp & Wiegand 1999:1778). This criterion has helped distinguishing between integrated and non-integrated outer texts (cf. Gouws 2001: 104-105; 2004a: 72; Gouws & Prinsloo 2005: 59). Regarding integrated outer texts Gouws (2002) states:

[Integrated outer texts] belong either to the class of so-called "help texts" with the function to assist the user with a more successful dictionary consultation procedure or, in as far as they contain a part of the lexicographic data presented in the dictionary, they are integrated texts from which the user can retrieve information regarding the subject matter of the dictionary. Both these types of outer texts stand in a relation with the genuine purpose of the dictionary. They either serve (as help texts) to explain the use of the dictionary to the user or they are, as a result of integration into the dictionary internal data distribution, directly integrated into the genuine purpose of the dictionary (cf. Gouws, 2004a: 73).

Reiterating the argument earlier set forth by Bergenholz and Tarp (2000: 198) according to which the functions constitute the driving principle of lexicography and no data should be included in the dictionary if its inclusion cannot be justified on the basis of the function it is called to fulfil (Bergenholz & Tarp 2003: 177; Tarp 2000: 198), Gouws (2007a, 2009a), in latter publications, pressed on with theoretical developments regarding an integrated model of outer texts, taking into consideration their participation in the fulfilment of lexicographic functions. In this trans-textual approach to outer texts Gouws (2007a: 81-83, 2009a: 70) proposes a distinction between function-adhering outer texts and non-function-adhering outer texts. Within outer texts, a function-adhering outer text is an outer text "by means of which a given lexicographic function can be achieved" while a non-function-adhering outer text does not serve to achieve any lexicographic function. A text containing a list of staff and contributors or a bibliography provided it not being referred to in the central list or in any other outer text, is a good example of a non-function-adhering outer text. Taking into account the role of outer texts in fulfilling the functions of the dictionary led Gouws (2007a: 81-83, 2009a: 70) to extending the notion of integrated outer text to dictionary functions and to differentiating integrated function-adhering outer texts from the non-integrated function-adhering outer texts. An integrated function-adhering outer text is integrated into the lexicographic function prevailing in the central list (e.g. user guide, grammatical data or central list data phased out to the outer texts) while a non-integrated function-adhering outer text does not, although it fulfils a secondary lexicographic function which is different from the lexicographic function catered for in the central list (Gouws 2007a: 82- 83). An integrated function-adhering outer text may or may not be integrated in the genuine purpose of the

dictionary (Gouws 2009a: 70).

However, in the light of the conception according to which the “genuine purpose is made up by the totality of functions of a given dictionary and the subject field(s) that it covers” (Bergenholtz & Tarp 2003: 176; Bergenholtz & Gouws 2007: 242), it will be difficult to imagine a function-adhering outer text which is not integrated in the genuine purpose of the dictionary. Following this line of thought, the scope of the genuine purpose goes beyond the central list as some functions of the dictionary can be realised in the outer texts, and the notion of integration and non-integration should be exclusively used with respect to primary versus secondary functions, not in relation to the genuine purpose of the dictionary.

The significance of the distinction between integrated function-adhering outer texts and non-integrated function-adhering outer texts is that it should guide lexicographers’ decisions regarding functions allocation across the textual structure of a specific dictionary, focusing on the use of outer texts in coordination with the central list to highlight the overall function(s) of the dictionary and to ensure the functional coherence of the dictionary. In the same way, outer texts should also be used to cater for secondary functions deemed necessary to satisfy the needs of a certain group of users, albeit they are not integrated in the function having to prevail in the central list (Gouws 2007a: 82- 83, 2009a: 70). For example, a bilingual dictionary having communication as a primary function could be made poly-functional by including in the back matter some cognitive-oriented texts on culture-dependent aspects which may be conveyed differently in respective languages from culture to culture or from country to country (cf. Bergenholtz & Tarp 2003: 177). Gouws (Personal communication, 27/06/2013) mentions another situation wherein outer texts could be used in a similar way. For example, the genuine purpose of a given monolingual LGP dictionary could be to provide the target user with a linguistic description of the given language. In doing so the dictionary may have a text reception and text production function. The default article does not contain etymological data. Etymological data regarding certain lexical items are given in a back matter text. This back matter text is an integrated outer text because it is integrated into the genuine purpose of the dictionary, i.e. to provide linguistic data. This back matter text has a cognitive function and, albeit integrated into the genuine purpose of the dictionary, it does not adhere to the function of the dictionary. It is therefore an integrated non-function adhering outer text.

If we have to adopt the theoretical developments outlined above regarding outer texts (Gouws 2007a), it is not quite clear whether the distinction between integrated and non-integrated outer texts could be

applied to non-function-adhering outer texts as is the case in Nkomo (2008: 169). Apparently, non-function-adhering outer texts cannot qualify in any way to the status of integrated outer texts (in the sense of integration into the genuine purpose), as they fulfil no function pertaining to the genuine purpose of the dictionary; in all respects, their function is not lexicographic. In addition, they cannot be integrated into the function realised by the central list, merely because they are not function-adhering. Thus, the description of “integrated non-function-adhering outer texts” as related to “the genuine purpose or lexicographic function of a dictionary as aids to optimum retrieval of relevant information” (Nkomo 2008: 169) seems quite problematic in that it is not clear how an outer text that does not realise a lexicographic function, be it a secondary one, could be integrated into the genuine purpose of the dictionary. On the other hand, help texts such as user guides are fully-fledged integrated function-adhering outer texts.

Outer texts play a decisive role in the textual structure of the dictionary as they present valuable opportunities for data distribution in a functionally coherent way. Some front matter texts (e.g. user guide) are used by the lexicographer to enhance both outer and inner access structures of the dictionary, provided the user consults them prior to dictionary consultation. However, this is not always the case as many users never read these introductory texts. In this thesis, the potentials of outer texts for proverbs and pragmatemes recording will be assessed.

### **2.3.4 The Macrostructure**

The term macrostructure refers to the ordered set of all items to be included in the dictionary as lemma signs (Gouws & Prinsloo 2005: 63). These items can either be first level treatment units or second level treatment units presented in the central list, but can also be presented in the outer texts (back and middle matter sections). A quick accessibility to the specific lemma sign within the macrostructure in a specific dictionary is determined by the type of article arrangement in the dictionary and the utilisation by the compilers of a rapid outer access structure. By arrangement, reference is made to either an alphabetical or a thematic arrangement of macrostructural elements and seldom to the casual arrangement principle (Nielsen 1994: 180) or the combination of both conceptual and alphabetical principles (analogical order, cf. Mpofu 2007). The following discussion is confined to alphabetical and thematic or conceptual macrostructures due to their relevance for this work.

#### **2.3.4.1 Alphabetical macrostructure**

Nielsen (1994, 1995a) discusses the alphabetical arrangement principle as one of the important aspects of dictionary structures. The alphabetical ordering is the common and most user-friendly approach used

in lexicography in ordering lemma signs in the macrostructure of the dictionary (Nielsen 1994: 182). Its popularity comes from the fact that the order of letters can easily be mastered by users. In this context, it is appropriate to quote Nielsen (1994: 181) who argues that the alphabetical arrangement principle “has the advantage that users knowing the alphabet of the source language will also know the order in which the lemmatised lexical units are placed.” He further claims that the fact that both conceptual and random arrangements cannot function optimally without recourse to additional alphabetical lists, supports the argument for this lemmatisation principle (Nielsen 1994: 182). In addition, the alphabetical principle is practical and timesaving for compilers who, with the aid of computers, can easily and very quickly compile a long list of words and sort it alphabetically. In addition, most, if not all, corpora software has alphabetical word sorting functionality.

However, the alphabetical listing is criticised for its weakness where the general alphabet of a language differs slightly from the access alphabet of a dictionary because of some “alphabet-external symbols”, such as empty spaces between multilexical items, hyphens, numerical signs or supra-segmental features which are inescapable in lemma listing (Nkomo 2008). In this situation, lexicographers have to expand the language alphabet by means of including those alphabetical features that are not provided in the general alphabet of the language. The need for a lexicographic access alphabet is articulated in Nielsen (1994) and Nkomo (2008: 143-146) among others. When these symbols and others are used in the access alphabet of a dictionary, lexicographers are duty bound to clearly inform the users of their implication on the outer-access structure of the dictionary in question.

Two types of dictionary articles have been identified, i.e. simplex versus complex articles in LGP dictionaries and simplex versus synopsis articles in LSP lexicography (Gouws 2003a: 93, 2003: 36). In dictionaries where articles display a complex structure, the lemmatisation of niched and nested lemmata (cf. Gouws 2001, 2002, 2003, 2007) constitutes one of the delicate aspects of lemmatisation with respect to maintaining a consistent alphabetical arrangement of lemmata. According to Gouws (2003: 40), in addition to the standard vertical alphabetical ordering, a second type of vertical alphabetical ordering is complemented by a horizontal ordering resulting from the inclusion of sublemmata attached to the main lemmata constituting the guiding element of a complex article containing clustered sublemmata. A lemma cluster may keep up the internal alphabetical ordering with respect to both vertical and horizontal ordering (niched lemmata). However, it may also deviate from the overall alphabetical ordering by including nested sublemmata. Due to the user-unfriendliness of both niching and nesting and with respect to target users, dictionary compilers should adopt a well-

balanced view on these text condensation strategies so as not to deter access to dictionary data. More particularly, nesting should be seldom, if ever used, especially if the dictionary targets an audience of lesser skilled users. In that context, a more simplified lemmatisation is preferable.

The outer access structure also concerns the lemmatisation of phrasemes and other multiword items. To tell the truth, theoretical lexicography is not yet quite advanced in this area although many insightful ideas have been given. The lemmatisation of lexical phrasemes such as idioms and other types of phrasemes (pragmatemes, collocations and proverbs) also poses many problems with respect to the role of the macrostructure in data distribution and accessibility. This thesis focuses on these kinds of items. It has been particularly noted that within both the functional theory and the general theory of lexicography and even in linguistic approaches to lexicography views diverge not only as to how those items should be recorded in the dictionary but also concerning their lexicographic status and the lexicographic validity of linguistically inspired classifications of phraseological items. Drawing on numerous theoretical contributions and practical applications, a model for lemmatisation of phrasemes in Kinyarwanda dictionaries has been sketched with respect to various lexicographic functions.

The use of a rapid outer access structure in connection with the alphabetical arrangement principle consists in using structural indicators, for example, thumb indexes and bold face capitals which facilitate a quick separation of different sections of the macrostructure and a quick access to the article stretch where the required lemma sign is likely to be accommodated. Another strategy is the inclusion of running heads at the top of each page of the dictionary, and the use of coloured pages or coloured edges of pages to separate the main wordlist sections in a bilingual dictionary or the inserted pictorial or study pages. Other important dimensions of a user-friendly and function-driven dictionary that will not be discussed in detail here are employing a poly-accessible macrostructure and the rapid outer access structure. A poly-accessible macrostructure implies the use of different access roads in one dictionary by including additional word lists or one or more indexes complementing the central list.

#### **2.3.4.2 Thematic macrostructure**

The motivation behind the use of a thematic macrostructure or systematic macrostructure (Nielsen 1995d) is the fact that it allows a systematic arrangement of lemmata in conceptually significant classes within a specific given subject discipline. As referred to by Nkomo (2008: 147), Nielsen (1995d: 195) articulates that lemmata in a systematic or thematic macrostructure are arranged “according to conceptual systems based on definitions, which among other things, serve to delimit the concepts relative to each other as well as to establish mutual relationships.” Thus a thematic macrostructure is

seen as an alternative to alphabetical ordering in LSP lexicography as by this way lexicographers ensure the representation of conceptual or cognitive relations within the lexical system of a given special subject field. Although dictionaries displaying the overall alphabetical arrangement at macrostructural level utilise cross-references (e.g. synonyms, antonyms, hyperonyms and hyponyms) to link conceptually related lemmata or senses of lemmata in the dictionary, an instance of thematic arrangement at microstructural level (Louw 2000: 121), the cross-referencing system is always selective and it generally disregards the systematic representation of concepts.

However, leaving aside its advantages, Nkomo (2008) points out that the thematic arrangement principle presents some non-trivial challenges that limit its utilisation in LSP lexicography which should be seriously considered by lexicographers when lemmatisation choices are to be made. One of these weaknesses is the fact that a dictionary presenting a thematic macrostructure requires users to have high encyclopaedic knowledge of the treated subject field to utilise it to the full. This compels lexicographers to make provision for an alphabetical access path for less experienced users by including an alphabetical page-indicating index of the macrostructure in the outer-texts in order to make the dictionary poly-accessible (Nkomo 2008: 148-149). Moreover, before concluding that the whole LSP lexicography cannot adopt the thematic macrostructure as lemmatisation procedure, Nkomo refers to Nielsen who finds that some subject areas, especially in social sciences where concepts are inextricably inter-linked and often overlapping, may rebel against a systematic classification of concepts. In addition to these difficulties, Nielsen (1995: 198) draws our attention to the fact that “systematic processing of comprehensive subject areas and their terminology is both a time-consuming and resource-demanding process” (as quoted by Nkomo 2008: 149). This is problematic in the sense that the fast evolution of knowledge and practices characterising most of the subject areas compels LSP lexicographers to complete their projects within a reasonable stretch of time in order to produce reference tools that reflect the state of the art knowledge in the fields in question. This thesis being directed at the general language (phraseological data) which privileges the alphabetical ordering, the thematic macrostructure is not its core approach. However, I will explore the potential of the thematic arrangement principle with respect to the lemmatisation of pragmatemes and proverbs in the back matter. It should also be noted that the onomasiological treatment of collocations under the keywords representing their bases is a sort of thematic procedure.

### 2.3.5 The microstructure

The foregoing discussion has made it clear that the macrostructure of a dictionary affects the overall access structure and reflects the complex nature of lexicographic texts. However, the complexity is also obvious in the structure of dictionary articles or the microstructure (Gouws 2003a: 93). The term microstructure refers to an ordering structure providing a hierarchical arrangement of items or data given as part of the treatment of the lemma sign within a dictionary article (Gouws & Prinsloo 2005: 64). Most data presented in the dictionary are accommodated in the microstructure that together with the lemma sign constitutes a dictionary article. I have earlier discussed problems posed by article clusters with respect to lemmatisation of their guiding elements. In this sub-section, emphasis is on other aspects of the article structure (microstructure) and their implication for the easy access to data and information retrieval.

When the right lemma sign has been identified and reached, the user has to navigate across the rest of the article or microstructure by means of the inner-access route. When conceptualising the structure of the dictionary, the lexicographer ought to make provision for a function-driven inner-access structure. In so doing, they have to ensure an adequate presentation of the main microstructural entries, i.e. the comment on form and the comment on semantics and an appropriate use of different structural indicators in order to avoid data congestion and to guarantee an unimpeded retrieval of information (cf. Gouws 2003, 2003a; Nkomo 2008). The implementation of the inner access structure relies on microstructural dimensions by means of which a rapid access to data encoded in the microstructure and a cost-effective information retrieval are achieved. Many publications have discussed these aspects *inter alia* Gouws (2003a), Gouws and Prinsloo (2005), Nielsen (1995b), Steyn (2004) and particularly many publications by Wiegand (cf. Welker 2009).

#### 2.3.5.1 Microstructure and data distribution structure

The data distribution structure refers to the strategic distribution or placement of data across different places in the dictionary so that all the parts of the dictionary contribute actively to the achievement of the genuine purpose of the dictionary (Steyn 2004: 288). As far as the microstructure is concerned, Wiegand's theory on dictionary texts proposes no less than eight types of microstructure (Welker 2009: 13) from which Gouws (2003a: 95) and Gouws and Prinsloo (2005) retain two significant dichotomies, namely the distinction between the obligatory microstructure and the extended obligatory microstructures on the one hand, and an integrated versus non-integrated microstructure on the other

hand (cf. Nkomo 2008: 161-162).

A compulsory microstructure is the default structure found in articles displaying the minimum number of entries (Nkomo 2008: 161) while, in addition to the types of entries accommodated in the default microstructure, an extended obligatory microstructure makes provision for additional data categories that may be necessary for the comprehensive lexicographic treatment of certain types of lexical items included as lemmata (Gouws 2003a: 95). Gouws and Prinsloo (2005: 141) insist that the “data distribution structure should make provision for additional items and data categories that might be extremely important in the treatment of certain lexical items.” In agreement with this recommendation, the following conclusion drawn by Nkomo (2008: 162) is very pertinent:

This means that a complete microstructural programme should have a slot for every item or data category that would be included in the dictionary. The filling in of the slots would then depend on whether a particular lexical item requires particular lexicographic treatment, which would be realised by the inclusion of a particular data category.

A prior design of a comprehensive data distribution programme helps the lexicographers because the possible number and types of entries and search areas to be incorporated in the dictionary being compiled is known. In an extended obligatory microstructure for example, lexicographers should make provision for the inclusion of inner text blocks or text boxes in order to accommodate e.g. usage notes, encyclopaedic, historical and cultural data which need additional emphasis or which could not be presented in the default compulsory article structure (Gouws 2001: 108-109; Gouws & Prinsloo 2010: 509-510). If used judiciously and only where necessary, this lexicographic device presents itself as an added value to the dictionary by ensuring a vivid presentation of those types of data. However, Gouws and Prinsloo (2010: 509) caution against an overexposure of text boxes or their use beyond the extended microstructure because they may lose value and hence their *raison d'être*. In addition, the scholars insist that “the positioning of a text box must be done in accordance with the type of microstructure prevailing in the dictionary and with cognisance to the scope of the text box in its addressing relation to the rest of the article” (Gouws & Prinsloo 2010: 502).

### **2.3.5.2 Microstructure and data architecture**

The presentation of microstructural entries in separate slots or search areas separated by white spaces, displays what is known as micro-architecture. Among other things, the micro-architecture makes provision for the distinction between the two major components of an article, i.e. the comment on form and the comment on semantics. In this way, the use of articles with a micro-architecture enhances the access to the data and information transfer. This makes the following recommendation drawn by Steyn

(2004: 289) very important:

The different microstructural categories should complement each other without increasing the degree of textual condensation in the article. It is therefore of great importance that the lexicographer already decides on the design and well-structured micro-architecture of the article during the planning stages in order to assist the user when he/she is negotiating the article internal search route. The lexicographer can use a lay-out sample with specimen articles to make sure that the micro-architecture leads to user-friendliness and readability.

Data architecture is another aspect of the inner-access structure. As indicated above, the use of a well-developed micro-architecture through the judicious use of white spaces contributes to indicate the boundaries of different items' search zones. However, the use of typographic and non-typographic structural indicators in the article-internal structure makes the dictionary consultation procedure a yielding one (Gouws 2003a: 94). According to Nkomo (2008: 165),

The former [typographic structural indicators] relate to different font types and styles, small caps, capital letters or even font colour by means of which specific items, data categories or search fields in the article may be identified and distinguished from each other. The latter [non-typographic structural indicators] are brief headings, symbols and signs such as punctuation marks, abbreviations, dots, circles and other shapes which are used to mark a specific field or data category in a specific dictionary. Both typographic and non-typographic structural indicators play an important role in the inner access structure of a dictionary.

As Gouws (2003: 35) states, "The genuine purpose of a structural indicator as a functional entry in a dictionary article is to assist the user in identifying and distinguishing the different items and in finding them as quickly as possible". He continues saying that structural indicators indicate the hierarchical relations holding between different items composing an article. Admittedly, a sophisticated and well-balanced use of structural indicators to make search zones and data categories stand out, contributes notably to an unimpeded access to the data presented in the dictionary (cf. Gouws 2001: 109; Steyn 2004: 290). Worthy of remark here is the recent suggestion by Gouws (2010) in favour of extending the use of structural indicators beyond the microstructural level and bringing some macrostructural items within the scope of structural indicators, where for instance a structural indicator could precede and identify text blocks containing vertically-ordered second level treatment units presented as grouped sublemmata attached to default articles, cf. Gouws (2010: 56). In this way, structural indicators contribute to the outer access structure of the dictionary.

In addition, an optimal retrieval of information is achieved if the order or position of entries within the microstructure reflects the functional priority of the dictionary, since entries in the first positions receive much more focus during dictionary consultation whilst those in last entries yield far less attention (cf. Steyn 2004: 290 referring to Gouws 1993: 37).

A well-developed data distribution programme and a well-spaced out micro-architecture make dictionary articles more user-friendly and enhance data access and information retrieval (cf. Gouws 2001: 109). In the same line of thought, the importance of a well-structured and standardised microstructure in reducing information costs becomes vividly evident when considering Nielsen's argument that (2008: 177):

Users place easy and quick access to data items in articles high on their list of priorities. By using standardized article structures in dictionaries, lexicographers make it easy for users to familiarize themselves with the layout of articles. They may divide the text of an article into fields, or zones, that each contains a specific type of data. Users can easily navigate their way through such standardized articles because it involves less effort or incurs a lower cost to find what they are looking for, or to realise that what they are looking for is not in the article.

### **2.3.5.3 Types of microstructures and addressing structures**

Hausmann and Wiegand (1989) have yet again introduced another distinction, i.e. between integrated, semi-integrated and non-integrated microstructures. Gouws and Prinsloo describe both the integrated and non-integrated microstructures and the way they should be put to good use to achieve a user-friendly presentation of microstructural items (Gouws 2003a, 2006; Gouws & Prinsloo 2005). This distinction refers to the closeness and directness of the relation between each definition or equivalent in the dictionary and the co-text elements provided as supporting data, e.g. illustrative sentences (Gouws 2003a: 95). Nkomo (2008: 162) explains, "An integrated microstructure displays a clear relationship between definitions or equivalents and their supporting co-text entries while a non-integrated microstructure does not." In a dictionary displaying an integrated microstructure, the translation equivalent or paraphrase of meaning appear in the same sub-comment on semantics with their respective co-texts. In a non-integrated microstructure, all sub-comments, i.e. equivalents or paraphrases are grouped in one text block followed by their co-texts in another separate text block. The ways in which they are grouped in the separate texts block helps an informed user to link a specific sub-comment to a corresponding co-text (Gouws 2006: 32).

A semi-integrated microstructure combines the properties of integrated and non-integrated microstructures. As in a non-integrated microstructure, entries in a semi-integrated microstructure are distributed in two main sections. The significant difference between a non-integrated and a semi-integrated microstructure is that each sub-comment in the first section is accompanied by a single co-text entry whilst additional co-text entries are supplemented in the second section. The relation between

each co-text and the relevant paraphrase or equivalent is clearly indicated. This structuring of the microstructure makes provision for two different functions in one article. The integrated co-texts assist users in both reception and production of texts while the non-integrated co-texts add additional data meant for the production function (Gouws 2006: 33). A semi-integrated microstructure presents substantial lexicographic potential for the presentation of notes in text boxes or culture-bound expressions within the scope of the microstructure. The use of data other than co-text entries would be a new way of using a semi-integrated microstructure.

The distinction between an integrated and a non-integrated microstructure reminds of the concept of addressing defined by Hausmann and Wiegand quoted by Welker (2009: 13) as “The way in which a form and information relating to that form are brought together is the addressing procedure [...]. Each information item [...] is addressed to a form called address [...].” (Hausmann & Wiegand 1989: 328) Therefore, the basis for distinction between integrated and non-integrated microstructures is the addressing structure. Gouws and Prinsloo (2005: 64) define this structure as “The relation between an entry and the treatment unit at which it is directed”, before explaining the three manifestations of the addressing structure, namely the lemmatic, the sublemmatic and the non-lemmatic addressing. With non-lemmatic addressing the address is neither the lemma nor the sub-lemma, but another type of microstructural element. Thus in correlation with integrated and non-integrated microstructures, a dictionary may display an article-internal direct or remote addressing structure respectively. However, it is of vital importance to avoid the occurrence of remote addressing as this has a negative impact on the retrieval of the needed information.

### 2.3.4 The mediostructure

In this sub-section, I will not provide a detailed theoretical discussion of the concept of mediostructure and its various applications in lexicography. Given here are only a brief description of its importance in data distribution and transfer and some typical occurrences of the mediostructure. In the foregoing discussion a dictionary has been described as a set of partial texts functioning as a whole in an interactive way to form one textbook or a frame structure (Kammerer & Wiegand 1998; Wiegand 1996.). Data distribution structure and data access structure address issues related to where and how to place data in different parts of a dictionary. Nevertheless, the lexicographers should give a complete account of different relationships prevailing among different types of data presented in a dictionary. It is with this background that modern-day lexicography has recognised the mediostructure or cross-referencing system as a critical structure of the dictionary (Nielsen 1995, 1999).

The mediostructure is commonly defined as a “system of cross-referencing which leads a user from a reference position to a reference address” (Gouws & Prinsloo 2005: 64) or “a network structure that consists of the set of relations existing between data by way of cross-references” (Nielsen 2008: 183).

Two types of cross-reference are identified, namely the explicit cross-reference and the implicit cross-reference (Gouws & Prinsloo 2005: 191). The main difference between explicit and implicit cross-reference entries lies in the fact that in the former the lexicographers compel the user to go from a reference position to the reference address by means of cross-reference markers, while in the latter the user is motivated by the need for extra information in addition to that which prompted the dictionary consultation process, without being advised to do so by the lexicographer (Nkomo 2008: 175).

Nkomo (2008: 174) argues that among other functions, the mediostructure facilitates data distribution while saving space and enhances access to data during dictionary consultation procedures, hence serving both lexicographers and users. Alberts (2007) argues that cross-referencing can be used to establish relations among different components of a dictionary and to create textual cohesion or interconnect the knowledge elements represented in different sectors of the dictionary on several levels of lexicographic description to form a network, by means of specific typographic and non-typographic structural markers. In addition, mediostructure as a guiding structure assists users accessing different data contained in a dictionary in order to achieve successful dictionary consultation (Gouws & Prinsloo 2005: 192), taking a user from an entry to another within the same article (internal reference address), from an article to another within the same central list, from the central list to the outer text or vice versa or from a reference position in the dictionary to an external source text used during the compilation (external reference address) (cf. Gouws 1999: 6; Gouws & Prinsloo 2005: 180 and Nkomo 2008: 177).

However, the unsuitable use of cross-references could have an adverse effect on the access structure and deter users during the consultation process, for example when they are referred to non-existing, wrong or unnecessary addresses. The same holds true when the lexicographer employs too many cross-reference markers so that the readability of articles become impeded by the resultant high textual condensation (Gouws & Prinsloo 2005: 192; Nkomo 2008: 177). Therefore, it can be argued that cross-reference is a helpful lexicographic device as long as the lexicographer uses it appropriately and moderately, and utilises the front matter wisely to acquaint the users with various cross-referencing procedures employed in the dictionary.

## 2.4 Conclusion

In sections 1.2 and 1.3 of this thesis I have indicated that the compilation of Kinyarwanda dictionaries is at all levels characterised by a lack of guidelines stemming from theoretical lexicography. This chapter has outlined some of the most important concepts of the functional theory of lexicography and the general theory of lexicography, and mapped the lexicographic functions onto the structural or textual theory of lexicography. Transformative in nature, the theory of lexicographic functions strictly envisages the lexicographic research and dictionary compilation from the viewpoint of the user. It postulates that everything in the dictionary is determined or reshaped by the situations in which user needs occur, the situations which in turn, determine the functions of the dictionary. Among concepts used in user research, this chapter has discussed the lexicographic functions, the user types and user situations, the genuine purpose of the dictionary and the access structure as well as other textual structures of the dictionary.

Finally, I have indicated how I intend to utilise various concepts and insights drawn from this theoretical framework. Phrasemes are relegated to a position of secondary importance in Kinyarwanda lexicography due to the challenges facing dictionary compilers in presenting phrasemes in dictionaries. The concepts presented in this chapter will be used in this thesis to formulate transformational models that could improve and make the lexicographic representation of phrasemes in future dictionaries of Kinyarwanda user-friendly. Although this research is not a typical research on dictionary users, it puts the latter at the centre of the debate where phraseology is envisaged from the perspective of the user, not from that of the “linguist” as has unfortunately been the case in all Kinyarwanda dictionaries. In this way, the concepts of dictionary function and access structure are the driving forces of this thesis. They are used in formulating relevant models for presentation and treatment of different types of phrasemes as illustrated in Chapters 4, 5, 6. However, besides the overall theoretical framework mapped out in this chapter, reference to scholars working in different theoretical perspectives of lexicography and linguistics will occasionally be made in relevant chapters.

## **CHAPTER 3: PHRASEMES IN KINYARWANDA LINGUISTICS AND LEXICOGRAPHY: THE EXTENT OF THE PROBLEM**

### **3.1 Introduction**

Some of the problems facing Kinyarwanda lexicographers are the lack of a clear definition of what a phraseme is and the lack of cognisance of the wide-ranging scope of the phraseological subset of Kinyarwanda lexicographic data. The notion of a phraseme is not a well-defined concept in Kinyarwanda lexicography. Neither is it in Kinyarwanda linguistics. In the first chapter it has been indicated that different types of phrasemes are being confused while others may be ignored. This misconception is mirrored in Kinyarwanda dictionaries and results in the mistreatment of the expressions in question. The issue is made worse by the lack of theoretical lexicography guidelines in the formulation of inclusion policies of phraseological data in Kinyarwanda dictionaries. Through a review of studies previously carried out on the phenomenon of Kinyarwanda phrasemes, this chapter will assess the problems and gaps of knowledge in the current treatment of phrasemes, using insights from theoretical lexicography. The main assumption is that a critical analysis of the treatment of phrasemes in major Kinyarwanda dictionaries makes it obvious that phrasemes are inappropriately treated. It is only after having weighed the gap of knowledge in this area that the problematic nature of the present research will be defined.

The present chapter is divided into four main sections. After the introduction (section 3.1), a working definition of the concept of phraseme using the most influential linguistic approach to phraseology developed in the framework of the Meaning-Text Theory, is presented in section 3.2. In section 3.3a critical review of linguistic and lexicographic literature pertaining to Kinyarwanda phrasemes is presented and the major ideas presented in the chapter with regard to its objective are recapped in section 3.4.

### **3.2 The concept of phraseme in the Meaning-Text Theory**

#### **3.2. 1 The Meaning-Text Theory**

The Meaning-text theory (MTT) is a linguistictheoretical framework, first put forward by Žolkovskij and Mel'čuk in Moscow in the 1960s and early 1970s and was initially meant for the construction of models for natural language processing (Žolkovskij & Mel'čuk 1965, 1966, 1967). However, over the last four decades, the theory has been extensively developed and refined by linguistic scholars in Russia, Canada and Europe, with respect to its coverage of linguistic phenomena and application in a

variety of research fields, and as an alternative to the Chomskian Generative and Transformational Theory. It is a holistic linguistic theory characterised in particular by the central position of the lexicon, the primacy of semantics, the importance of the communicative structure, and the reliance on dependencies at all levels (Mel'čuk 1988, 1997, 2012; Mel'čuk, Clas & Polguère 1995). Its main objective is to formulate a Meaning-Text model that can be applicable to natural language processing. It is in keeping with formal grammars and presents itself as a model structured in a number of representation levels: semantic, syntactic, morphological and phonological (Jousse 2003: 11). Regarding lexicology, the study of phrasemes is part of the lexicological framework of the MTT, i.e. the Explanatory and Combinatorial Lexicology (ECL).

### **3.2.2 Definition of the concept of phraseme**

In the framework of ECL, a phraseme is defined as “a phrase which is not free” (Mel'čuk 1998:27), i.e. a phrase which is not unrestrictedly constructed. In this study, the term *phraseme* as defined by Mel'čuk is used to globally refer to different kinds of phraseological word combinations in which at least one of the lexical constituents is restrictedly selected (Mel'čuk 2010).

To understand the importance of this definition, there is a need to be reminded of the role of the production perspective in MTT. The name of this theory, Meaning-Text Theory, exemplifies well the beginning and the outcome of the linguistic model, i.e. the Meaning-Text Model (MTM), and the orientation of linguistic research, from the meaning to the text (text production or synthesis). The Meaning-Text Theory is production-oriented and envisages the study of linguistic expressions from this perspective. In this way, any linguistic analysis in the MTT has its start in the Semantic Representation (= SemR) and results in text, the Phonetic Representation (=PhonR). In line with the general orientation of his theory (Mel'čuk, Clas & Polguère 1995), Mel'čuk believes that any discussion about phrasemes makes sense if the latter are considered from the writer/speaker's point of view, that is, in terms of their production/construction, rather than in terms of their interpretation or reception by the hearer/reader (Mel'čuk 1998:24), because the activity of producing a text by the speaker is more linguistic than is the comprehension, by the hearer/reader, of a sense from a given text (Mel'čuk 2003: 20). This is linguistically pertinent although in lexicography, the lexicon can be envisaged from both production and reception perspectives.

However, due to the fact that text reception relies on linguistic and extralinguistic skills and knowledge, it is not likely to provide the theoretical guidelines for a linguistic macro-classification of phrasemes; rather, they can be invoked in sub-classification of specific classes of phrasemes as will be shown in

the relevant chapters. The comprehension of a text does not presuppose the knowledge of all steps that have led to the production of the text. It requires only the understanding of the used expressions, the grammatical rules involved in the text and to some extent the encyclopaedic knowledge of the subject field discussed in the text. In our case, the text reception orientation cannot help distinguishing free phrases from phrasemes or reveal why compositional phrasemes such as collocations are still unpredictable. Therefore, a text reception perspective will later be invoked in the sub-classification of idioms, proverbs and collocations only at internal level, to determine those that necessitate a compulsory treatment in text reception dictionary and those that do not.

According to Mel'čuk, (1998:24-25) three major representations are involved in the production of a linguistic message. First, the speaker starts with the Conceptual Representation [ConceptR] of the situation he wants to express. The ConceptR is a mental reflection of perceived reality, of the speaker's encyclopaedic knowledge relevant to the situation in question, of his intention, preferences, wishes and goals, and of his ideas about the addressee.

The following step is the construction of the Semantic Representation [SemR] of his intended utterance, based on the initial ConceptR. He/she does it in accordance with the Concepts-Meaning Model of his/her language L [CMM (L)], which maps elements and configurations of the ConceptR onto elements and configurations of the corresponding SemR. The SemR specifies the common meaning or semantic invariant of a set of synonymous or paraphrasal utterances (Mel'čuk: In press). At this level, he/she has the meaning or the signified.

The further step, from a given SemR, the speaker constructs, through a series of other steps, the Phonetic Representation [PhonR] of the utterance. He/she does it following the Meaning-Text Model of his/her language[MTM (L)], which maps elements and configurations of the SemR onto all actual linguistic elements that make up the corresponding actual utterance. At this level, the speaker has the meaning and the phonetic string of the utterance or the signifier. Therefore, an utterance is produced in two major steps using two models and involving three major representations.

It claims that the phraseological nature of a complex expression lies in the constraints exerted on the selection of the signified and signifier (constituent lexemes forming the phrase) at the level of conceptual representation and semantic representation. Seen from this point of view, the mystery of formation and distinction of various types of phrasemes can be penetrated. Through this theoretical framework, a clear separation is first established between phrasemes and free combinations.

According to Mel'čuk (1998: 27): “For a phrase to be free means freedom of selection (of its signified and its signifier – with respect to the given ConceptR, that is, in the ultimate analysis, to the given situation; and of its signifier – with respect to the corresponding SemR) and freedom of combination (of its components: according to their own signifieds and syntactics plus general rules of L).”

Thus, each one of the components of a free syntagm can be substituted by any other synonym without altering the meaning of the syntagm, provided special properties of that synonym are taken care of (being of the same part-of-speech of the initial component) (Mel'čuk 2010: 2).

As distinct from free phrases, phrasemes violate this basic principle of unrestricted selection of components. A syntagm is a phraseme or non-free combination only if at least one of its constituents is restrictedly selected (not freely selected) according to the rest of the components. In this way, none of the phraseme's components can be replaced by any other synonym, except some few cases where an expression can admit one or two variants.

When this violation occurs in the transition between the ConceptR and SemR of a linguistic expression, where the speaker chooses a given SemR in function of the initial ConceptR, the outcome is a pragmateme or a pragmatic phraseme (Mel'čuk 1998: 27, 2010: 2; Mel'čuk: In press). Regarding their lexicographic treatment, pragmatemes, though pragmatically non-compositional, are semantically compositional and therefore there is no need for semantic definitions of pragmatemes. However, on the pragmatic level, pragmatemes are non-compositional, they are also unpredictable, hence the necessity to record them in monolingual and bilingual dictionaries and to indicate among other things the contexts of their use.

When the speaker chooses a syntactic representation for a given SemR without being predetermined by a specific situation, the result is a semantic phraseme. Semantic phrasemes are mainly idioms and collocations. They are distinguished based on semantic compositionality. In collocations, one of the components is restrictedly selected in function of the other which is freely selected, while the selection of components of an idiom is made fully restricted (Mel'čuk: In press). Therefore, four classes, i.e. collocations, idioms, proverbs and pragmatemes emerge from Kinyarwanda phrasemes, that linguists had mistakenly mixed up into a catch-all class called “*inshóberamáhaángá*” (idioms) in Kinyarwanda (Nikuze 2006, 2008). Some examples of these classes of phrasemes are provided below.

### **Idioms:**

- *Kumara imyaambi mu rugé* (to use some thing up, lit. to use up all arrows in a quiver)

- *Gutámika umuuntu undí* (to set against each other, to cause a disagreement between, lit. to put some body into the mouth of somebody else)
- *Gufáta umwuúka* (to go for a walk, to relax, lit. to take air)
- *Kuririra mu myóotsi* (to have problems/difficulties in an unfavourable place, to be hopeless, lit. to cry in smoke)

**Collocations:**

- *Kwiíruka amasígamáana* (to run as fast as your legs could carry you)
- *Kugeenda nk'ímwambi* (to go very fast)
- *Gukóra icyáaha* (to commit a crime)
- *Imvúra y'ínuunduura* (a heavy rain)

**Proverbs:**

- *Abahiígi beénshi bayobya imbwá uburarí* (Too many cooks spoil the broth, lit. Too many hunters mislead dogs): Understand: Many commanders are misleading.
- *Ak'ímuhaná kaaza imvúra ihíse* (A bird in the hands is worth two in the bush/Tomorrow is another day, lit. A help from outside comes after it has stopped raining): Understand: It is good to be self-reliant and not rely on external assistance which often comes late.

**Pragmatemes:**

- *Birabújjwe kunywéera itaábi hánó!* (No smoking)
- *Muraaho!* (Hello!)
- *Amahírwé masa!* (Good lucky)
- *Murakóze!* (Thank you)

According to the founder of the Meaning-Text Theory, phrasemes should be allocated a place in the dictionary according to their respective types and corresponding lexicographic status (Mel'čuk 2010). Therefore, the MTT definition of the concept of phraseme provides the basis for the lexicographic treatment of phrasemes in that it recognises the lexical status of idioms as first level treatment units and shows that collocations and proverbs as well as pragmatemes are too much phraseological to be treated as mere illustrative sentences in a general dictionary, but as first or secondary level treatment units in their own right, as each of them can be a subject of lemmatic or sublemmatic addressing. Mel'čuk (2010) insists that, whereas idioms are lexical units and must thus be included in individual articles as first level treatment units, it is different with collocations and pragmatemes and he suggests to present them as secondary level treatment units. However, due to the linguistic bias of the Meaning-Text theory approach to lexicography, this theory has not developed valid lexicographic principles for the compilation of user-friendly dictionaries. Moreover, it does not take note of typical lexicographic considerations such as dictionary functions and users' needs that along with linguistic factors are taken into account when designing the access structure of the dictionary. This research emphasises the need for implementing relevant lexicographic guiding principles when presenting phrasemes in Kinyarwanda dictionaries.

### 3.3 Phrasemes and Kinyarwanda studies: the extent of the problem

#### 3.3.1 Kinyarwanda phrasemes in linguistic studies

In a number of linguistic and ethnolinguistic publications on Kinyarwanda, the concept of phraseme is presented as a catch-all concept covering idioms, collocations as well as other simple forms of Kinyarwanda oral literature such as riddles, proverbs and “wellerisms”. In this case, one can mention Bigirumwami (1967) who included proverbs, riddles, and what he called "sayings" in his collection. With this in mind, we can also see the publication of Kayigana (1971 reprinted in 2007) that provided proverbs and explained expressions at the same time. That same year, Overschelde published a collection of 1150 sayings (amareénga in Kinyarwanda). In this collection, every headword was followed by an explanation in French. Father P. Schumacher (1956) had also listed some phrasemes in his *Dictionnaire Runyarwanda-Français/Français-Runyarwanda*. However, like in Bigirumwami, what he calls idioms is a mixture of wellerisms, phrasemes and paremiological formulas. There have also been few attempts to theoretically address some aspects of phraseology in Kinyarwanda, but these linguistic reflections have remained at the level of raising questions rather than providing answers (Bizimana 2002; Kagame 1977; Kimenyi 1999, 2000; Uwilingiyimana 2000; Munyakazi 1992; Nkusi 1986, 1986a).

For instance, unsatisfied with the lack of clear-cut separation between idioms and proverbs, Munyakazi (1992: 231-241) has proposed a set of criteria that could help in distinguishing these two language features. In his paper, Munyakazi considers all Kinyarwanda phrasemes as being idioms or lexicalised phrases which are known to be non-compositional. He ignores other types of phrasemes, i.e. collocations, pragmatemes, etc. It could be argued that collocations and pragmatemes are not phraseological expressions because they are semantically compositional, which means that the individual constituents of a collocation or a pragmateme maintain their own meanings. This conception is due to a traditional idiom bias in phraseological research. As it has been argued in sub-section 3.2.2, the phraseological nature of complex expressions depends not only on the semantic compositionality principle, but also on lexical restrictions for collocations and pragmatic restrictions for pragmatemes. Therefore, the concept of phraseme as it is used by Mel'čuk and other scholars following his theoretical framework is taken in a broad sense, encompassing semantically non-compositional phrasemes (idioms, proverbs) as well as restricted co-occurrence and pragmatic phrases that are semantically compositional.

It can also be observed that contrary to what is said above about the lack of universality in phrasemes'

meanings, collocations can have universal meanings, because it is acknowledged that some general meanings expressed in all languages tend to be expressed by collocations (Polguère 2003:137). In addition, the operations involving syntactic modification are very restricted in the case of noun-idioms. In the case of idioms displaying the Verb + Noun structure, these syntactic modifications are a function of syntactic flexibility inherent to the verb and characterize similar verb-collocations in Kinyarwanda.

However, this linguist has found other serious weaknesses in the typologies proposed by his predecessors, weaknesses that greatly compromise the definition of the concept of a phraseme in Kinyarwanda. Indeed, Munyakazi (1992: 235) wonders why expressions such as *gutéera isékuru* (to hobble or limp along, literally to throw a mortar), *kugwa gituumo* (to surprise, lit. to fall hurriedly on somebody or something) are considered phrasemes (*inshóberamáhaángá*) while the same cannot be said of *kubá mu máazi abíra* (to be in awkward situation, literally to be in boiling water) or *gutúunga ibyaa Míreenge* (to be very rich, wealthy, literally to possess the goods of Miréenge).

For him, there is also a need to distinguish between the plethoric terminology used to refer to phrasemes such as *incáamareénga* or *amareénga* (veiled expression/language), *inshóberamáhaángá* (idiomatic expressions), and *imiganí* (proverbs). He proposed the term *amabaango atánaze* (frozen expressions) to designate phraseme proverbs exclusively. It is important to remember that Munyakazi does not refer explicitly to other types of phrasemes such as collocations and pragmatemes. In this way the idiom bias recurs in Munyakazi's work.

Another ethnographic work regarding phrasemes is a collection published in two volumes entitled *Ibirali by'insigamigani I, II* by Benoît Mulihano in collaboration with the Ministry of Education (Mulihano & Mineduc 1980) and later on with the Ministry of Higher Learning Education and Scientific Research (Mulihano & Minesupres 1986). The purpose, as the title suggests (On the origin of our proverbial sayings), was to recover or trace back the history of famous characters immortalized by collective memory in idiomatic expressions containing their names or names of places where they have performed memorable actions. These stories are believed to be informative regarding the origin of the expressions or in other words, their etymology.

Nevertheless, soon after this work had been released it became the subject of considerable controversy. Thus, Nkusi (1986a) did not hesitate to consider the collection of Mulihano and Mineduc (1980) and Mulihano and Minesupres (1986) as folk etymology. Nkusi showed with striking arguments, that the method used by Mulihano and co-authors should be supplemented by an etymological approach which

is likely to postulate an intrinsic linguistic relation between the meaning and the form of the expression by studying the morpho-semantic structures of those expressions. Therefore, Nkusi (1986: 12) suggests completing and reviewing the collection of Mulihano and Mineduc (1980) and Mulihano and Minesupress (1986) in order to base the origin of popular expressions on historical evidence.

In this perspective, an etymological analysis of Kinyarwanda lexical items, including multiword expressions, has been attempted by Uwilingiyimana (2000: 31-45). This study includes a section on formal structures of Kinyarwanda compounds and phrasal items. Lexical units are subdivided into three important categories: simple lexical items, complex lexical items and marginal lexical items. The theoretical framework of this study was inspired by a scholarly approach to etymology initiated by Guiraud (1967, 1979) in France during the seventies. Uwilingiyimana (2000:17) points out that most of Kinyarwanda items are complex, interlinked and juxtaposed, which makes the acquisition of and access to the lexicon difficult for language learners and users.

In his study, Uwilingiyimana (2000) proposes an innovative morpho-semantic classification of Kinyarwanda lexical items. It examines complex items where constituents are linked by a connective pronoun with a sense of possession such as *ikiziba cy'índa* (abdomen, lit. puddle of stomach), *uruti rw'umugóongo* (vertebral column, lit. big tree of the back); phrasemes whose constituents are linked by a coordinator conjunction “na” (carrying the meanings “and”, “proximity”, “big”), e.g. *urujuyá n'úruúza* (comings and goings), *agahuumbi n'ágahuúmbagiza* (uncountable hundreds of thousands); and interrogative expressions with *hé* meaning *where*, e.g. *akaári aha kaajya hé* (incorrigible, lit. where is going what is here) in the expression *kwiigira akaári aha kaajya hé* (to make oneself incorrigible). There are also phrasemes containing proper names of persons or places, e.g. *gukízwa na Mbúga* (to be hardly reconciled, lit. to be saved by Mr. Mbuga), *kuruha uwa Kavuna* (to work hard vainly, lit. to make an effort like that of Mr. Kavuna), *kwíiraaza i Nyaánza* (to feign not to know, lit. to-ask-for-lodging at Nyaánza<sup>2</sup>).

However, apart from the phrasemes that contain proper names, the phraseology of Kinyarwanda in its diversity has not been comprehensively catered for in this study. Furthermore, the only lexicological study of phrasemes is not helpful as far as their lexicographic treatment is concerned.

This sub-section cannot be concluded without mentioning another prominent figure of Kinyarwanda linguistics. In two articles on cow and journey related metaphors in Kinyarwanda, Kimenyi (1999,

---

<sup>2</sup>Nyaánza is the name of many villages in Rwanda including the former capital of the ancient Kingdom of Rwanda.

2000) suggested a cognitive approach to investigating semantic mechanisms at work in idioms. He went on emphasising the role of metaphor, metonymy, de-iconicity and cultural symbolism in the lexicalisation of idioms. For him, idioms are like other tropes; the only difference is that an idiom is a multi-lexemic construction made of two or more constituents in deep or surface syntactic relation. His approach, when dealing with this kind of linguistic expressions, is helpful to the study of phraseological expressions in cognitive perspective. Nevertheless, the only psycho-cultural side of those expressions does not allow an explanation of all the mechanisms used in the generation or formation of all idioms. This is, for example, the case of idioms wherein constituents do not display a conceptual relation with the meaning. In addition, although such an approach could contribute to a lesser extent to improve the semantic description of phrasemes, it does not help resolving the issue of lexicographic presentation of phrasemes in general.

In Kinyarwanda, semantically meaningless or vague verbs combine with predicative nouns to form a good number of collocational phrasemes. These supporting verbs bring a communicational structure (time, mode and aspect) to predicative nouns and make them able to express processes, events or actions. A similar category of achievement verbs has also been identified, (cf. Mel'čuk 2004). For a long time expressions containing support verbs have confused Kinyarwanda linguists. About these phrasemes formed on semantically empty verbs, Kagame (1977: 117-118) says that, paradoxically, neither the first nor the second term leaves traces in the resulting meaning, while acknowledging earlier that this verb provides a contextual meaning that needs to be specified by a keyword.

This confusion is due to ignorance of collocations and mechanisms leading to the formation of this type of phrasemes. Therefore, what Kagame (1977) missed is that these expressions are collocations. For instance, in the expression *gutéera ishyeengo* (to chat in amusing way), the second word *ishyeengo* (fun) that is a noun that is a predicate with one argument (*ishyeengo rya X*: fun by X). Therefore, in this expression, the verb *gutéera* verbalizes the noun which is the semantic base of the whole construction. As to the verb, it is a syntactic support that provides the noun with aspectual and communicational structure proper to the verb. It contains an aspectual meaning of a provoked movement of an object from one point to another. It is indeed, a collocation which is equivalent to *gushyeenga* (to fun). The same could be said of *gutéera imbabázi* (to be an object of sympathy) = *kubábaza* and *gutéera itáma* (to put food in the mouth at one go, *itáma* (cheek), lit. to make cheek swelling or to put into cheek) = *gutámira*, etc.

In the last decade, the above-mentioned expressions of Kinyarwanda have been identified as support

verb collocations (Nikuze 2006, 2008). These studies focused on formulating a classification of phrasemes using definitional criteria developed within the framework of the MTT. A model for lexicographic representation of Kinyarwanda collocations is outlined in Chapter 5.

Another study worth mentioning in this review is by Bizimana (2002:116-119) who devotes a cursory analysis of Kinyarwanda phrasemes, classifying them in conceptual domains according to the object referred to by the nouns found in the expressions. These domains are human, animal, parts of the body, place, plant, miscellaneous things. He also alludes to the transparency and opacity and briefly the gradable nature of compositionality in the expressions. He also indicates that in a small number of phrasemes, there are lexemes which are never used in any other context, except in those expressions. However, his thematic stance would have been useful had it used the anthropological or cognitive approaches. Such approaches are useful in understanding metaphorical and metonymic mechanisms that underpin the exchange and transfer of semantic characteristics between humans, animals, objects, plants, etc.

### **3.3.2 Phrasemes in the metalexicography and lexicography of Kinyarwanda**

It is of utmost importance to consider the way phrasemes have been treated by Kinyarwanda (meta-) lexicographers, in the light of modern approaches of lexicography, and more precisely the two underpinning lexicographic theories of this study, i.e. the function theory of lexicography and the theory of dictionary structures. In the following sub-section, the problematic issues discussed in the foregoing section are illustrated through a review of important dictionaries of Kinyarwanda with reference to insights from theoretical lexicography. It is assumed that along with other factors, the prevailing confusion between idioms and collocations has something to do with the concept of a heavy word.

#### **3.3.2.1 Phrasemes in the metalexicography of Kinyarwanda**

In the framework of metalexicography, proposals exist as to the presentation of phrasemes in the dictionary. For example, regarding the lemmatisation of idioms, Gouws and Prinsloo (2005:88) state that they are fully-fledged lexical items and given that idioms cannot be easily accommodated as primary level treatment units, they suggest to present an idiom as a secondary level treatment unit in an article attached to the article of the constituent taken as guiding element of the idiom. In this approach, idioms are given a macrostructural status albeit presented as nested sublemmata in a special text block attached to the article of the main lemma (Gouws 2010a). Yet again, it should be remembered that the problem of selecting the guiding element of an idiom is far from being agreed on by scholars in

lexicography. Besides, what is said about idioms could equally concern collocations and pragmatemes because they are often treated in almost the same way although they can also be used as illustrative materials.

With regard to the choice of the guiding element, Simon Bizimana, one of the most prominent Kinyarwanda scholars in lexicography and linguistics, proposes the morphological stability as guiding criteria for identifying the so-called strong word or guiding element of the idiom. In his book on lexicology, lexicography and semantics of Kinyarwanda, he advises that phrasemes should be described in the article of the main and stable lexeme (key word) among their constituents. He also suggests that a cross-reference from the other constituents of the expression should direct the user to the main word where the idiom is recorded and treated (Bizimana 2002:134). This conception is taken up again in the front matter of the first edition of his *Inkoranya y'Îkinyarwanda mu Kinyarwanda* (Kinyarwanda Monolingual Dictionary), wherein he argues that a stable lexeme is the one which does not undergo flexional or conjugation-related modifications, which is not a conjunction, a connector or a determiner (IRST 2002: xviii). Obviously, this scholar prefers morphological stability to semantic stability, excluding therefore, the option of describing the expressions in the articles of verbs. However, there are serious arguments in favour of a contrary hypothesis. These arguments lead to considering both the noun and the verb as possible candidates for phraseme description, according to their respective lexico-semantic status and role in the combination.

André Coupez, the Belgian scholar whose work has influenced Kinyarwanda lexicography in many respects is inclined to favour a slightly different option. In a chapter devoted to methodological issues in the lexicography of Bantu languages (Coupez 1985), he notices that if the word with its lexical morphemes provides the basis of the dictionary, the latter also hosts the grammatical morphemes and phrasemes. He says that idioms are autonomous lexical units, or phrases whose meaning is not deductible from the constituent words. He therefore advocates describing them in the article of one of their constituents with cross-references from the articles that have the remaining words as lemma signs. He indicated that the word chosen for this purpose should be the most semantically precise, and in this perspective some vague verbs such as *kugira* (to act, to do, to say, to have, to think) whose meaning is very wide and enters into dozens of phrases, are not favourite candidates to host phrasemes in their articles Coupez (1985:103). This position comes to the fore in the dictionary that he co-authored and which was published twenty years later (DIFO2005: xxiv) wherein the option of mentioning phrasemes in the articles of all lexemes constituting the expression for being eventually described or cross-

referenced to the lexeme under which it is described, is reasserted. However, his position remains unclear, at least in practice, as to the selection of the guiding element. Thus, as the Coupezian approach seems to be incomplete, other issues that need to be looked into are raised. These issues concern the lexical status of phrasemes.

It is apparent that when Coupez discusses phrasemes, he thinks about both idioms and collocations because these two types of phrasemes had been confused in Kinyarwanda dictionaries and metalexicography. In fact, what Bizimana and Coupez refer to as restricted constituents and semantically heavy constituents are different ways of designating the same thing because they converge on including phrasemes (idioms and collocations) in the article of the nominal constituent taken as the guiding element of the phrasemes. However, if this policy can be maintained with a bilexical expression displaying a *V+N structure*, a difficulty still exists with an expression with more than two lexemes, where the verb is followed by an expansion of many other lexemes on the one hand, and with binominal phrasemes in which two nouns are not in a determination relation and with phrasemes wherein all constituent words are verbs on the other hand.

Irrespective of their microstructural status, I concur that support/achievement verbs should be included in the articles of the verbs for text production purposes. What about collocations where the verb is a semantic base and the noun serves as a modifier playing an adverbial role? In this case, the verb is most appropriate to receive this kind of expression because the meaning of the verb is the most important in the expression, e.g. *kugeenda amasígamáana* (to walk fast). In this expression, the verb *kugeenda* (to walk) is freely selected while *amasígamáana* (fast, lit. in a way to get ahead of God) is selected to express a meaning **Y** according to the meaning of the verb *kugeenda*, and cannot be replaced by another word without affecting the correctness of the expression. This may be different from *gukóra amahano* or *kuréenga umuziro* (to do what is prohibited, to break a taboo) where both nouns are the bases which are verbalized by means of support verbs. Here, the noun is the most important semantic node and on the *linguistic basis* the inclusion of the collocation in the article of the noun could be justified, but not in the former case. Similarly, *gufáta mpiri* (take an opponent alive) and *kugeenda buruundú* (to leave for ever) always keep up the basic meaning of the verbs. And what about expressions where there are only two verbs, *kwíishyira ukízana* (to enjoy freedom, to be free), or the expression where the noun is substituted by an object pronoun, like in *kuruca ukarumira* (to keep quiet, to remain silent, lit. to cut one's tongue and swallow it) in which *-rú-* substitutes *ururími* (tongue)? Should this expression be treated under *ururími*, under *gucá* or *kumira*? These questions show that this

debate is far from being closed.

Furthermore, it should be remembered that a dictionary is a utility tool designed to meet different user needs and dictionary compilers should negotiate lexicographic functions when devising a user-friendly lemmatisation policy. A dictionary can be exclusively monofunctional, e.g. for either text reception or text production, whereas others could be bi- or multifunctional. Indeed, the role of a dictionary is not only to assist in the search for the meaning of a word that the user does not understand in a text, but also to assist in finding an expression conveying the message he/she has in mind and intends to express. For example, which word that combines with *kugeenda* (walk) expresses the meaning of walking fast (semantic information)? Naturally, in the encoding perspective, the basic concept or semantic nucleus comes to mind first, before the modifying or determining word. And nowadays, where multilingualism and the need to work in a context of foreign languages is inevitable, dictionaries are more and more being used for encoding and communicative purposes not only in language teaching, but also in translation and professional writing activities. Therefore, this new lexicographic paradigm should be taken into account if one wants to make a dictionary a utilitarian, multipurpose and communicative tool, encouraging an intensive use of language. In keeping with this view, Larivière (1998: 189-190) emphasises the importance of presenting specialised collocations used in finance language in both articles of the base and collocator in a multi-functional LSP dictionary to assist users who might consult it either for text reception or text production related needs. Unfortunately, the wide-ranging nature of Kinyarwanda phrasemes, the diverse dictionary functions and the user needs are not taken into account when deciding on the lemmatisation of phrasemes in Kinyarwanda dictionaries.

Sublemmatisation of phrasemes has been the practice in Kinyarwanda lexicography. In general, the noun is used as if it is always semantically more significant or precise than the verb. It seems that the morphological criterion dominates here because the noun is less variable than the verb. The only advantage of this approach is that it avoids overloading the articles of verbs that in nature are already more polysemic, hence helping to balance the size of the dictionary's articles. The polysemy of verbs and the tendency to monosemy of nouns are natural. I assume that users are also aware of this fact which is reflected in their daily manipulation of those lexical items. Nobody would want to reverse this state of things. Therefore, it is theoretically irrelevant to record phrasemes under nouns merely for the sake of changing this situation if it is not justified by considerations of user needs.

Further, the presumed semantic significance of constituents is impossible in idioms because they are non-compositional expressions. It is only possible in compositional phrasemes such as collocations and

pragmatemes. The sense of an idiom has nothing to do with the individual meaning of its constituents. The search of a heavier or precise constituent to be used as guiding element of phrasemes results from an abusive generalisation entailed by the confusion of compositional and non-compositional phrasemes in Kinyarwanda lexicography. Due to the lack of a clear-cut distinction of various types of phrasemes, dictionary compilers think that what is peculiar to one type – e.g. semantic heaviness or precision of one constituent in a collocation – is also characteristic of all phrasemes.

As far as Kinyarwanda is concerned, a thorough classification of phrasemes is obviously necessary if one wants to establish a sound method for lexicographic representation of phrasemes. Complementarily, it is absolutely essential, for the lexicographer, to set priorities and determine in advance who will be the targeted users of the dictionary and what they should expect from it in connection with their needs and with respect to phraseological items.

### **3.3.2.2 The lexicographic representation of phrasemes in current Kinyarwanda dictionaries**

The foregoing sub-section has already identified most of the problematic issues related to phrasemes in metalexicographic and linguistic works. Given that the ultimate purpose of this research is to propose a lexicographic theoretical model which would spearhead a better treatment of phrasemes in dictionaries, it is time to look at how phrasemes are treated in the most frequently used dictionaries of Kinyarwanda. This will provide a concrete expression of the problems mentioned above for which the present research is expected to propose solutions. This sub-section is dedicated to a critical and contemplative analysis of the inclusion and treatment of phrasemes in the major dictionaries of Kinyarwanda. It discusses challenges and problems posed by the current inclusion practices and how the latter undermine the easy accessibility of lexicographic data and the lack of cognisance of user needs. It should be borne in mind that this presentation will be limited to the analysis of facts as they appear throughout dictionaries and that some assumptions were made as to the difficulties the user might have when looking for a phraseme in the dictionary. The verification of these assumptions through research in dictionary use would have been more instructive, but such investigation is beyond the boundaries set for the present study. Thus, this section discusses and illustrates only aspects related to this study.

Moreover, this review does not cover all dictionaries which have been made; it deliberately only focuses on four general dictionaries which are considered to be the most important in terms of volume, contents and elaboration methods, and which are inclusive enough and systematic to allow the observation of diverse lexicographic and linguistic phenomena. A small introductory background will be given of each one followed by two dictionary articles that will be presented and discussed for each

dictionary focusing on the presentation of phrasemes.

- ***Dictionnaire Rwandais-Français de l'Institut National de Recherche Scientifique* by Father Irénée JACOB**

This bilingual monodirectional Kinyarwanda-French dictionary is a result of lexical research that was started in 1965 at the former Institut National de Recherche Scientifique/INRS (the current IRST) and the Centre d'Etude des Langues Africaines (CELA) where Father Irénée Jacob adapted the lexical database compiled by linguists of the former INRS to produce the present dictionary published in three volumes in 1983, 1985, 1987 respectively. It was published as a test edition of the more comprehensive dictionary projected by the INRS (*Dictionnaire de l'INRS*).

The macrostructure is ordered alphabetically, the lemma is written in its full form, but like in most Kinyarwanda dictionaries, ordered according to the first letter of the root which is given in bold. The lemma is followed by an indication of its grammatical category in French (perfect aspect for verbs, pronoun prefix and class number for nouns), a cross-reference to a variant or derivational form where applicable, a meaning followed by usage examples in Kinyarwanda in italic fonts, their translations in French, and eventually the Kinyarwanda phrasemes labelled *Loc* (locution, idiom) with their definitions in French, and the same applies to proverbs (marked prov.). Where there is more than one sense (polysemic lexemes), different senses are marked with Arabic numerals, while roman numerals are used to indicate the order of homonyms in the case of homonymic pairs. The orthography used in this dictionary is slightly different from the current official orthography of Kinyarwanda.

In the front matter of *Dictionnaire rwandais-français de l'INRS* (Jacob 1985), no information is given regarding the treatment of phrasemes and therefore, their inclusion and treatment policy seems arbitrary and inconsistent. “Locutions” as idioms and collocations are jointly referred to in this dictionary, are treated in the articles of either the first or the second constituent. In general, they are mentioned as a specific category (loc) in the microstructure or merely as illustrations without any indication of phraseology. Where they are explicitly indicated, they are listed either under relevant senses or under the last sub-comment on semantics. In addition to the absence of a consistently and uniformly implemented inclusion policy, the inner-access to idioms in Jacob (1985) is rendered more complicated by their being mixed with semantic items without being allocated a separate text slot within the microstructure and without being included as sublemmata. The frequent use of proverbs to illustrate the use of the proper senses of words can be questioned because in many cases the meaning conveyed by a given word used in a proverb differs significantly from the same word used in a free sentence. For

example, the proverb: *Ahó inkoóngora zitalí ibishwí biratanaga* (Lit. *You have to cut your coat according to your cloth*) is provided as the illustrative sentence of the lemma sign *igishwí* (sparrow). However, the word *igishwí* is used in this proverb in an ad hoc metaphoric sense in such a way that its ornithological sense is not directly understood in the meaning of the proverb. The study of the following dictionary articles highlights a number of other issues.

#### ❖ Article of the lemma *isóni* (Jacob 1985 : 203-204)

**isóni**, (pl.). 1° Honte, humiliation (que ressent celui dont une mauvaise action est dénoncée, ou dont un ridicule apparaît). *Baamutaangaarije mu ruhamé kó yasaambanyije akáana k'ágakoóbwa nóone yagize isóni zó kujya ahagaragara* : on lui a imputé publiquement le viol d'une petite fille, et à présent, il a honte de se montrer en public. 2° Timidité, modestie, pudeur, retenu. *Yagize isóni zó kwiicara hamwé n'ábaándi kubéera kó yaambáye inshabali* : il est gêné de s'asseoir avec les autres parce qu'il est vêtu de haillons. *Isóni zikuliisha amarozi* (prov) : la timidité te fait avaler le poison (la timidité cause parfois les malheurs). *Umukwé w'iisóni ahera muu mfúruká* (prov.) : un gendre timide reste dans le coin de la maison (la timidité prive de certaines jouissances). 3° Crainte révérentielle, respect. *Nagize isóni zó kweegeera báliýa batégetsi mbóna kó baabá báfité ibyó biivúganira ubwaábo* : je n'ai pas osé aborder ces hautes personnalités car je pensais qu'elles avaient à traiter entre elles. *Igituúntu cy'ibukurú kiliibwa isóni* (prov.) : la tuberculose chez les grands se mange par crainte révérentielle (mieux vaut risquer la maladie que la colère d'un grand. – Si un personnage important, pris de tuberculose, t'invite à partager avec lui, tu n'oses pas refuser, au risque d'attraper la maladie.). 4° Franges de la couronne royale *igisíingo*. **Loc.** *guhísha isóni/ kubíika isóni* : se, coucher, dormir. *Gukóra ibitéeye isóni* : commettre des actes impudiques. *Gukórwa n'isóni* : avoir honte, être humilié ; être indigné, s'indigner de. *Gukórwa n'íiz'íamóoko* : être un objet de risée. *Gukóza isóni* : confondre (qqn), l'humilier. *Gushíra isóni* : être éhonté, irrespectueux, effronté ; refuser d'obéir. *Gutá isóni* : être ou devenir éhonté, insolent, effronté. *Gutééra isóni* : être humiliant, confondant, inconvenant. *Kugira isóni* : avoir honte, ne pas oser. *Kulyá isóni* : être excessivement insolent ou effronté. *Kuzinnya* [kunnya isóni, déféquer la honte] : être éhonté, ne pas avoir honte de commettre des actes répréhensibles. *Inkíngi y'iisóni* : pilier situé à droite de la pièce où l'on couche, dans la maison traditionnelle. *Uruugi rw'iisóni* : panneau situé près de l'entrée de l'alcove (au travers duquel la jeune mariée regarde les gens qui sont déhors, lorsqu'elle est en période de claustration).

In this article, the word *isóni* (shame, being ashamed of doing something) has generated all the expressions mentioned after the sub-comment n°4. However, these expressions have no relation with that sub-comment; rather, they relate to the meanings n° 1, 2, 3. At least 8 out of 14 expressions are collocations and they should have been distributed under relevant senses instead of listing them at the end of the article. For idioms, their being listed under the last sub-comment can be justified; however this practice is not applied in a consistent way because in other polysemic articles they are included in a different way, along with specific senses. This proves the lack of a suitable methodology for the placement of phrasemes in this dictionary.

#### ❖ Articles of the homonymic lemma sign *gutééra* (Jacob 1985: 321-323)

**gutééra**, -ye I (cfr *gutá*). 1° Jeter, lancer (vers un objectif déterminé). *Rutwé yamuteeye icúmu undí*

*alúzibukiira* : Rutwe lui a jeté une lance et l'autre l'a escquivée. *Nanyuze imbere y'íiwníhitira abáana baawe baanteera amabuye uzáabahané* : lorsque je suis passé devant toi, tes enfants m'ont lancé des pierres, tu devras les punir. 2° Semer (à la volée) ; planter, mettre en terre. *Amasaká n'úburó babiteera babímiisha* : le sorgho et l'éleusine se sement à la volée. *Uteera ibishyúimbo bikamera bímaze imiínsi munaáni, bikazaasaaruurwa bímaze améezi atatu* : quand on plante des haricots, ils germent après huit jours et se récoltent après trois mois. *Ubusáanzwe insína zikuunda gutéérwa ku muhiindo* : en général les bananiers se plantent au début de la petite saison des pluies. *Uúhiingga imyuumbati cyáangwá ibishéke atééra ingéli* : celui qui cultive du manioc ou de la canne à sucre plante des boutures. Loc. *gutéérwa ukamera* : traîner longtemps quelque part.

**gutééra**, -ye II (cfr **gutá**), ficher un objet pointu dans, percer au moyen de. Gutéera umuuntu icúmu : percer qqn d'une lance. Gutéera umuntu icyúuma : enfoncer un couteau dans le corps de qqn. Gutéera umuuntu urushíingé : faire une piqûre à qqn, lui faire une injection. Gutéera umusúmaári : enfoncer un clou.

**gutééra**, -ye II (cfr **gutá**), pondre (animaux ovipares). Mfite inkokó itéera amagí atatu ku muúnsi : j'ai une poule qui pond trois œufs par jour.

**gutééra**, -ye IV (cfr **gutá**), jouer à certains jeux (traditionnels ou étrangers) où le matériel est lancé à une certaine distance (petite ou grande selon le cas). Gutéera uruzíga : pratiquer le jeu ce cerceau. Gutéera umupiíra : jouer au foot ball. Gutéera amakalíta : jouer aux cartes. Gutéera urusiímbi : jouer aux jeux de hasard (urusímbi).

**gutééra**, -ye V (en parlant des cucurbitacées), fructifier. Rwáa ruyúzi Gahiinda yaruranduye rwáali rútaangiye gutééra : Gahinda a déraciné cette courge alors qu'elle avait déjà commencé à produire. Urujyá gutéera rubanza urutúutú (prov.) : le (plant de couge) qui va produire commence par fleurir (tout événement a ses signes avant-coureurs).

**gutééra**, -ye VI (cfr **gutá**), frapper avec une partie déterminée du corps. Gutéera umugelí : donner un coup de pied ou de patte, ruer. Gutéera iheémbe : encorner. Gutéera ingumí /gutééra ikoófe : donner un coup de poing.

**gutééra**, -ye VII (cfr **gutá**), battre (coeur) ; avoir des pulsions (artère). Iyó umuuntu amazé gucá umutíma ntíwoongéra gutéera : quand une personne vient de rendre l'âme, son cœur ne bat plus.

**gutééra**, -ye VIII (cfr **gutá**), fixer, coller une chose sur une autre. Gutéera ibifuungo : coudre des boutons (sur un vêtement). Gutéera ikiremo : repiécer. Gutéear ibára : orner d'un dessin. Gutéeara ikiraká : réparer une chambre à air à l'aide d'une rustine. Gutéera iraángi : peindre. Locgutéerwa hó uundí : avoir besoin de l'intervention de qqn pour faire qqch.

**gutééra**, -ye IX (cfr **gutá**), mettre un support à. Mfasha iyi módochá tuyitéera ijekí tuboné ukó dukuurá mó umupiíra watobótse tuwuhomé : aide-moi à soulever cette voiture au moyen d'un cric en vue d'enlever le pneu crevé et de le réparer. Kugira ngo inzu itazágwa bayiteera inkiíngi imbere : on met des piliers à l'intérieur d'une maison pour l'empêcher de s'écrouler. Iki gitooke kigiiye kugwa uragiteere injáliro : ce régime de banane menace de tomber, mets-lui un support.

**gutééra**, -ye X (cfr **gutá**), disposer (un siège), le placer de telle ou telle manière. Ko mutaraatééra intébe kaándi abatumírwa báli haáfi yó kugera aha ! vous n'avez pas encore placé les sièges alors que les invités sont sur le point d'arriver.

**gutééra**, -ye XI (cfr **gutá**). 1° Attaquer. Bavuga kó Rwáabugili yalí umwaámi uhóra muu ntaambara, kó yatéeye Ubunyabuungo, agateera Inkóre n'áhaándi heénshi : on raconte que le roi Rwabugili était tout le temps en guerre, qu'il a attaqué le Bunyabungo, l'Ankole et beaucoup d'autres régions encore. Ahó

umugabo ateeréwe ni hó ariindíra (prov.) : où un homme est attaqué, c'est là même qu'il se défend (la meilleur défense est celle qui se produit au moment même de l'attaque). Ntaa wutéera uwó ataátase (prov.) : nul n'attaque celui qu'il n'a pas espionné (toute attaque doit être préparée). Loc. unteeéra cg uraánteeye ! je regrette, cela ne me regarde pas du tout !, dommage ! Umunwa wa rúguru uti aya matá yali aályooshye uw'eépfo uti untéere jye nali níibeereye inyuma y'ícyáánsi (prov.) : la lèvre supérieure dit : ce lait était savoureux ; la lèvre inférieure réplique : dommage ! moi, j'étais hors du pot (on ne témoigne que de ce que l'on connaît). 2° Chasser, déloger (qqn) de ses biens pour se les approprier. (...) 3° Tenter d'avoir part aux biens d'autrui alors qu'on n'y a pas droit, essayer d'en profiter. Kó daatá yasíze adúhaaye iminaáni y'úbutaka twéese unteeeye mu wáanjyé bígeenze bíté ? avant de s'éteindre, notre père nous a donné à tous une part de terrain. Pourquoi veux-tu donc empiéter sur la mienne ?

**gutéera**, -ye XII (en parlant d'un événement nouveau : épidémie, famine, sécheresse, mode de vie, etc.), se manifester ou apparaître en plusieurs endroits, se diffuser, se répandre. Ubushiita bwaáratéeye bumara abaantu : la variole s'est déclarée et a décimé la population. Kó amapfá ateeyé ibishyfimbo byáacu bítaréera tuzaahaaha hé ? voici que la sécheresse commence à sévir alors que nos haricots ne sont pas encore mûrs, où irons-nous donc chercher des vivres ? Ahó gusuka inweéri biteeréye abakoóbwa bó mulí aaka karéeré ntíboóngeye gutéga amasuunzu : depuis que la mode de cheveux tressés s'est répandue, les jeunes filles de cette région ne portent plus de houpes. Inkweeto zaáratéeye ntáa wutákizaambaye : les souliers sont à la mode, tout le monde en porte.

**gutéera**, -ye XIII Parvenir à un endroit en commençant par. Nagiye gukúura ibiraayi umulimá nywuturuka ku kibúno nteera ku mutwé ntárúuzuza igiteebo : je suis allé arracher des pommes de terre ; j'ai commencé par le début du champ et je suis parvenu au bout sans avoir rempli un panier.

**gutéera**, -ye XIV (cfr gutéureka). 1° (en parlant de qqn, qqch), être constitué de telle ou telle façon, avoir telle ou telle forme ou proportion. Umuuntu utéeye néezá agira igikúundiro : une personne bien faite plaît. 2° Avoir telle ou telle comportement. Wáa mukoóbwa yaantútse. – Mureke ní kó ateeyé : cette fille m'a insulté.-Ne t'en fais pas, elle est comme ça. Loc. ateeye ukwé : il est bizarre, il n'est pas comme les autres.

**gutéera**, -ye XV 1° Causer, occasionner, provoquer, entraîner. Kó nduuví ubabáye wabiteewe n'íikí ? je te vois triste, quelle en est donc la cause ? Amagaambo wamu bwiíye yamuteeye uburaakali : les paroles que tu lui as adressées l'ont fâché. Amapfá ateera inzara : la sécheresse provoque la famine. Umwanda uteera indwáara : la malpropreté rend malade. (...) 2° Inspirer un sentiment, suggérer une idée. Ijaambo umbiíye liinteeye igitéekerezo : ton mot me suggère une idée.

**gutéera**, -ye XVI (syn. Guseembura 2°), mettre la levure dans la boisson. Amaárwa atéeje urwáagwá rukazé cyaane amena umutwé : la bière de sorgho que l'on a fait fermenter avec de la bière de banane très forte donne des maux de tête. Yashígishe inzogá yiibagirwa kuyíteera ntíyaabira : en préparant de la bière, il a oublié d'y mettre de la levure et la bière n'a pas fermenté.

**gutéera**, -ye XVII Entonner (un chant, une récitation, une prière), donner le ton ; chanter les couplets. Iyó tulifimba hateera batatu abaándi bakiikiliza : quand nous chantons, trois d'entre nous chantent les couplets et les autres reprennent le refrain. Loc. gutéera ukíiykiliza : vivre solitaire, peiner seul dans la vie.

The compiler of this dictionary considers the forms *gutéera* as homonyms where Roman figures (I to XVII) are used as structural indicators to identify the various homonyms. In most cases, the verb *gutéera* can be regarded as a light verb functioning as collocator in collocations. It is interesting to note that in most cases, the lemma is defined with reference to a word indicated in parentheses without

which the lemma cannot be defined. That word precisely explains the contextual use of the lemma. It can play the role of subject or object but in all cases it is the base of the collocation of which *gutéera* is the collocator. In addition, the illustrating phrases given are collocations. From a text production perspective, these collocations should be treated under the article of their respective bases instead of appearing exclusively in the article of the collocator.

The compiler himself realised this problem when confronted with the remarkable number of collocations (he uses the term locution everywhere) in which the verb *gutéera* occurs. In a text-block appearing under the last homonym, an additional 81 words which collocate with the verb *gutéera* are given. A note to the users explains in the following words their confusion with respect to the extent of the homonymic status of the verb *gutéera* (Jacob 1985: 323):

pour d'autres locutions, faute de pouvoir faire un tri sûr parmi les homonyms *gutéera*, raportez-vous aux mots suivants [For other expressions, failing an accurate distribution among the homonyms, you will have to refer to the article of the following words]: *imyaáma*, *icyáankokó*, *imyaánya*, *amabára*, *urubáliro*, *imbóni*, *ikibúno*, *ibuye*, *urubwa*, *ububyáará*, *imbyíno*, *inda*, *indiishyi*, *indómo*, *ingáamba*, *ingáta*, *umugelí*, *urugiímbu*, *umugóongo*, *ibigwá*, (*im*)*hakanizi*, *ubuhaké*, *amaháané*, *umuhará*, *heejuru*, (*im*)*huúndu*, *ijísho*, *injoongo*, *injóongo*, *ibika*, *inkiíngi*, *kanó*, *ikoontaro*, *igikúmwe*, *inkíunga*, *innyo*, *umunnyuúngu*, *inogó*, *inóni*, *nyaanteebo*, *urunyó*, *umunyúgwé*, *igipaández*, *ipaási*, *ikiraká*, *kureka*, *burekwe*, *ikiremo*, *kuréenga*, *ikireenge*, *ikiroro*, *indíurú*, *imirwi*, *umulyoóngó*, *isaamburuma*, *isékuru*, *sée-ntili*, *shiinguura II*, *shuumbú*, *inshuiúnzi*, *ishwiíma*, *ishyeengo*, *amashyíga*, *urusiímbi*, *abasogi*, *isóni*, *ubutaka I*, *amatáko*, *itáma*, *intáambwé*, *agatí*, *intoré*, *ibitúgu*, *intsítso*, *ubwuúzu*, *iví*, *amaví*, *waaraza*, *amáazi*, *inzúzi*, *izúuru* [Italics is their].

The lexicographers advise the user to refer to the respective articles of these lemmata wherein additional expressions are listed. This would have been the same with more of the expressions treated above as illustrative sentences; they would have been listed under their respective bases, not under the collocators. Actually, what Jacob takes as homonyms are different collocations containing this verb as grammatical support. There are no homonyms. Therefore, most of the seventeen senses of the verb *gutéera* should have been treated in the respective article of the verbs which are the bases of the collocation. For example, in *Gutéera*, -ye VI, its meaning is expressed according to the part of the body or the object used to hit or to stab. Consequently, as a collocative verb, it may have been treated in the articles of the noun that combine with it in collocation, because they are the semantic keys through which the language user obtains the derived phrasemes. The problem of treating *Gutéera* in a homonymic way is evidenced by the fact that some similar meaning is found in different homonyms or a given meaning is repeated in the treatment of different homonyms. For instance, the following senses are quite close so that they should not be considered as different meaning units, although the meaning

defined in the second subcomment of the first homonym could be seen as an exception.

**gutéera, -ye I 1° to throw, to fling, to throw sth to sb 2° to plant, to sow**

**gutéera, -ye II to hit or stab sb with a sharp object (lance, knife, injection)**

**gutéera, -ye VI to kick, to punch sb**

Actually, both *Gutéera* I 1°, *Gutéera* II and *Gutéera* VI are used to convey the same meaning, what changes is the types of instrument used to hit and the fact that that object used is thrown away, held or is part of the person. But these semantic precisions are not characteristics of the verb *gutéera*, rather, they are intrinsic semantic properties of the respective nouns which collocate with *gutéera*. For example, when using a lance (*icúmu*), one can throw it up or maintain it in your hand if the target is very close. This explains why the same collocation *gutéera icúmu* is used in both *Gutéera* II° and *Gutéera* II., e.g.:

Rutwé **yamuteeye icúmu** undí alízibukiira (Rutwé flinged a lance to him and he dodged it).

**Gutéera** umuuntu **icúmu** (to hit sb with a lance)

However, when aiming a kick at somebody (*gutéera* VI), one uses a part of his/her body and the part in question remains inseparable from him/her and cannot be thrown. On the other side, in the first homonym (*gutéera, -ye* I), one cannot see the closeness between the first (1°) sub-comment on semantics and the second (2°) sub-comment on semantics which has an agricultural sense (1° to throw, to fling, to throw something to somebody vs 2° to plant, to sow).

These abstract combinatorial meanings are differently expressed according to the words which collocate with the verb. Thus, the lexico-semantic content of this verb can be described in terms of collocations. This concerns not only the 17 articles presented above, but also those expressions constructed with any of the 81 words mentioned in the text block on page 323. The number of homonyms would have been reduced to the minimum. However, would a treatment of *gutéera* as polysemous lexical item not have been better? Where there are no semantic differences the collocations included in a single subcomment on semantics could have illustrated different uses of the word. This dictionary has opted for an approach of splitting instead of lumping. Of course, in a dictionary meant for text reception for example, the splitting approach can be justified when there are semantic differences between different collocational uses of the verb. The different options in the representation of collocations in the Kinyarwanda dictionary with respect to lexicographic functions are discussed in

Chapter 5. However, this dictionary contains no explicit indication of its function(s), although one could regard it as a polyfunctional dictionary with text reception and text translation as its dominant functions. In fact, there is no monofunctional dictionary in the lexicography of Kinyarwanda.

- ***Dictionnaire Rwanda-Rwanda / Rwanda-Français***

This dictionary is seen as the mother of the previous dictionary, although it was released twenty years after the dictionary of Father Irénée Jacob. As indicated above, the dictionary published by Jacob was adapted from the lexical database of the INRS dictionary. This database has been enriched and used to compile the *Dictionnaire Rwanda-Rwanda / Rwanda-Français*, published for the first time in 2005. The preparation of the second edition is at an advanced stage at the Languages Department of the Institute of Scientific and Technological Research of Rwanda. Both Belgian and Rwandan scholars have collaborated on this project over a period of 40 years. It is a hybrid dictionary: a monolingual dictionary (Kin-Kin) with a bilingual dimension (Kin-Fr). Because it is the most comprehensive dictionary of Kinyarwanda (2895 pages in 3 volumes) it is also called DIFO (*Dictionnaire Fondamental*) and will be referred to as such in the following discussions.

Like all Kinyarwanda dictionaries, the frame structure of the DIFO displays a front matter and a central list. The back matter section is never used in Kinyarwanda general language dictionaries. The front matter contains the front page, the preface, the foreword, the list of abbreviations and acronyms used in the dictionary, the list of special extra-alphabetical signs, 19 pages of introduction presenting among other things the access structure of the dictionary, and finally, the bibliography. Unfortunately, this introduction is not only complex for lay users but it also does not reflect the hybrid dimension of the dictionary because it is exclusively monolingual French. Moreover, it only appears in the first volume and this compels the user to refer to it every time he/she comes up against a problem when using the second or third volume. Regarding the central list, the macrostructure of the DIFO is alphabetically arranged according to word roots. It contains 25000 articles and about 50000 lexical units coordinated by a cross-reference system. Complex articles are divided into sub-articles coordinated by the main lemmata. The microstructure of a DIFO article is divided into different paragraphs or sub-comments. Each paragraph containing the treatment of a specific sense is marked by a square bullet followed by a letter of the alphabet, indicating the order of the sense (instead of numbers). The sub-comment on semantics contains a paraphrase of the sense in Kinyarwanda, followed by its equivalent (often a translation) in French, itself followed by illustrative sentences in Kinyarwanda with their translation in French.

As regards the inclusion policy for phrasemes in DIFO, the “locutions” (idioms and collocations) are mentioned in the articles of all of the content or lexical words it contains for being either described or cross-referenced to the lemma under which it is described. In principle, the idiom is described in the microstructure of the constituent which is semantically “restricted” (Coupez, Kamanzi, Bizimana, Sematama, Rwabukumba & Ntazinda et al. 2005: xxiv-xxv). This set criterion excludes the verbal constituent as idiom host just because it is likely to be highly polysemic and unstable due to inflection. Verbal idioms are listed under nominal, adverbial and pronominal constituents (V+N, V+Adv, V+Pr), that are often placed in the second position. However, it is not clear as to which criteria have guided lexicographers when determining the most restricted constituent in binominal idioms (N+con+N/N+Pron+N) where both nouns are equally semantically restricted in principle, or in the idioms in which all constituents are verbs (e.g. V+Pron: *Kwíikora hó:* lit. to touch on o.s., to bring trouble upon oneself; V+V: *gutééra ukíiyikiriza:* to have nobody to assist you; V+V+V: *guhenera uúzi guhéengeeza:* to harm somebody bearing you a grudge).

This practice overlooks the communicational and syntactic-semantic role of the verb within the idiom, which determines how to use/integrate the idiom in the sentence. In addition, it hinders a clear and user-friendly lexicographic presentation of idioms in as much as it does not lend itself to the alphabetical ordering of idioms due to the position of the guiding constituent. Moreover, a constituent that has been “restricted” does not reveal much about the semantic weight of that constituent in the whole idiom. The other component may be semantically the most dominant. Furthermore, for an average user of the dictionary, the concept of “*restricted constituent*” is somehow misleading and challenging because it requires the use of extensive semantic knowledge that may fail him/her in the end. Finally, if the verb qualifies for sublemmatisation of idioms with only verbs as content constituents, why is it not used in recording the remainder of verbal idioms?

As emphasised in section 3.3.2.1, the practice of the semantic listing of idioms in the article of the heavy elements poses many problems as lexicographers rely strongly on their own intuition when determining the corresponding meaning of the idioms while some of them are opaque. As Mulhall (2010: 1355) has observed, “identifying the ‘most important’ or ‘semantically heaviest’ element in an idiom is a largely subjective and complex decision and is practically impossible for those with a non-compositional meaning due to the imposed figurative sense on each lexical component.” In addition, putting idioms and collocations into one category strengthens this theoretical bias because, due to their compositionality, collocations exhibit regular correspondence between their forms and the respective

meanings of the lemma that Kinyarwanda lexicographers generalise to both types of phrasemes, despite the well-known lexicographic principle of treating different types of data in different data categories. Collocations are not lexical items while idioms are multiword lexical items. Of course, collocations need to be recorded in the dictionary because, although compositional and transparent in most cases, they are not predictable. Moreover, the importance of collocations in a text production perspective of a general language dictionary – in an LSP text production dictionary, collocations may be equally important – and the role of lexicographic data discrimination in facilitating access to data presented in the dictionary (cf. Bergenholtz & Gouws 2010; Nielsen 2008) cannot be overemphasised here.

The inclusion of idioms in the DIFO is complex in another way. Most idioms are included in the description of non-idiomatic meanings, or in other words, distributed among relevant senses of the main lemma. Obviously, those idioms are treated as microstructural data. The attempt to relate idioms to paraphrases of the lemma is questionable provided the non-compositionality of idioms and their lexical status. They cannot be put in a dependency position to the lemma paraphrase since they are lexical units in their own right. A limited number of idioms that include non-word or single-distribution lexical items functioning only in those idioms are treated individually. Surprisingly, despite the lack of semantic content for the second component, it is lemmatised and receives the idioms being described. Individual treatment is also a solution where lexicographers have failed to link a totally opaque idiom to a non-idiomatic sense of the main lemma. In this case, the idiom is presented as a sense of the main lemma and is listed and marked as directly attached to the main lemma. Nevertheless, a meticulous analysis of some lemmata appearing in many phrasemes (*umutíma*: heart, *umutwé*: head, *ugutwí*: ear, *igití*: tree, *umwuúka*: air, *inká*: cow, *umuuntu*: human being, etc.) reveals that the distribution of idioms to presumably corresponding senses is inconsistent and often done in an arbitrary way.

For example, the “locutions” (idioms and collocations) are generally linked to non-idiomatic senses of the main lemma along with co-text entries of relevant senses, differing from other co-text data (paraphrase, translation, illustrative sentences) in that each “locution” has its own paragraph following the preceding entry of the same subcomment on semantics and is presented as a treatment unit in a procedure of non-lemmatic addressing. However, contrary to the rest of Kinyarwanda dictionaries in which idioms are attached to paraphrases without a distinguishing non-typographic structural indicator, each idiom in the DIFO (2005) is introduced by the non-typographical structural marker “●” followed by the typographical structural indicator *Loc* (abbreviation of French term locution = idiom). This ensures to a small extent a quick inner-access to idioms. Therefore, contrary to Jacob’s dictionary

which lists the phrasemes in the same text slot of the last sub-comment and which does not separate the microstructural subcomments, each subcomment on semantics and each phraseme constitute a separate paragraph in the DIFO (2005). However, as it is stated in the introduction of the DIFO (2005: xii), the idioms are not presented as sublemmata; rather, they are presented article-internally as microstructural items. For example in the following article from DIFO (2005), the phraseme Gukóreesha ikigáango (Loc Gukóreesha ---) is presented within the comment on semantics.

- GÁANGO 7,8 ■ Imbáragá nyíinshi cg ubuníni buvaanzé nō gukomera.** *Grande force musculaire, robustesse.* Yabóonye abahínzi bafitέ ibigáango aseezerera abó baari básaanganywe b'íngwaágwaane kugira ngo ubuhíinge bwé bwihúute. *Ayant trouvé des cultivateurs qui ont une plus --, il a congédié les habituels qui étaient faibles, afin que ses champs soient cultivés rapidement.* | < -BARAGA 10.
- Loc **Gukóreesha ---:** Gushyíra hó agahato. *User de violence, au propre et au figuré.* | < -BÁKWE 5; -HATO 12.

The phrasemes or “locutions” that display no relation with the senses of the lemma taken as their guiding element are presented as “idiomatic senses” of the lemma so chosen, where they stand themselves as representing non-lemmatic addresses. In some articles however, the phrasemes which would have been placed near relevant subcomments on semantics on account of semantic relatedness, are isolated from them and treated similarly as ‘idiomatic sense’ of the lemma within a polysemous article, because of the lexicographer’s failure to relate them to the relevant senses. They are not presented as sublemmata at all. This presentation isolates the phrasemes from the senses of the lemma and to a certain extent brings them to a position of salience but does not elevate them to the level of macrostructural presentation as either main lemmata or sublemmata. These non-lemmatic addressing procedures correspond to what Hausmann refer to as “integration”, i.e. integration of phrasemes into the polysemic structure of the article and “annexation”, i.e. attachment of the phrasemes to the polysemic structure of the lemma (Hausmann 1990: 63). Therefore, whether phrasemes are treated in the vicinity of semantic subcomments or listed as a separate sense of the lemma chosen as the guiding element, they enjoy a treatment unit status even though they are not treated as sublemmata because they are still presented within the microstructure of a non-phraseological lemma. In addition, whereas the “integration” of phrasemes is used as the default presentation of phrasemes, the decision regarding the use of “annexation” is taken not on account of the lexicographer’s endeavour to ensure an easy access to phrasemes but further to the lexicographer’s failure to match the phrasemes with relevant senses.

Nevertheless, what is a problem is that neither integration nor annexation addressing is used in a

consistent way. In some articles, phrasemes are placed alongside the main lemma's non-idiomatic senses while in other articles they are given an individual status or treated as idiomatic senses of the main lemma. The same article (*-sóni*) studied in Jacob's dictionary has been used to explicate this inconsistence.

**-SÓNI 10iisó ■ A Ikimwaaro gikora úmt cg ubwiígaye umt yiyúumva mó hári ikt kibí cg igikorwá kidakwiyyé yakozé kikamenyeekana.** Honte, humiliation que ressent celui dont une mauvaise action est dénoncée ou dont le ridicule apparaît. Baamutaangaarije mu ruhamé kó yasáambanije akáana k ágakoóbwa nóone yagize -- zó kuja ahagáragara. On lui a imputé publiquement le viol d'une petite fille, si bien qu'il a -- de se montrer en public. | Syn -bebyá 10 | < -GAY- C.

■ B Ubwiýubaahé bushiingíye ku bwitoozzi cg ku kwáanga umugayo cg buvaanzé n ákt k úubwóoba buuzá mu umt bugatuma yíigeengeseera ntaábe yáatiinyuuka gukóra cg kuvúga ibt ibi n íibi bidakwiyyé cg kujya aht aha n aáha. Pudeur, réserve, timidité liée à la modestie, à la retenue, etc. Mukáamabano yabáaye inkumi árisko usaanga afite -- iyo áhoobérana n ábahúungu. M. est devenu une grande fille mais elle éprouve toujours de la -- quand elle embrasse les garçons. Yagize -- zó kwiicara hamwé n ábaándi kubéera kó yaambáye inshabari. Il a eu -- de s'asseoir avec les autres parce qu'il est vêtu de haillons. | Pv 1.937 -- zikuriisha amarozi. La -- te fait avaler le poison. **Compr** la timidité cause parfois des malheurs. | Pv 3.493 Umukwé wá -- ahera muu mfúruká. Un gendre timide reste dans le coin de la maison. **Compr** la timidité prive de certaines jouissances. | < -ÓOBA 14 | > -GANIR-; HAASÍ E loc; -HOOMBEKEER-; -HÚ 11 A loc; -HÚRIRÁ 7 II; -JÚNUNJAMÉ 7; -KÓONJ- I B; -MÉEMEETEREZ- A; -SUSÚ 3; -TÍNY- C; -ÚUBIKANGÓHE 1/11.

■ C Ugutíinya gushiingíye ku bwuúbahe cg ubwóoba umt agiríra abatégetsi bugatuma yáanakora n íbyaámugwa náabí. Crainte révérencielle. Nagize -- zó kweegeera báriýa batégetsi mbóna kó baabá báfite ibyó biivúganira ubwaábo. J'ai éprouvé la -- d'aborder ces hautes personnalités car je pensais qu'elles avaient à traiter entre elles. | Pv 1.050 Igituúntu cy íbukurú kiriirwa --. La tuberculose chez les grands se mange par --. **Compr** mieux vaut risquer la maladie que la colère d'un grand. | < -ÓOBA 14.

■ D Inshúunda z íikáamba zipfuka ámáaso. Franges de la couronne royale -káamba 5. | < -CÚUNDA 11; -KÁAMBA 5.

■ E Loc Guhísha cg kubíika -- : Kuryáama ugasiinziira. Se coucher et dormir. | < -RYÁAM-.

• Loc Gukóra ibitéeye -- : Gusáambana cg gukóra ibt bigayítse nk ifbyo. Commettre l'adultère ou d'autres actes impudiques. | < -SÁAMBAN-.

• Loc Gukórwa ná -- a) Kuderera, kugira ipfúnwe, gufátwa n íkimwaaro. Avoir honte; être humilié. | < -SHWAARW-. b) Gutaangaazwa cg kubábazwa nó kubóna ibt bítageneze úkó byaagoombága kumera. Être indigné, être saisi d'indignation, s'indigner de. | < -BÁBAR-; -TAANGAAR-.

• Loc Gukórwa n íiz áamóoko : Gusekeerwa n ábt mu ruhamé. Subir la risée publique. | < -JEGERER-.

• Loc Gukóza -- : Kumwaariza umt mu ruhame úgáragaza ububí bwé cg ibibí yakozé. Confondre, humilier qqn, rendre confus. | < -SÉBY-.

• Loc Gupfá -- : V syn -sóni 12 loc A Respecter.

• Loc Gushíra -- : V syn -suuzugur- B Désobéir à.

• Loc Gutá -- : Kutágira ákt k ímbebyá na búsa; kúubahuka bikabíje. Être ou devenir éhonté, insolent. | < -TÍNYUUK- A.

• Loc Gutéera -- : Kuba íkt kídakwiyyé, kígayitse, cyáatuma umt asúuzugurika. Être humiliant, confondant, inconvenant. | < -GAY-.

• Loc Inkiíngi yá -- : Inkiíngi y úurwúuririro y íburyó. Pilier situé à droite de la pièce où se trouve le lit dans la maison traditionnelle. | < -KIÍNGI 9.

• Loc Kugira -- : Kutúubáhuka gukóra ibt ibi n íibi. Avoir honte de; ne pas oser.

• Loc Kuryá -- : Gushíra -- by íkiréengá, kutágira íkt wuubáha cg utiinyá bibá hó. Être insolent à l'excès; être effronté. | < -TÍNYUUK- A.

• Loc Kuzinnya : Kutágira ímbebyá na búsa cg kudátiinyá na gáto, gukóra ibt bikwiyyé kwáamaganwa. Être éhonté; ne pas avoir honte de commettre des actes répréhensibles. | < -TÍNYUUK- A.

• Loc Uruugi rwá -- : Akuugi kabá haáfi y úurwúuririro umugení yiikiíngá mo ákaruungurukira mo ábt mu gihe aba ágítiinya. Panneau situé à proximité de l'entrée de l'alcôve du lit et à travers lequel la jeune mariée regarde les gens qui sont dehors lorsqu'elle est encore dans la période de claustration. | < -KIÍNGI 9.

The microstructure of this article contains 5 subcomments indicated by the letters **A-B-C-D-E**. The idiom provided in the sub-comment **E** is treated as a sense of the main lemma because the compiler failed to link this idiom to any of the four senses of the lemma. This is understandable because the sense that the word “*isóni*” bears in the idiom *guhísha/kubíka ~ [isóni]* (to go into bed and sleep) is not found in any of the previous senses; it is a separate sense probably meaning *eyes* but exclusively used in this sole construction, hence the lexicographers’ decision to define not the sense as such but the phraseme in which it is found. Still, this idiom is opaque because this fifth sense of the word *isóni* does not prevail in other contexts.

Individual treatment was used for the rest of the idioms and collocations presented after the sub-comment **E**. However, collocations should not be treated in the same article slot with idioms. For example, the phrasemes *gukóra ibitéeye isóni* (to commit adultery or other sexual immoralities) and *kugira isóni* (to be/feel ashamed] are not idioms; they are collocations. Therefore, they should be linked to sub-comments A and B respectively. Only the phrasemes *guhísha/kubíka isóni* (to sleep) and *gupfá agasóni* (to have respect for) can be identified as idioms and should be listed as nests if lexicographers opt to treat idioms as secondary level treatment units.

Regarding collocational support verbs such as *gutéera*, the practice of DIFO (2005:2501-2504) is not different from what has been indicated in Jacob (1985) and the same approach has been followed in all dictionaries which have been developed using the database of DIFO, including *Inkoranya y'ikinyarwaanda mu kinyarwaanda* (monolingual dictionary), first published by the IRST in 2002. The second edition (2011) is considered in the following discussion.

- **The inclusion of phrasemes in *Inkoranya y'ikinyarwaanda mu kinyarwaanda* (2011) and *Inkoranya iciriritse y'ikinyarwanda* (2010)**

In the second edition of *Inkoranya y'ikinyarwaanda mu kinyarwaanda* published in 2011, henceforth referred to as DIMO (dictionnaire monolingue/monolingual dictionary), and *Inkoranya iciriritse y'ikinyarwanda* called DIEL (dictionnaire élémentaire/Kinyarwanda elementary dictionary) published in 2010, the inclusion of phrasemes displays a more systematic treatment policy. Globally, they are listed in the microstructure of the constituent seen as the “heavier” part, like in DIFO (generally the nominal, pronoun, adjective or adverb). In DIFO, a phraseme containing a single distribution lexeme is listed in the comment on semantics of the single distribution component; that is, the non-word is

lemmatised in order to receive the idiom, without any additional semantic description. However, as in Jacob (1985), phrasemes and particularly idioms are not treated in a separate text block; rather, they are mixed with other microstructural entries. This constitutes a serious impediment to their accessibility. According to Mulhall (2010: 1355), this “diminishes the importance of idioms by placing them in a subordinate role to words, which misconstrues their equal status as units of meaning in the lexicon.” To complicate matters even further, the typographical structural marker *Isb* used in DIMO to indicate an idiom, is not bolded or highlighted so as to be visible and can thus be confused with the words prior to that. Moreover, the listing of idioms under the second constituent coupled with the distribution of idioms to different paraphrases make an overall alphabetical ordering of idioms impossible. This is worsened by the remote addressing procedure of some idioms presented as treatment units within the microstructures of derivatives functioning as non-lemmatic addresses in a dense article. Without a clear alphabetical ordering of idioms in a special search zone, their accessibility is compromised, especially when they are presented within the microstructure of a dense article. This is all the more worrisome since Kinyarwanda lexical items are highly polysemic. Consider the following article in DIMO:

***umuGABO 1 Umuuntu uwô arî wê wêese ukûze wô mu gitsîna cy'âbahuûngu.*** Iki kibiîndi cyaaterurwa n'ûmugabo w'îmirya ntîcyaaaterurwa n'ûbonetse wêese Isb ***Gutéerwahô undî: Gushobora gukôra ikiintu arî ukô usatikanyije n'ûndî muuntu.*** Isb ***Gutwâara ubugabo: Iyô bavugâ inkâ, kugeenda imbere y'iziîndi.*** Isb ***Umugabo mu mwâana: Umugabo mugufî bikabîje.*** **2 Umuuntu wô mu gitsîna cy'âbahuûngu ufite umugorê cyâangwâ wamwîigeze.** Jye n'ûmugabo waanjeje ejô tuzaaza kugusuura. Jyan ***Umugabo wâabo: Umuvâandimwê w'ûmugabo w'ûmuuntu iyô arî umuhuûngu.*** Jyan ***Umugabo wâacu: Umuvâandimwê w'ûmugabo waanjeje iyô arî umuhuûngu.*** Isb ***Mugâbo wa mâmâ: Amagaambo umuuntu ushyîkirana n'ûundi amuhamâgaza atâraakâye.*** **3 Umuuntu ukîraanura ababûuranyi akabigira avûga ukô ibyô bapfâ byaâgeenze kubêera kô yabihagazehô.** Urubaânza rwâacu ruzaaciibwa urukîiko rûmaze kûumva imvûgo y'âbagabo. Iht umuhamyâ. Isb ***Gutâanga umugabo: a) Gutâanga ibimeneetso mu rubaânza. b) Gukôra imibarê unyûze iyîndi nzira, ugêenzura kô ibyô wakozê bituungâanye.*** **4 Umuuntu bashiima ímyiifatire cyâangwâ ibikorwâ.** Isb ***Gukôra iyô bwaâbaga: Gukôra ukô ushobôye kôose.*** Isb ***Gutwâara ubugabo: Iyô bavugâ inkâ, guhôra igêenda imbere y'iziîndi.*** Isb ***Gutwâaza ubugabo: a) Kwêerekana ubutwâari. b) Kurwarwaanya.*** ■ ***imiGABO Imigaâmbi y'îngirakamaro umuuntu yiyyêmeza gukûrikirana kugeza igihe imugejêje ku kiintu yiifûuza gukôra cyâangwâ kugerahô.*** Mfite imigabo yô kûubaka inzu arîko sîinzî kô nzâabôna imigozî n'îmigaanda. Isb ***Gukûura umuuntu imigabo mu nda: Kumûtiinyiisha cyâangwâ kumugaamburuza.*** ■ ***iGABO Rb int ijabo.*** ■ ***inGABO I Umuuntu ujyâ cyâangwâ urî ku rugaâamba arwâana.*** Ingabo zitabaaye ku nkîiko, umugabâ wâazo aziraangaaje imbere. ■ ***inGABO II 1 Mu ntaambara zô haambere, igikôreesho biîkiingiraga amacûmu n'îmyaambi.*** Ingabo bayibaaza mu gihooondohondo nô mu mukô. Isb ***Guhêra mu ngabo yasâbyôombî: a) Guhâahaanwa n'âbaantu ukabura ahô weerêkera cyâangwâ kubâ mu kaâga ukabura ukô ukiîgobootoramô. b) Guhêra ku cyûzeere cyô kuroonka.*** Jyan ***Ingabo y'ûgutwî: Igicê cy'ûgutwî kimezê nk'ikibabi.*** **2 Ikinyagû baatemyê mu muziinga.**

In this article, the square bullet structural indicators marks non-lemmatic addresses (e.g. ■ *imiGABO*, ■ *inGABO I*, ■ *inGABO II*) that would be presented as a sublemmata although they are not actually presented as such because they remain in the scope of the microstructure of the lemma *umuGABO*. Since the three lexicalised derivatives are accessed via the lemma sign *umuGABO*, it is likely to be more problematic to access phrasemes here highlighted in grey background in the microstructures of these derivatives

than it is to do with phrasemes presented in black background in the lemmatic zone of the microstructure of the lemma umuGABO.

DIEL is a simplified version of DIMO. Its main peculiarity is in word-based lemmatisation of lemmata and in removing diacritic signs, i.e. tones. Apart from this, the data presentation policy remains unchanged. These two dictionaries share the specificity of being bi-functional, that is they are both meant for text reception and text production. Surprising as this may seem, the relevance of collocations and pragmatemes in text production is not recognised in these dictionaries.

From this sub-section, the following important conclusion can be drawn. Phrasemes have been recorded in dictionaries exclusively based on morpho-syntactic criteria, irrespective of the situations of dictionary consultation and the user perspective. In some dictionaries, phrasemes have been recorded in the microstructure without use of a specific structural indicator as if they are illustrative materials in the comment on semantics of the guiding element. They have not been included as fully-fledged lexical items that are legitimate lemma candidates. The access-related problems caused by the presentation of phrasemes as microstructural items should compel the lexicographers of Kinyarwanda to explore innovative alternatives or procedures for the presentation of phrasemes in general printed dictionaries of Kinyarwanda.

### **3.3.2.3 Semantic description of phrasemes in Kinyarwanda dictionaries**

Nkomo (2008: 126) stipulates that the comment on semantics contains data categories intended to help the user to understand the sense of the lemma and its proper usage comprehensively. These include definitions or paraphrases of meaning, equivalents, etymology, semantic relations, usage examples, lexicographic labels and pictorial illustrations. It is normal that the types of data categories included in the article of a dictionary depend on the functions of the dictionary. Of the data categories mentioned above, paraphrases/equivalents, cross-references and usage examples are the most ubiquitous in Kinyarwanda lexicography. However, only paraphrases of meaning and equivalents are provided in the comment on semantics of idioms. This applies also to some collocations which are treated as idioms. The following discussion focuses on idioms, bearing in mind that idioms constitute a catch-all category in Kinyarwanda dictionaries.

Regarding idioms' paraphrases of meaning in monolingual dictionaries and equivalents in bilingual dictionaries, it is assumed that there is no big difference between idioms and lexemes regarding definition formulation. What can be said, is that the paraphrase of meaning should be as descriptive as

possible if a text reception function prevails in the dictionary and as discursive as possible in the case of text production. Lexicographers should avoid using synonyms or paraphrases when defining idioms, especially in monolingual general public dictionaries with a text decoding function.

However, two significant trends in defining idioms should be stressed. The first trend, and the most compelling in general lexicographic works, views idioms as non-compositional multi-lexical items that must be described with figurative paraphrases of meaning. This approach is largely applied in Kinyarwanda dictionaries. In the DIFO for instance, each idiom is followed by a figurative explanation in Kinyarwanda and its translation in French. It is not specified whether the French definition is an equivalent or a paraphrase. In Jacob (1985), the treatment of idioms is less developed. As noted above, the idioms are used as illustrations or recorded explicitly as idioms. In both cases, they are followed by French translations.

In DIMO (Bizimana & Kayumba et al. 2011), idioms are explained in Kinyarwanda. If the idioms have more than one sense, their paraphrases of meaning are numbered with lowercase letters. For example, the idiom *gucá amaraká* (lit. to cut the pharynges) has up to four senses (Bizimana & Kayumba et al. 2011: 470): *a) to communicate bad news; b) [hunters] to look out for game/ an animal, taking distant positions one from another; c) [acrid beer, hot food, spices] to hurt sb's throat; d) to cost a great deal.* Understanding the polysemic relationships among these senses is a seemingly impossible task. I do not intend to discuss in detail the forms of paraphrases of meaning.

The second approach on the semantic description of idioms seeks to unearth the motivation behind the formation of idioms and to give an account of the etymological meaning of idioms. According to Mojela (2004: 333-334), the role of etymology in describing idioms is “the determination of the relationships between the literal and figurative meaning of the lexical items” by providing “the basic meaning of the expression, the relationship between the literal meaning of the expression and its real (i.e. figurative) meaning, and the etymological background (meaning) of the individual words in an expression”. As one can easily conclude, this approach is not practical in compiling communication-directed general public dictionaries. Rather, it is suitable for idiom dictionaries with a cognitive function. What is more, the etymological analysis of idioms may be time-consuming and cumbersome in general dictionary compilation; it may also be useless for some transparent idioms and impossible for some opaque idioms, unless historical accounts help. The compilation of Kinyarwanda idioms by Mulihano and Kayigana are to be seen in this framework (Mulihano & Mineduc 1980; Mulihano & Minesupres 1986; Kayigana 1971, 2007). The first publication (Mulihano & Mineduc 1980 and

Mulihano & Minesupres 1986) privileges the historical narration without paying attention to rhetorical and linguistic mechanisms behind the expressions. The latter focuses on history, culture, and folklore to explain the origin and the meaning of idioms. However, the explanations of idioms do not fulfil any communication function, but a cognitive function as they provide only encyclopaedic knowledge about the origin of idioms. These data are of no practical value in communication-oriented dictionaries but they can be useful in cognitive-oriented print and electronic dictionaries. Therefore, the collections of phrasemes mentioned above can be considered as cognitive-oriented lexicographic works (cf. Bergenholz & Bothma 2011: 64-65), although none of the compilers indicates the underlying function having guided his work.

To conclude this section, it can be said that the semantic description of idioms in Kinyarwanda dictionaries is quite satisfactory for decoding purposes. But the absence of syntactic properties, of illustrative items and usage labels specifying how to incorporate idioms in sentences and making explicit the stylistic particularities of idioms, coupled with the lack of grammatical data categories (lack of syntactic characterisation in the form of parts of speech), make the current treatment of idioms in Kinyarwanda dictionaries inefficient in text production tasks. It is important to know whether the subject of a verbal idiom involves a human, an artefact or an animal; it is equally useful to know the stylistic register within which a given idiom is used. In this connection, Tarp (2004a: 314-316) thinks that in addition to textual usage examples, the explicit indication of “*syntactic properties*” or “*syntactic mini-rules*” is of great importance if the dictionary is meant for text production in foreign language. Where space is not a problem, the use of as many illustrative sentences as possible is important, particularly in dictionaries with idioms as subject matter. Similarly, the indication of syntactic classes (parts of speech) is essential for the proper use of idioms in encoding tasks, especially for exocentric idioms (see in subsection 4.3.3.3).

### **3.4. Conclusion**

In this chapter existing dictionaries of Kinyarwanda were criticised with the objective of assessing the extent to which the representation of phrasemes in these dictionaries subscribes to relevant principles of lexicography. In this respect, it underscored the need to recognise the wide-ranging nature of Kinyarwanda phraseology and to devise sound lexicographic strategies for inclusion and treatment of different types of phrasemes in an explicit way. Besides defining the concept of phraseme following the MTT approach to phraseology, this chapter has above all emphasised the complexity of problems facing Kinyarwanda lexicography with respect to phraseology. Among other things, it has been made

clear that Kinyarwanda lexicography lacks a coherent lexicographically determined approach in which the target users and their needs as well as the function(s) of the dictionary, play an important and determining role. Moreover, it has underlined the lack of a lexicographically inspired policy in the treatment of phrasemes in dictionaries and the relegation of phrasemes to a position of secondary importance in the lexicography of Kinyarwanda. In addition, the idiom bias or confusion of various types of phrasemes in one category called *inshóberamáhaánga* (idioms) has been discussed. Unfortunately, this mixing of different types of phrasemes into idioms (*inshóberamáháanga*) has led to their misidentification and mistreatment in dictionaries and in other meta-lexicographic works. Due to a lack of a sound classification of phrasemes, different types of linguistic expressions are confused and treated in the same way in dictionaries. For example, the distinction between idioms, pragmatemes and collocations is recent and is not yet reflected in Kinyarwanda dictionaries. These three types of phrasemes are thus intermixed in Kinyarwanda dictionaries.

For many years, Kinyarwanda lexicography has utilised the analysis and reception standpoint, rather than the perspective of text synthesis or production. In a text reception perspective, most phraseological researchers pay too much attention to the properties of opacity or fixedness of phrasemes, excluding what is transparent or compositional from the field of phraseology. For them, phrasemes are characterized by a higher opacity and fixedness and what is not an idiom cannot be called a phraseme. This is a narrow perception of phraseology, an idiom-biased approach which confines the scope of phraseology to the sole category of idioms. Thus, because most collocations and pragmatemes are transparent and not fixed, they go unnoticed and are not included in a dictionary, unless they are merely used as illustrative data along with other free phrases. Generally, it can be concluded that Kinyarwanda dictionaries do not take full advantage of insights from theoretical lexicography as far as the representation of different types of phrasemes is concerned. Therefore, in the following chapters, the lexicographic representation model for idioms, proverbs, collocations and pragmatemes is formulated in response to the shortcomings pinpointed in this chapter.

## CHAPTER 4: A MODEL FOR LEXICOGRAPHIC REPRESENTATION OF KINYARWANDA IDIOMS AND PROVERBS

### 4.1 Introduction

Not only is the lexicon composed of free morphemes, simple and compounded words but it also includes lexical strings more complex than words, such as idioms (Rey & Chantreau 1989). Idioms are as essential as simple words and compounds, and nobody can learn and use a language irrespective of them. Therefore, a lexicon-based lexicography must account for the lexical status of idioms and address the issue of their lemmatisation, when determining the inclusion policy of different lemma candidates.

Mojela has pointed out that the interest of lexicographic and lexicological research on idioms and proverbs lies in their possible lemmatisation and inclusion in lexicographic works (Mojela 2002: 332-333). Over the last three decades, researchers in lexicography have assessed the strengths and weaknesses of either lemmatisation or sublemmatisation of idioms (Gouws 1991, 1996, 2007, 2010; Louw 2006; Mojela 2002; Mulhall 2010; Vrbinc & Vrbinc 2011; *inter alia*). In accordance with this research the objective of the present chapter is to devise and propose a theoretically motivated model for the lexicographic treatment of Kinyarwanda idioms and proverbs, utilising the theory of functional lexicography as well as insights from Wiegand's general theory of lexicography. Reference is made to the metalexicographic work of scholars working in these theoretical frameworks. Furthermore, some linguistic references are also mentioned when relevant linguistic issues pertaining to the identification, classification, presentation and use of Kinyarwanda idioms are discussed.

The rest of the chapter is structured as follows. In section 4.2 the working definitions of the concepts of idiom and proverb are provided. This is followed by a proposal in section 4.3 of a function-driven model for the presentation and treatment of idioms in Kinyarwanda dictionaries. The challenges of the lexicographic treatment of idioms within the Kinyarwanda context are discussed and different alternatives according to the function of the planned dictionary are proposed. Then in section 4.4 the inclusion of proverbs in a dictionary is dealt with. Different theoretical proposals for the inclusion and treatment of proverbs in a dictionary are illustrated and their user-friendliness evaluated. Finally, section 4.4 contains some concluding remarks.

### 4.2 The concepts of idiom and proverb

#### 4.2.1 Definition of idiom

Following a particular line of thought found in the MTT, an idiom is defined as a lexical non-

compositional phraseme or phraseological expression (Mel'čuk 2011). The notion of semantic compositionality implies that the sense of a lexical string A+B is the sum of the individual meanings of A and B, and non-compositionality means the opposite, although the constituent words may play a role in decoding the meaning of an idiom (Mel'čuk: In press; Vega-Moreno 2001). Therefore, idioms are phraseological expressions violating the principle of compositionality. The non-compositionality appears to play a decisive role in characterising idioms, no matter whether the latter are transparent or not. The idiom is a pre-constructed phrase and is taken as such. Whereas the user unrestrictedly constructs its meaning, the selection of the lexical items to construct the expression to verbalize it, does not depend upon the language user, but rather, it is restrictedly made. As discussed in Chapter 3 and 5, idioms differ from collocations by the fact that one of the components of a collocation is unrestrictedly selected for its meaning, while the other one is chosen to express a given meaning contingent to the former, hence their compositionality. In the case of an idiom, however, its sense is different from the sense of each of its individual constituents as understood on its own. In other words, it is not the sum of the senses of its constituents. For instance, the senses of the idioms *to kick the bucket* (to die) or *number one* (urination) have no compositional relation to the free-combination senses of its lexical components, respectively *kick* and *bucket* and *number* and *one*. Therefore, idioms have to be learnt as such, hence their importance in dictionaries.

In the context of Kinyarwanda, an idiom has two properties: it appears as a pre-constructed syntagm and has a proper sense that does not result from the calculation of meanings of its components. Being structurally pre-constructed does not imply that idioms are either binary as collocations or discursive as proverbs. Rather, it entails that items making an idiom have to appear intact in any sentence in which it is used and that insertion of another extraneous element is impossible in most cases, although the order and morphology of words in idioms may change further to morpho-semantic and morpho-syntactic operations (passivization, flexional morpho-syntactic operations, clefting, etc.). For a limited number of verbal idioms (V+N), the nominal constituents have no autonomy in the Kinyarwanda lexicon, that is “words that occur only as components of multiword lexical items” (Gouws 1991: 75). These idioms are “non-word idioms” (Mulhall 2010: 1358). Though these idioms have this particularity of including unique distribution lexical items, apart from this, they are in no way different from other full idioms.

Unlike some proverbs, idioms do not have literal and figurative meanings at the same time. What people usually call an idiom’s figurative meaning (as if there is a literal meaning) should be regarded as the only primary or literal meaning of the idiom, because the compositional meaning is not a lexical

meaning, but a textual one having nothing to do with the idiom, but the free phrase involving the same constituent words as the idiom.

#### **4.2.2 Difference between idioms and proverbs**

Scholarly attempts to distinguish proverbs from idioms have resulted in highbrow and sometimes contradicting ideas. For example, Munyakazi (1992: 231-241) attempts to distinguish proverbs from idioms through a set of contrastive criteria that he believes could help in distinguishing the two phraseological phenomena. According to this scholar, the idiom and the proverb oppose one another by the following pairs of distinctive features:

- The content of a proverb aims at providing an everlasting moral lesson; no such intention is understood in an idiom.
- A proverb contains collective, normative and universal wisdom; the meaning of the idiom is general, language-specific and not universal.
- A proverb has a double interpretation: literal meaning and figurative or symbolic meaning; an idiom has a non-compositional sense and its compositional meaning is not relevant.
- A proverb has an unchanging structure which does not allow formal modification and is always inserted in the discourse as a complete piece of information; an idiom is a ready-made expression allowing various syntactic modifications without changing the meaning. An idiom is always used in a broader sentence and is not syntactically independent.
- A proverb usually has a dual structure in which the first part exposes a factual context and the latter the consequence or advice; an idiom has a linear structure although with many lexical constituents. It does not have a binary or dual structure.

On the other side, Mel'čuk (In press) claims that opaque proverbs constitute a sub-set of full idioms because both forms are characterized by semantic irregularity, i.e. the impossibility to equate their meaning from their constituents. For other phraseologists, however, the distinction between the idiom and the proverb lies in the fact that the proverb is a complete message because it can work independently as a sentence, while the idiom cannot work as a full sentence, but must be incorporated in a sentence as a part of it. However, some idioms have a finite form or sentence-like form (propositional idioms or speech formulae). That is why Dobrovolskij and Filipenko (2007: 721) suggest differentiating idioms from proverbs not on the syntactic level, but according to semiotic,

semantic and pragmatic parameters. They emphasise that unlike speech formulae, a proverb expresses a universal truth relating to a “shared knowledge about rules governing social behaviour and human coexistence”. They further affirm that the semantics of proverbs have the “illocutionary force of recommendation” in that “they provide a moral support for an argument or action by reference to a generalised proposition.” Finally, they say that proverbs have a discursive autonomy contrary to sentence idioms that are often embedded in a broader discourse (Dobrovolskij & Filipenko 2007: 721). The proverb itself constitutes a meaningful and independent discourse irrespective of its inclusion in a more inclusive discourse. Along the same line, Crépeau and Bizimana (1979: 6) point out that the proverb provide a rule of life by ordering a given behaviour or formulating an interdiction, by recommending the preferential or the exemplary social behaviour and by ridiculing a deviant behaviour.

Thus, it turns out that the idiomaticity, i.e. the semantic discrepancy between the regular meaning and the real meaning (others say figurative meaning) is the sole shared phraseological characteristic between idioms and proverbs while the normative aspect is peculiar to proverbs. In addition, whereas the literal meaning of the idiom may sound uncommon, both the literal and the figurative meanings of the proverb are true and the figurative meaning derives from the literal one by metaphorical or metonymic analogy. Therefore, due to their discursive structure and pragmatic orientation, proverbs will be discussed as a separate category of phrasemes although according to the MTT, a proverb can be an idiom. Lexicographically, it is more profitable on the side of the dictionary user to have proverbs included as a separate and well-delineated data category rather than to mix them with other types of data.

When discussing characteristics of the proverb, Dobrovolskij and Filipenko (2007) establish a difference between proverbs and speech formulae and relate the latter to idioms. In our case, these speech formulae are in no way idioms; their finite form status relate them to proverbial phrases called pithy sayings and it is appropriate to treat them in the back matter together with proverbs. Therefore, without wanting to undermine the linguistic and literary classifications of paremiological utterances, in the following discussion the term proverb is used in a broader sense encompassing also pithy sayings provided these phrases can have a similar lexicographic treatment.

## 4.3 Lexicographic representation of idioms in Kinyarwanda dictionaries: a transformative account

### 4.3.1 Working assumptions

The status of idioms and other phrasemes has long been, and remains a polemic issue in both lexicography and metalexicography. In this regard, the salient question is not to know whether they should be covered in the dictionary. That is understood. Rather, it is about how to do it in a way to cater for the specific needs of the potential users of the dictionary and in keeping with its respective function. To put it differently, should idioms and other phrasemes be listed as primary or secondary level treatment units, and how should this be done in a consistent manner without undermining the user's needs. However, including idioms in a dictionary is one thing, describing them is another. Therefore, it is necessary, if only for an instant, to look into the challenges facing users with respect to lexicographic data assigned to idioms in various Kinyarwanda dictionaries and how to improve their treatment.

With respect to the presentation of idioms within the microstructure, Vrbinc and Vrbinc (2011: 257) identified and enumerated four different models of microstructural treatment of idioms.

- Listing of idioms at the end of the article within a special section devised for idioms: this model is close to the macrostructural listing as secondary level treatment units.
- Listing idioms in definitional zones together with other senses of the main lemma, based on frequency: use of an idiom as a paraphrase of the main lemma (direct lemmatic-addressing).
- Listing idioms at the end of the article without a structural indicator (a case of non-integrated and remote addressing) or for more than one idiom, in a section labelled idioms or phrases where they are sub-divided into groups according to the relation between the idioms and different polysemes of the main lemma: a kind of semi-integrated microstructural treatment.
- Idioms are used as example phrases to illustrate a metaphorical sense of one of the senses of the lemma: this is a strict microstructural treatment of idioms where idioms are integrated in the article in question in a remote lemmatic addressing relation with the lemma that it illustrates.

The first three are referred to as ‘direct treatment methods’ while the last is being considered as an ‘indirect approach’ (Vrbinc & Vrbinc 2011: 257). The drawback of microstructural listing of idioms is that it does not allow a comprehensive treatment of idioms, albeit that they are lexical items in the same way as single word lexical units. This constitutes an impediment to data accessibility and retrieval of information needed by users (Gouws 2007: 63). The rationale for the presentation strategy of idioms

that is proposed in this research is based on the following two postulates. First, the lemmatisation strategy to apply and the extent of data to provide for idioms in future dictionaries should be determined by users' needs and lexicographic functions prevailing in the dictionary in question. Secondly, idioms are according to Gouws, lexical items in their own right and not entries addressed to another lemma sign (Gouws 2010a: 55). In accordance with the proposal formulated by Gouws in the same publication (2010a), idioms should be lemmatised in the macrostructure of a dictionary. In the Kinyarwanda dictionaries, this can be achieved using the first content word as reduced guiding element for inclusion of an idiom.

#### **4.3.2 Idioms and lexicographic functions**

In view of the first principles mentioned above, the selection of idioms to be included in a dictionary and their presentation should be done according to dictionary functions. Two types of functions can be envisaged with respect to Kinyarwanda idioms, i.e. a communication-oriented function and a cognitive function. As indicated earlier, the range of communicative activities during which users can seek dictionary assistance includes text production and correction in L1 or L2, text reception in L1 or L2 and text translation from L1 into L2 or from L2 into L1. Text production and text reception tasks may be oral or written. Although simultaneous interpretation does not allow the interpreter to consult a dictionary, the interpretation of pre-recorded texts leaves room for dictionary use and could also be seen as involving communicative lexicographic functions. In this section, I exclusively look at text reception and text production in the mother tongue. Thus, the scope of application of this model is the general language monolingual dictionaries. Regarding cognitive functions, two options are envisaged according to whether one is dealing with a general language dictionary or with a dictionary of idioms or of multiword expressions. They are meant to provide users with assistance in learning idiomatic proficiency (form and meaning) and cultural as well as encyclopaedic knowledge (origin, history, rhetoric figures, etc.).

##### **4.3.2.1 Idioms in a dictionary with a text reception function**

Text reception related needs arise in a situation of written text decoding and oral conversation where the user meets an idiom of which he/she does not understand the meaning and therefore consults a dictionary to seek help. When negotiating the inclusion of idioms in a dictionary tailored for text production as the main or only function, lexicographers should bear in mind the descriptive nature of such a dictionary or part of that dictionary.

The descriptive principle in lexicography is discussed by Bergenoltz (2003: 71) and Bergenoltz and Gouws (2010: 34) who say that “descriptive lexicography refers to the use of data from a linguistic survey and a text investigation in practical lexicographic work” as empirical basis. They point out a wide range of possibilities for the empirical basis of a dictionary including (a) introspection, (b) analysis of a linguistic survey, (c) involvement of descriptions in existing dictionaries, grammars, monographs, articles, etc., (d) analysis of a number of examples randomly chosen from random texts (corresponding with the practice of dictionary making before the age of computers), (e) analysis of a specifically constructed text corpus, and (f) analysis of usage found in texts in the examined language in all available websites on the internet (Bergenoltz & Gouws 2010: 34).

The descriptive approach takes an empirical perspective on lexicographic description, endeavoring to present and describe different variants without judging or recommending which one should be the best or appropriate form, neither does it prohibit any of the presented variants (Tarp & Gouws 2008: 238). It leaves users to judge for themselves. Bergenoltz and Gouws (2010: 39) go further by saying that an open, total, strong and explicit description is the best approach to use when negotiating reception in general language dictionaries.

In a monofunctional reception dictionary, the typical destination of dictionary consultation is the item giving the definition of the lemma and of the idiom in our case. This concern should be reflected in the selection of candidate items for a reception dictionary. Due to the singularity of the needs and the type of data needed to satisfy the need, i.e. the paraphrase of meaning of a given lemma, the types of data which do not contribute largely to the attainment of the reception function (e.g. collocations, illustrative sentences) should be eschewed so “to guarantee a rapid access that is not impeded by unnecessary data or that leads to information stress or even information death” (Bergenoltz & Gouws 2010: 41). When consulting a dictionary in order to satisfy text-decoding needs, the ideal thing would be for the user to find only what has prompted his consultation of the dictionary and nothing else. In this case, only data presenting the lemma, the paraphrase of meaning of the lemma, the non-compositional idioms and other multiword expressions and their definitions will be enough and appropriate to ensure a successful consultation and to assist users to resolve text reception problems.

The profile of users also plays a significant role in data selection. Although it is true that a reception dictionary for adult mother-tongue speakers should also focus on infrequent, archaic, polysemic and idiomatic or multiword lexical items, lexicographers should consider it as a necessity also to lemmatize frequent words, especially when non-native speakers or young learners are concerned (Bergenoltz &

Gouws 2010). In either case, the lexicographers have to formulate with caution their assumptions regarding the native and non-native speakers' proficiency in L1 before determining the extent to which frequent words will be included or not. For example, in Rwanda where millions of people grew up in exile with no opportunity to use their mother tongue some words which are regarded as frequently used by certain people may be unknown to others.

The empirical selection of lemma candidates for a text reception dictionary is thus informed by the lexicographic function, the (socio-) linguistic profiles of specific users and their needs. Nevertheless, the function-driven selection of treatment units, their explicit description and the types of users are not the sole factors of rapid access to data in a reception dictionary; the data should be structured in a way such that they are accessed in an unimpeded way by both the skilled and the less skilled users. For example, it is understood that the structure and micro-architecture of articles in a reception dictionary will be non-extended in contrast to the articles of text production dictionaries. However, the types of data to be included in the dictionary and the extent to which they have to be described, reveal much less about the outer and inner distribution of data or idioms for one to be specific with regard to the article structure. Before looking at the organisation of idioms in a dictionary frame structure, it is equally important to look at the requirements of text production and cognitive functions with respect to data selection.

#### **4.3.2.2 Idioms in a dictionary with a text production function**

Lexicographic needs related to production of texts arise when the user wants to ascertain the meaning, the register, the morphology and the syntactic use of an idiom he/she already knows in order to use it appropriately in a sentence or to validate its use in a sentence when correcting texts. Its needs are summarized in ascertaining the appropriateness of a given word or phraseme to a specific context. Therefore, a text production dictionary does not put emphasis on comprehensively defining the lemma; the minimum definition of a word is enough. The lexicographers' endeavour should rather be to include in each dictionary article, all items from which users will get text production assistance. Bergenholz and Gouws (2010: 43-47) indicate among other items that should be negotiated in a text production dictionary e.g. the lemma sign or treatment unit, pronunciation data (very important if the dictionary has spoken text production as a function or if it targets foreign language speakers), grammatical data, a minimum paraphrase of meaning of the lemma or a translation equivalent in bilingual dictionaries, one or more example sentences, one or more collocations, cross-referenced synonyms and antonyms, idioms and other fixed expressions provided with items indicating their definitions and typical use and pragmatic or stylistic guidance on the use of words. How these types of data assist users in their text

production activities is not discussed in this work. However, it is important that the selection of items and the scope of treatment assigned to these items should be in keeping with the needs of the intended users. These decisions should also be influenced by the envisaged descriptive, prescriptive or proscriptive approach of the dictionary (Bergenholtz & Gouws 2010).

The descriptive approach in lexicography has been discussed in connection with text reception dictionaries and should be used in a text production dictionary provided only one variant is concerned. However, where the need to make a choice between two or more variants arises, lexicographers should envisage prescription and proscription as viable options. When prescription is applied, the lexicographers may envisage the following possibilities (Bergenholtz 2003: 74; Bergenholtz & Gouws 2010: 35):

- (a) a specific linguistic variant is explicitly prohibited,
- (b) one or more linguistic variants are explicitly prescribed, thus prohibiting all other non-mentioned variants, and
- (c) a specific linguistic variant is explicitly prescribed (as opposed to prescription (b), this involves a new word, new spelling, new pronunciation, new inflection or neologism).

The feature characterising the prescriptive approach is explicit prohibition or implicit prohibition by omitting a given variant and the explicit prescription of one or more variants or a neological form. Unlike the descriptive approach which objectively explains on the empirical basis how a language element is actually used in the variety of its forms, the prescriptive approach indicates how the user should use it by prescribing or prohibiting (Bergenholtz 2003: 68; Tarp & Gouws 2008: 237). Because language policies are in the nature of things prescriptive, it is a foregone conclusion that prescriptive lexicography is the domain in which language policy-related ideologies come to the fore. A state-of-the-art and in-depth discussion of the interpenetrations between lexicography and language policy issues is provided by Bergenholtz and Gouws (2006) and Tarp and Gouws (2008) and will not be reproduced in this chapter.

In the proscriptive principle, the lexicographers present data resulting from the empirical basis and show all variants while recommending the use of only one of them (Bergenholtz & Gouws 2010: 35). Therefore, proscription seems to be appropriate for text production purposes because it displays all variants and makes a recommendation allowing the user to make an informed choice. In a prescriptive approach, the user can be discouraged if he/she consults the dictionary and does not find the form he/she was looking up because it has been prohibited by omission. In a more helpful way in a proscription approach, such a variant is shown with a cross-reference leading the user to the

recommended variant or form. In this way, even often misspellings of words or wrong plural forms should be included in the dictionary for the sake of helping users who could use them as dictionary consultation strings to reach the right forms. Various pragmatic indications play a decisive role in text production dictionaries with a prescriptive background (Bergenholtz & Gouws 2010).

#### **4.3.2.3 Idioms in a dictionary with a cognitive function**

With respect to idioms it is assumed that the most knowledge-related needs concern the motivation behind the formation of an idiom or the need for knowing the relation between the form and the content. This motivation can be found in the culture, in the history or in the rhetoric procedures involved in the formation of idioms. In brief, it is the lexicographer's commitment to assist users to know more about idioms, which is at the basis of cognitive functions. Dictionaries with a cognitive function are not yet discussed in relation to description, prescription and proscription approaches (Bergenholtz & Gouws 2010: 49). It can only be assumed that from the encyclopaedic point of view, description is the most helpful principle in conveying knowledge-related data, because it gives a thorough account of the state of things in the world.

The above discussion regarding the implication of descriptive, prescriptive and proscriptive principles in catering for lexicographic functions has been an oversimplification of these concepts because, for example, description is not an absolute concept; there is rather different forms of description depending on the scope of the empirical basis. Thus, this discussion has deliberately eschewed discussing various forms of these principles, hoping that this will not undermine in any way their being relevant to the functional perspective adopted in this work.

#### **4.3.3 Towards the lemmatisation of idioms in Kinyarwanda dictionaries**

Kinyarwanda lexicography has been dominated by the traditional word-based approach by lemmatising only words and handling phraseology data as marginal elements treated in the microstructure together with other co-textual items. Therefore, although idioms have long been recognised as lexical units in the same way as mono-lexemic lexical items (Gouws 1996) their lemmatisation in many dictionaries remains problematic, due to the word bias (Gouws 1996) still prevailing in many dictionaries and the contemplative approach adopted by many lexicographers who replicate the traditional user-unfriendly models. As far as the presentation of Kinyarwanda idioms as macrostructural treatment units is concerned, two options are being envisaged: sublemmatisation of idioms (4.3.3.1) and straight lemmatisation of idioms (4.3.3.2). The sublemmatisation of idioms consists in presenting them as vertically-arranged secondary level treatment units while straight macrostructural listing of idioms

consists in presenting them as primary level treatment units, just like simple lexical items.

#### **4.3.3.1 Sublemmatisation of idioms**

For idioms, not only is sublemmatisation seen as a space-saving strategy, but for some lexicographers, it is also seen as the most direct way of representing idioms so to render them quickly accessible to average dictionary users (Vrbinc & Vrbinc 2011: 257). Sublemmatisation of idioms is seen as an inclusion strategy whereby instead of listing idioms as microstructural entries or as first level treatment units, the lexicographer lists and describes them as secondary level treatment units attached to the article of “the lemma sign functioning as reduced guiding element of the article (...),” resulting in a macrostructural slot representing a lemma clustering (Gouws 2010a: 55). Various sublemmatisation strategies have been identified so far. For example, Vrbinc and Vrbinc (2011: 251) show that in most English monolingual dictionaries, the ordering policy of idioms varies according to lexicographers and the level of fixedness of the idiom’s constituents. In most dictionaries, idioms are included and treated in the article of the first content word of the idiom. However, in the case of varying idioms where the first content word may be replaced by another without undermining the idiomaticity and meaning of the string, the idiom being presented is included in the article of the following content fixed word. The alternative words are written and separated by slashes where the idiom is presented. For idioms containing “delexical” verbs such as *to be*, *to take*, *to have* etc. and adjectives in pre-position, some dictionaries direct the user to the following content noun. Grammatical idioms containing non-lexical words are included in the article of the first word of the idiom. However, some dictionaries consistently follow the traditional approach of including the idiom in the article of the first noun it contains. A theoretical discussion regarding the inclusion of idioms as sublemmata is a relatively recent activity in metalexicography. It cannot be traced back beyond the advent and development of the concept of sinuous lemma files in metalexicography.

- **The concept of a sinuous lemma file**

Gouws (2003: 40) stipulates that the alphabetical ordering of lemmata does not imply that all lemmata are homogeneously presented. As far as lemma presentation is concerned, he distinguishes two schemas, i.e. the presentation of lexical items as main lemmata in the vertically alphabetical ordering of the macrostructure (straight alphabetical ordering or vertical ordering) or as sublemmata in the horizontal ordering of macrostructure elements (Gouws 2002: 134). The second procedure differs from the straight vertical ordering in that it leads to “a sinuous lemma file, resulting from the inclusion of sublemmata, ordered horizontally in lemma clusters” (Gouws 2003: 40), although a vertical presentation of sublemmata in the form of a grouped or clustered ordering of nested lemmata is also

feasible (Gouws 2010a: 57). A short observation of Kinyarwanda dictionaries shows that in most cases lexicographers conform to a strict vertically ordered listing of lemma candidates, supplemented by the horizontal listing of niched derivatives and the presentation of idioms in nests in the microstructure.

- **The concept of niching and nesting**

Taking his cue from Hausmann and Wiegand (1989: 336), Gouws (2002: 134-135) defines the two inclusion strategies at work in sublemmatisation, i.e. niching and nesting when writing that “niching is a strict alphabetical clustering of lemmata or articles which may or may not be semantically related, whereas nesting is a clustering of lemmata or articles which displays a deviation from the rules of strict alphabetical ordering.” In a later publication, Gouws (2003: 40) explains further the functions of niching as a macrostructural device as follows:

Niched lemmata adhere to a straight alphabetical ordering with respect to both the horizontal and the vertical ordering, i.e. the internal and the external ordering. The lemmata entered within the niche display an internal alphabetical ordering and they also precede the next vertically ordered main lemma alphabetically. This type of cluster merely illustrates a deviation in the direction of macrostructure, i.e. horizontal instead of vertical, but does not imply any deviation from the prevailing straight alphabetical ordering

Needless to say, a sinuous lemma file (niching and nesting) is a suitable textual condensation strategy in the context of derivatives, nominal compounds and other multiword items (Gouws 2002, 2003, 2007; Louw 2006). As one can read in the introduction of *Inkoranya y'ikinyarwaanda mu kinyarwaanda* (Monolingual Kinyarwanda Dictionary, DIMO), the dictionary article may be either simple or complex. A complex article is made of a main article followed by one or more sub-articles treating derivatives of the main lemmatised noun, adjective or pronoun (Bizimana & Kayumba et al. 2011). Clustering of compound lemmata is not applied in Kinyarwanda. Compounds are written conjunctively and hence deserve individual articles. Thus, as far as stem-based lemmatisation is concerned, niching of derivatives is the most used textual condensation strategy in Kinyarwanda dictionaries. In the following example taken from DIMO, the strict alphabetical ordering is based on stem-lemmatisation (uppercase) even though prefixes (in lowercase) and endings of the word are also written. Conversely, were the dictionary alphabetical ordering be based on the word-lemmatisation, the derivatives in this article would display the characteristics of a nest instead of a niche.

***kuGABIZA (...)***

***umuGABO 1 Umuuntu uwô arî wê wêese ukûze wô mu gitsîna cy'âbahuûngu.*** Iki kibiîndi cyaaterurwa n'ûmugabo w'imirya ntîcyaaterurwa n'ûbonetse wêese. Isb ***Gutêerwahô undî: Gushobora gukôra ikiintu arî ukô ufatîkanyije n'ûundî muuntu.*** Isb ***Gutwâara ubugabo: Iyô***

bavugâ inkâ, kugeenda imbere y'îziîndi. Isb **Umugabo mu mwâana**: Umugabo mugufî bikabîje. 2 **Umuuntu wô mu gitsîna cy'âbahuûngu ufite umugorê cyâangwâ wamwîigeze**. Jye n'ûmugabo waanjye ejô tuzaaza kugusuura. Jyan **Umugabo wâabo**: Umuvâandimwê w'ûmugabo w'ûmuuntu iyô arî umuhuûngu. Jyan **Umugabo wâacu**: Umuvâandimwê w'ûmugabo waanjye iyô arî umuhuûngu. Jyan **Umugabo wâanyu**: Umuvâandimwê w'umugabo waawe iyô arî umuhuûngu. Isb **Mugâbo wa mâmâ**: Amagaambo umuuntu ushyîkirana n'ûundi amuhamâgaza atâraakâye. 3 **Umuuntu ukîraanura ababûuranyi akabigira avûga ukô ibyô bapfâ byaâgeenze kubêera kô yabîhagazehô**. Urubaânza rwâacu ruzaaciibwa urukîko rûmaze kûumva imvûgo y'âbagabo. Iht umuhamyâ. Isb **Gutâanga umugabo**: a) Gutâanga ibimenyeetso mu rubaânza. b) Gukôra imibarê unyûze iyîndi nzira, ugêenzuura kô ibyô wakozê bituungâanye. 4 **Umuuntu bashiima ímyïfatire cyâangwâ ibikorwâ**. Isb **Gukôra iyô bwaâbaga**: Gukôra ukô ushobôye kôose. Isb **Gutwâara ubugabo**: Iyô bavugâ inkâ, guhôra igêenda imbere y'îziîndi. Isb **Gutwâaza ubugabo**: a) Kwêerekana ubutwâari. b) Kurwarwaanya. ■ **imiGABO Imigaâmbi y'îngirakamaro** umuuntu yiyyêmeza gukûrikirana kugeza igihe imugejêje ku kiintu yiifûuza gukôra cyâangwâ kugerahô. Mfite imigabo yô kûubaka inzu arîko sîinzî kô nzâabôna imigozî n'îmigaanda. Isb **Gukûura umuuntu imigabo mu nda**: Kumûtiinyiisha cyâangwâ kumugaamburuza. ■ **iGABO Rb imt ijabo**. ■ **inGABO I Umuuntu ujyâ cyâangwâ urî ku rugaâmba arwâana**. Ingabo zitabaaye ku nkîiko, umugabâ wâazo aziraangaaje imbere. ■ **inGABO II 1 Mu ntaambara zô haambere, igikôreesho biîkiingiraga amacûmu n'îmyaambi**. Ingabo bayibaaza mu gihondohoondo nô mu mukô. Isb **Guhêra mu ngabo yasâbyôombî**: a) Guhâahaanwa n'âbaantu ukabura ahô weerêkera cyâangwâ kubâ mu kaâga ukabura ukô ukiîgobootoramô. b) Guhêra ku cyiîzeere cyô kuroonka. Jyan **Ingabo y'ûgutwî**: Igicê cy'ûgutwî kimezê nk'ikibabi. 2 **Ikinyagû baatemyê mu muziingga**. ■ **muGÂBO Shy Umuuntu ukôreeshwa imirimo ya rubaanda rwa giseesêka**. Baa mûgabo bamwê batwaarê igitaka bâgishyira mu muhaânda abaândi bagisaânze, nahô abafundi buubâke. ■ **nGABO Ubwôoko bw'îgitî gikomôoka murî Amerikâ gisâ na ngabo-itsiinze**. ■ **kiGABO** cyâangwâ **ruGABOUmr**. Ku buryô bukwiiyê, butuungâanye. Kugena kigabo. ■ **ubuGABO**. Ijaambo ryô kûubaha rivugâ igitsîna cy'ûmuhuûngu. Baafâtanye bararwaana Gicâ amukaamuura ubugabo. Iht ubuhuûngu. ■ **GABO Ibz iny Rb umukôma 2 jyan; igitsîna jyan 3**.

### ***nGABO-ITSÎINZE (...)***

In this article, the lemma umuGABO contains nine annexed articles of its derivates preceded by the structural markers ■. They are annexed to the last subcomment on semantics of the lemma umugabo. Each derivative is described in one or more definitions eventually followed by phrasemes (see the phraseme indicators *Isb* and *Jyan*). In the articles of some derivatives, cross-reference is made to variants (■ *iGABO Rb imt ijabo* means *See variant ijabo*) or to a relevant collocation base, for instance in the last niche where **GABO** as collocate is cross-referred to collocations with **umukôma 2** and **igitsîna 3** as bases respectively. With such a dense article, the user will undoubtedly encounter problems in gaining access to data presented in dictionary. The idioms presented in the ‘sub-microstructures’ of annexed derivatives presented as non-lemmatic addresses within the overall microstructure of the article of the lemma taken as the guiding element (i.e. umuGABO) have been

underlined, in order to highlight problems the user may experience when struggling to access them. If the access to idioms presented as integrated or annexed tertiary level treatment units pose problems, then those presented as fourth level treatment units within the microstructure of a tertiary level treatment unit functioning as non-lemmatic address within the overall microstructure of the lemma taken as the guiding element, raise even more concerns. It is obvious that access to those idioms is made more difficult because the user first has to locate the guiding element or the overall lemma sign, then the derivative within a congested article before gaining access to the idiom. Within such a dense article this would not be easy.

A similar approach to lexicographic representation of derivatives has been used in the *Dictionnaire Rwanda-Rwanda et Rwanda-Français* or DIFO. Thus, in the DIFO, even the most skilled user will have to be patient to gain access to an idiom presented as a third or fourth level treatment unit. By way of example, consider the modal idiom *bubí na bwízá* (at all costs). This idiom is presented as a third level treatment unit in the article of the opening word *bubí* which functions as a single distribution modal, i.e. it is used everywhere but in this idiom. However, the modal *bubí* (class 14) is the second to last in a cluster of seven articles of derivatives attached to the end of the article of the adjective *-bi* (bad). Spatially, *bubí* (modal cl.14) is located about one and a half page down from the main lemma *-bi* (adjective) which is the overall guiding element. Again, once the right derivative is identified, the user has to locate the idiom within the microstructure of the derivative in question. Such an access road is not cost-effective unless the user is well trained in linguistics. Because *bubí* occurs exclusively in the idiom *bubí na bwízá* and nowhere else it would be better to lemmatize it as a primary level treatment unit. On the other hand, no empirical study has been conducted in support of the alleged benefits of the systematic niching or nesting of derivatives for users of Kinyarwanda dictionaries in their efforts to gain access to data presented in the dictionary. For instance, when the user needs data on the lemma *inGABOI* or *inGABOII*, it is not certain that this lemma will necessarily evoke in the user's mind the lemma *umuGABO* (in at least one of its various senses) used as guiding element in the above-mentioned article. It is likely that he/she may never think about it. In fact, the two lemmata refer to two quite distinct entities or referents so much so that the first one does not automatically remind the user of the second and vice versa.

It does not suffice to explain a lemmatisation procedure in the front matter of the dictionary; it should also be brought within everyone's reach, which is impossible when linguistic pre-analysis is a prerequisite to a successful dictionary consultation. In order to spare the user a fruitless dictionary

consultation, the idioms in text production dictionaries should be moved to the straight macrostructural level when the opening word is a single distribution derivative. In text reception dictionaries, they should be recorded as secondary level treatment units attached to the article of the appropriate guiding element. This is due to the fact that in text reception perspective the user may fail to recognise a word combination as idiom or may feel that the uncommon single distribution word is the cause of his/her failure to decode the idiom. In text production, the user knows the idiom and wants to verify its sense or its adequacy in a given context. Thus, presenting idioms as primary level treatment units in a text production dictionary will allow the compiler to provide as many data categories as required in text production.

Therefore, sublemmatisation of derivatives as either niches or nests should not be negotiated irrespective of the idioms access-related constraints that such a lemmatisation strategy entails. In this respect, the recommendation formulated by Gouws regarding these strategies is more pertinent than elsewhere. Scrutinizing how sublemmatisation is used in different dictionaries, the scholar advises that niching and nesting as textual condensation strategies should be applied with caution, mainly due to the risk of violating the basic principle of alphabetical ordering. This caution especially applies when bringing together morpho-semantically related items, leading to a hybrid ordering system (alphabetical and morpho-semantic ordering) in one dictionary (Gouws 2002: 136). Further, he concludes that niching as a space-saving strategy should not be used “at the cost of the users in their attempt to gain access to the data presented in the dictionary” (Gouws 2002: 156). In other words, niching and nesting should not be viewed merely as a space-saving strategy; they have also to be seen as lemmatisation strategies in their own right. Therefore, when making decisions regarding lexicographic representation of idioms the lexicographers should endeavour to use various sublemmatisation procedures in a way consistent with the lexicographic function(s) assigned to the intended dictionary and take care of the needs of the specific group of users in different situations of dictionary consultation.

- **Niching idioms in a dictionary for text reception**

Taking the notion of a sinuous lemma file a step further, Gouws suggests that sublemmatisation, and precise nesting should be used to include idioms in a dictionary as secondary level nested treatment units (Gouws 2007, 2010, 2010a). In this approach, the lexicographers should list idioms as secondary level treatment units within a text block attached to the article of the lemma representing the guiding element of idioms, generally the first content word in the idiom. The only difference between nesting

and the straight vertical macrostructural listing of idioms lies in the possible horizontal listing and in the deviation from the straight alphabetical ordering because the guiding element of the internal ordering is different from the reduced guiding element of sublemmatisation. Obviously, the option of nesting is motivated by the particular structure of some languages in which the word in the first position of idioms are neither content nor stable words, and by the long-established practice of treating the idiom in the article of the so-called heavy or core word which often is not the word in the first position (Louw 2006). Theoretically, this is a viable approach, because phasing out idioms from the microstructure to the macrostructural list could result in increased visibility and improved and comprehensive treatment, in comparison with the microstructural recording that diminishes and undermines the lexical status of idioms. For example, Gouws (2007: 62) states that:

By including idioms in a separate text block, attached to or included in the article of a lemma representing a keyword taken from the idiom, the idiom does get the status of a macrostructural treatment unit, i.e. a horizontally ordered sublemma. (...) Idioms are sublemmata because they can only be accessed via the entrance lemma...

Gouws reiterated this suggestion saying that “by attaching a text block with these nested articles to the article of the vertically ordered lemma the fixed expressions get a position of more salience that will enhance the access possibilities” (Gouws 2010a: 56).

In Kinyarwanda, despite the traditional conception of semantically heavy constituents, the question of determining the word that should be taken as the guiding element does not crop up, because all idioms have a content word in the first position. Thus, it could be suggested to use the word in the first position as guiding element for the idiom, instead of using an approach that compels users to proceed to pre-consultation semantic analysis of the words in the idioms. Many users, even the most skilled ones, may not have such knowledge. It has often been stressed that this practice complicates the accessibility to data for the user and prevents the lexicographer giving a comprehensive treatment of idioms. In the sublemmatisation model, idioms should be presented in alphabetical arrangement according to the first and following constituents. In this view, lemmatised verbs will host verbal idioms; lemmatised adverbs will host adverbial idioms while lemmatised nouns will host nominal idioms. The idioms as data category will be placed at the end of the article of the main lemma.

Notwithstanding the foregoing, even in the case where the word in the first position or the opening word could be used, it should be noted that the word in the first position could differ slightly from the word taken as default main lemma. The verb may contain lexicalised infixes and suffixes. The case of

idioms containing verbs with fixed infixes or suffixes raises additional concerns when the verb used as guiding element of the idiom is the default form. For example, if the verb in the first position of the idiom has additional lexical suffixes, a skilled user can still relate its form to the default verb taken as the main lemma of the article to which the text block containing the idioms is attached. Thus he/she can easily find the article in question and jump to the idiom text block. However, relating a verb with a lexicalised reflexive or object infix to the default verb can be much more complicated due to subsequent morpho-phonological or inflectional changes in the orthography of the verb as it is used in the idiom. The verb appearing in such an idiom is often in single distribution. Access to idioms can be impeded if the words taken as the guiding elements of those idioms do not occur in the exact forms in those idioms. The following is an excerpt of the article of the lemma *kwīishyira* in the DIMO.

**kwīISHYIRa Ibz Isb Kwīishyira heezâ:** Kwiivuga nēezâ cyāangwâ kwīiboneza ku buryō utagarāgarahô ikibî. Isb **Kwīishyira ukīizana:** Gukôra ibiintu ntâa mususû, ntâa cyo wiikâanga. Isb **Kwīishyirahô:** Kwīiyemeza ku baândi kubêera ubwfiraasi. Iht kwīishyuungika. (...).

**Translation:** [KWISHYIRA /kwīishyira/ v Def **Idm kwishyira heza:** to praise o.s. **Idm Kwishyira ukizana:** to act freely. **Idm Kwishyiraho:** to be unapproachable]

For example, in the idioms *kwīishyira heezá*, *kwīishyira ukīzana*, *kwīishyirahó*, the verb *kwīishyira* is a reflexive form of the verb *gushyíra* (to put sth somewhere). Recording these idioms in a horizontally arranged text block attached to the lemma sign representing the verb *gushyíra* could make it difficult for many users to access those idioms because it will compel them to operate a transitional morphological reanalysis of *kwīishyira* into *gushyíra* in the pre-consultation phase. In the article reproduced above, there is no semantic treatment beside the typographic abbreviation (Ibz, Imburabuzi: defective) given as the sole comment on the form.

Because the verb *kwīishyira* (a lexicalised reflexive form of *gushyíra*) is restricted to idiomatic use and occurs only in these idioms, it should be lemmatised as guiding element of the dictionary article hosting the above-mentioned idioms as sublemmata. In a dictionary designed for text-decoding needs, this presentation could assist the users who may undertake dictionary consultation using the opening word of the idiom as search string following their failure to recognise the word group as idiom or due to low reference skills. Finding the opening word of the idiom in the dictionary will undoubtedly guide the user through an appropriate access structure to the ultimate destination of dictionary consultation, i.e. the idiom in question. In such a case, the treatment of the main lemma could not go beyond items providing its grammatical category and indicating that it is a form that is restricted in its occurrence to

be part of an idiom.

In the article with the rudimentary treatment the form, here *kwíshyira*, is not the real lemma but a lemma part functioning as reduced guiding element in the article entrance position. The idioms become the real lemmata and actual treatment units. In addition, this presentation has the advantage of maintaining the straight alphabetical ordering in both vertical and horizontal directions. This is an application of *niching* to the lemmatisation of idioms, made possible by the use of the word in the first position as guiding element for both outer and inner alphabetical ordering of idioms. The lexicographic presentation should be of such a nature that the user could directly jump to the required idiom in the vertical list if he/she does not go through the guiding element, here *kwíshyira*. In this regard, the idioms should be presented as vertically-ordered clustered sublemmata within the text block with each idiom displaying its own horizontal treatment while maintaining the internal vertical structure of the niche, instead of presenting them in a horizontally-ordered niche. An appropriate non-typographic structural indicator identifying a cluster of idiomatic lemmata (e.g. the Kinyarwanda abbreviation ISB) should precede the text slot. Not only will the use of structural indicators beyond the microstructural level quickly direct the user to the vertically-ordered idioms' text slot in the macrostructure, but it will also prevent them from wasting time searching a non-idiomatic lemma in that text block. In other words, extending the use of structural indicators to macrostructural level will spare the users from reading the text block when the subject of consultation is not an idiom. Similar proposals have been outlined in Gouws work (2010: 56-58) in the framework of the macrostructural presentation of fixed expressions as nested sublemmata. If applied consistently, this representation policy can help the user to find the idioms as sublemmata attached to the article of the relevant verb – as indicated in the example presented below.

In the following section, I have adapted the above-mentioned article to show what the lemmatisation of idioms would look like in a dictionary having text reception as the main function. As one can see, the verb *kwishyira /kwíshyira/* is maintained as a guiding element of the rudimentary article to which the idiom text block is attached. Then the idioms are alphabetically arranged with the special text block preceded by the non-typographic structural indicator ISB. However, marking out the top delimitation of the idiom text block is not enough; the user should be able to know how far below a given textual slot extends. Thus, idiomatic lemmata are presented as indented items in comparison with the normal text in order to help the user at first glance to mark out the boundaries of the text block from the top to the bottom and easily navigate across the macrostructural structure of the dictionary. While shifting the text

to the right the cleared white space functions as an implicit typographic structural indicator, assisting the user to interpret with fewer efforts the outer access structure of the dictionary, as far as the idiomatic zone is concerned.

### **KWISHYIRA /kwíshyira/ sh Ibz**

**ISB: KWISHYIRA HEZAIsb/sh Kwivuga neza cyangwa kwibonekeza ku buryo utagaragaraho ikibi.**

**KWISHYIRA UKIZANAIsb/sh Gukora ibantu nta mususu, nta cyo wikanga.**

**KWISHYIRAHOIsb/sh Kwiyemera ku bandi kubera ubwirasi.**

**KWISHYIRAMOIsb/sh 1 Gufata umuntu cyangwa ikintu nabi cyangwa neza ntushobore guhindura. 2 Kwanga umuntu n'yo yakora neza ntibikunyure. 3 Gukurikirana ikintu kugeza ubwo ukigezeho.**

**KWISHYITSAs** Kwihwanya kugira ngo ugushe neza umuntu.

In a text reception dictionary, the treatment of idioms should focus on essential items that are required to assist the user resolving text reception related problems, i.e. the paraphrases of meaning. Examples are not needed unless they are used to make an abstract sense or a grammatical sense more understandable. Similarly, there is no need to indicate semantic relations. The indication of pronunciation should be provided only where the phonological transcription differs from the common rendition of the lemma and participates in distinguishing two words having a similar spelling in the common writing system. In all cases, pronunciation is not essential in a text reception dictionary; it plays a role only when it allows distinguishing between words in adjacent positions that have similar writing and different pronunciation and meaning (e.g. *ikirere /ikireere/*: banana tree leaf versus *ikirere /ikiréeré/*: space, universe).

However, surprising as this may seem, a Kinyarwanda idiom may also be composed of one graphic word when the second component has become part of the idiomatic verb in the form of integrated proclitics or enclitics. The straight vertical alphabetical ordering of idioms cannot prevail when the expansion or modifications affects the left side of the stem of the verb. Therefore, although niching is a viable option for the presentation of many idioms, it cannot accommodate idioms with verbs displaying morphological deformations in comparison with the default verb, e.g. the idiom *kutîrûibwahô* attached to the article with the lemma sign *kwîiryâ* as guiding element in the DIMO:

**kwÎIRYA, -îriye 1 (...). 2 (...). Isb *Kutîrûibwahô*: Kudâshobôkera uâbonetse wêese. Isb *Kwîiryâ ukûmara*: Kuruha wêenynî udâfîte undî ukûungaanira. Isb *Kwîiryâ icyâara*: Kwîishiima hâri icyô wiizêeye kuroonka.**

Of the three idioms which are attached to the above-mentioned article with the label Isb as structural indicators, one of them, i.e. *kutîrîibwahô*, deviates from the straight alphabetical ordering even when the stem-based arrangement is taken into account, whilst *kwîirya ukiimara* and *kwîirya icyâara* do not. Whereas niching prevails in the article displaying adherence to alphabetical ordering, the above-outlined article violates the alphabetical principle and the dictionary compilers have opted for nesting, at the cost of the user in his effort to gain access to data presented in the article. Indeed, in order to access an idiom displaying dissimilarities in comparison with the default lemma taken as reduced guiding element, the user must undertake a pre-consultation phase that does not always end successfully, due to complex morpho-phonological changes that the idiom may have undergone and the possible failure of the user to relate it to the lemma taken as guiding element of the idiom. Therefore, when dictionary compilers decide to implement nesting in order to represent idioms as in the above-mentioned article, the access-related problems that confront the user are often a matter of morphological reanalysis entailed by formal discrepancies between nested idioms and the guiding element of the article to which the nest is attached. This re-analysis is undertaken in the pre-consultation phase and supposes a certain level of linguistic knowledge in order to lead the user to an accurate conclusion.

Therefore, although nesting is a viable approach where niching does not function, it does not ensure an unproblematic access to idiomatic data. In a dictionary with text reception as unique function the idioms displaying a deviation to alphabetical ordering, here *kutîrîibwahô*, could be moved to the macrostructure to constitute an independent primary level treatment unit. A similar presentation should be assigned to the exocentric and grammatical idioms in which the word in the first position is a rudimentary word formed through deletion of the initial morpheme of the default lexical items or a word formed through conversion or functional/syntactic shift, cf. Balteiro (2007: 15), e.g. *bufíuni na buhoró*: unfairly, *epfó na rûguru*: in an awkward position, *haasi heejuru*: on equal terms, *patá na ruúgi*: inseparable, *mbará na mbariro*: for long, *kabiíndi na buúki*: definitely, *mpirú na nyoni*: definitely, *curí na mbeba*: deeply, *mu méenyo ya rûbaámba*: in awkward position, *mabyí na nnyo*: equally bad, *mutí wa mperezayó*: definitely, *bubí na bwíizá*: at all costs, whether you (they) like it or not, willy-nilly, *kubéera kó*: because, *kugira ngo*: in order, *ibirí amaámbu*: on the contrary, etc.).

These noun-based, prepositional or verbal idioms function as opaque modifying collocates in collocations. In the sentence, they fulfil the syntactic functions that are not characteristic of their grammatical categories but of other grammatical categories, namely adverb, conjunction and adjective.

Therefore, they are the subject of dictionary consultation for both text reception and text production needs. Their being lemmatised as first level treatment units will not undermine their accessibility because knowledge of the default form of the opening word or any other constituent of the idiom is not required for the idiom look up to be successful. In addition, apart from the fact that such knowledge is useless to dictionary consultation, it may also be lacking among many other users. Furthermore, these idioms are syntactically stable, i.e. the constituents of the idiom do not admit a variability or an insertion of free words and they always appear as united sequence in a text. This will result in some types of idioms being treated as niched sublemmata while other are elevated to the macrostructure as primary level treatment units in the same dictionary.

It is true that niching represents by far the most user-friendly textual condensation strategy in terms of idiom presentation. However, niching is not an appropriate lexicographic representation of idioms in text production dictionaries where the user knows the idiomatic nature of the multiword expression and primarily focuses on the idiom as indivisible search string during the consultation process. Therefore, the sublemmatisation of the idiom is not deemed necessary in a dictionary meant for text production. In addition, in text production dictionaries the lemmatisation of single distribution verbs with a one-way cross-reference to idioms are mere nice-to-haves because those idioms could be lemmatised as main entries without going through a rudimentary article of a word without independent lexical existence.

In view of the preceding shortcomings resulting from the use of inappropriate lexicographic procedures in the representation of idioms in a Kinyarwanda dictionary designed for text production use, the straight vertical lemmatisation of idioms as primary treatment units is worthy to be explored.

#### **4.3.3.2 Straight lemmatisation of idioms in a text production dictionary**

The main issue at stake when negotiating the lemmatisation of phraseology items remains whether the dictionary user achieves the purpose of his dictionary consultation in a successful and cost-effective manner. The lemmatisation of idioms as first level treatment units is not new in Bantu lexicography. This approach has been successfully applied in revising a Shona dictionary, *Duramazwi Guru ReChishona* (Chimhundu 2001) as a way to respond to users' needs (Chabata 2002; Mpofu 2002, 2007). Moreover, multiword items are lemmatised in Kinyarwanda LSP dictionaries without problems, e.g. Kabagema (2011). It is assumed that a similar approach could be applied without problems with the view of making the treatment of idioms in Kinyarwanda text production dictionaries more user-driven.

The following example shows a group of articles taken from the DIMO, where the middle article is a single distribution lemma (note the mark **Ibz** as abbreviated form of the term **Imbúrabuzi**, i.e. defective word or non-word after the single distribution lemma kwÎISHYIRa) containing as comment on semantics only the idioms preceded by the structural marker **Isb** in Kinyarwanda. In this dictionary, although lemma signs are written in full form, they are alphabetically arranged according to the first letter of the stem.

**kwÎISHYIINGIRa**, ye **Kwîijyaana ku mugabo by'ûmukoôbwa udâtaanzwê n'îwaâbo akaroongorwa ku buryô budahuujê n'ûmucô karaânde w'ubukwê.** Kô iwaâcu baânze kô unsabâ nzâakwiîshyiringirehô? Iht gucikira 5; kwîijyaana 2.

**kwÎISHYIRa**, ze Ibz Isb **Kwîishyira heezâ**: *Kwîivuga nêezâ cyâangwâ kwîiboneza ku buryô utagarâgarahô ikibî.* Isb **Kwîishyira ukûzana**: *Gukôra ibiintu ntâa mususû, ntâa cyo wiikâanga.* Isb **Kwîishyirahô**: *Kwîiyemeza ku baândi kubêera ubwîiraasi.* Iht **kwîishyuungika**. Isb **Kwîishyiramô**: a) *Gufâta umuuntu cyâangwâ ikiintu nâabî ntûishobore guhiindura.* b) *Kwâanga umuuntu n'îiyô yaakorâ nêezâ ntibûkunyure.* c) *Gukûrikirana ikiintu kugeza ubwô ukigezêhô.*

**kwÎISHYIIRa**, riye **Gukurura ikiintu ucyîiganiishahô**.

Below is the adapted article illustrating what the alphabetical arrangement and treatment of the idioms would look like in a text production dictionary. The lemma *Kwîishyira* has been abandoned as guiding element of the article and the four idioms that were attached to the article have been elevated to the level of primary treatment units. Because of alphabetical ordering, the last lemma sign *kwÎISHYIIRa* has moved from the last position to the second position after *kwÎISHYIINGIRa* and before the first idiom.

**KWISHYINGIRA** /kwîishyiringira/ SH Cy0 Iyo umukobwa yishyingiye yijyana ku mugabo adatanzwe n'iwabo akarongorwa ku buryo budahuje n'umuco karande w'ubukwe. Ko iwacu banze ko unsaba nzakwishyingireho? Rb Iht gucikira 5; kwijyana 2.

**KWISHYIRA** /kwîishyiira/ SH Iyo umuntu yishyira akurura ikintu acyiganishaho. Ayo meza yaterure wishyira haboneke inzira.

**KWISHYIRA HEZA** /kwîishyira heezâ/ ISB SH Cy0 Iyo umuntu yishyira heza yivuga neza cyangwa akibonekeza ku buryo atagaragaraho ikibi. Umuntu uhora yishyira heza ntibamukunda.

**KWISHYIRA UKIZANA** /kwîishyira ukûzana/ ISB SH Cy0 Iyo wishyira ukizana ukora ibintu nta mususu, nta cyo wikanga. Amahoro atuma abaturage bishyira bakizana. Rb Iht kwigenga

**KWISHYIRAHÓ** /kwîishyirahó/ ISB SH Cy0 Kwishyiraho ni ukwiyemera ku bandi kubera ubwirasi. Umva ko wishyizeho, wasimbuka inzu ntuvunike. Rb Iht kwirata; *kwishyungika*.

**KWISHYIRAMÓ** /kwîishyiramó/ ISB SH Cy1 1 Kwishyiramo ni ugufata umuntu cyangwa ikintu nabi cyangwa neza ntushobore guhindura. Wishyizemo ko nkwanga kandi ndengana. Rb Ibs kwikuramo. 2 Kwishyiramo ni ukwanga umuntu n'ijo yakora neza ntibikunyure. Kwishyiramo umuntu si byiza. 3 Kwishyiramo ni ugukurikirana ikintu kugeza ubwo

## **ukigezeho.**

In the above proposed presentation, the focus is on a hypothetical text production dictionary. As shown in these proposed articles, many more items are needed to provide users with relevant information when proceeding with text production, be it in written or spoken discourse. In the following paragraphs some categories of data that should be provided in a text-production dictionary as a way of treatment of idioms as well as other categories of lemma are discussed.

- **Provision of items giving pronunciation**

The phonological transcription supports oral text production but it has a lesser value for written text encoding. These transcriptions should however be kept, unless the dictionary compilers exclude spoken texts from the scope of the dictionary. In addition, when two words with a similar orthography but different pronunciations are lemmatised in adjacent positions, the user can refer to the pronunciation data to differentiate between them. On the other hand, where the pronunciation of idioms is predictable being that of the citation form of each component, idioms should be listed without indication of pronunciation. However, it should be provided when one of the elements of the idiom is a single distribution form that is given nowhere else in the dictionary, as in the case of *Kwiishyira* in the articles of the four idioms provided above. Apart from the pronunciation, the provision of grammatical data is very important.

- **Provision of grammatical and stylistic labels**

For the idioms in this instance, the abbreviated structural indicators ISB (standing for the term meaning idiom in Kinyarwanda) followed by SH (standing for the term for verb in Kinyarwanda), are highlighted to show the grammatical category or the syntactic function of the idiom and indicating that they are verbal idioms. In Kinyarwanda, most idioms play either a verbal or a nominal function. In these idioms, the part of speech is the same with that of the lexeme functioning as head of the idiom. However, there are idioms whose parts of speech cannot be deduced from the parts of speech of lexemes functioning as their heads. Their parts of speech are determined by the role they play in the sentence, irrespective of the parts of speech of the lexemes functioning as syntactic heads. The issue of identifying the parts of speech in such idioms is discussed in sub-section 4.3.3.3.

Besides items indicating the parts of speech of idioms, the lexicographers should also provide items giving different style levels where necessary, in order to ensure an optimal retrieval of the information by the user. Different language registers have been identified for Kinyarwanda expressions. The

lexicographers could use any of the following abbreviated labels (in brackets) to indicate the style or language level of the addressed idiom, provided such abbreviation is explained in the list of abbreviations in the front matter, e.g.:

- Neutral style: Imvúgo isáanzwe (Szwe)
- Formal/polite style: Imvúgo yiiiýbashye (Ubah) / ikinyábupfúra (pfura)
- Technical/specialised style: imvúgo ya gíhaánga (Ghg)
- Figurative style: imvúgo y'íkigereranyo (Ger)
- Pejorative style: Imvúgo ipfóobyá (Pfoby)
- Standard style: imvúgo gikwiirá (Kwir)
- Argotic style: imvúgo z'úturéeré (Rere)
- Insult: gutúkana (Tuk)
- Obscene style: imvúgo nyandagazi (Nyand)
- Colloquial: ikimenyáne (Meny)
- Vulgar: Nyámusózi (Nys)
- Slang: imvúgo y'úruféfeko (Fef)
- Neologism: gishyá (Shy)
- Borrowed: gitíirano (Tir)

The difference between some of these stylistic indicators is not clear and is not discussed here. What we have to admit is that stylistic data categories are not provided for in monolingual Kinyarwanda dictionaries with the exception of the indication of neologisms, archaisms and borrowed forms in DIEL (2010) and DIMO (2011). The only dictionary indicating both the register and the special field of activity is DIFO (2005). Yet again, only colloquial, formal, insulting and vulgar styles are indicated in DIFO. With respect to idioms, not all registers mentioned above need to be indicated. The indication of language style must be done in keeping with the function of the dictionary. It is assumed that in a general monolingual dictionary with text production as the main function, apart from the neutral style which is marked by omission, only colloquial, formal, obsolete, pejorative and insulting idioms should be marked.

#### • Provision of transitivity

Another important item provides data regarding the transitivity of the addressed treatment unit. The marker Cy (abbreviation of *icyuzuzo*, complement) is used followed by a number (Cy0, Cy1, ...). The number zero means the verb/idiomatic verb is intransitive or does not require an object, while the numbers 1, 2 indicate that the verb/idiomatic verb is followed by one or two objects respectively. From the text production perspective, this is crucial because it informs the user of the syntactic configuration in which the lemma occurs.

- **Provision of paraphrase of meaningin a text-production dictionary**

Both text reception and text production dictionaries require a comprehensive treatment of the meaning of the lexical item represented by the lemma sign. However, in a monolingual text production dictionary, the lexicographer should use discursive explanations or full-sentence definitions (Rundell 2006) instead of descriptive paraphrases of meaning that are appropriate in text reception dictionaries. Using discursive lexicographic definitions or full-sentence definitions (FSDs) is by far the most natural and user-friendly way of lemma description in a monolingual text production dictionary because this kind of definition provides not only the meaning of the lexical item represented by the lemma sign; it also indicates the pragmatic context of that lexical item, here being the idiom (cf. Burkhanov 2003: 103; Rundell 2006: 325).

- **Provision of illustrative items**

In addition to discursive definitions, example sentences play a vital role in informing the user on the typical uses of the lemma and idioms. Illustrative items are supplementary items within the comment on semantics in the above-proposed articles. If an idiom is used as single-distribution collocator in a collocation, the example should consist in giving that collocation. For example, the idiom *kibúno mpá amaguru* (as fast as his legs could carry him) is only used as a speed intensifier of the basis *kwíiruka* (to run) in the collocation *kwíiruka kibúno mpá amaguru* (to run as fast as his legs could carry him). Because this idiom has a unique occurrence in this collocation, the latter should be provided as typical illustration of the idiom even if it has been treated as a collocation in the article of the headword *kwíiruka*. Furthermore, because this type of idiom cannot be defined without referring to the word it accompanies, the definition given is to a certain extent the definition of the whole collocation.

- **Use of mediostructural items**

Last but not least, the provision of antonyms and synonyms, where relevant by means of cross-references, presents many advantages in text production dictionaries. In the articles above, the indicators *Rb* (*Reba*: see) that followed the Kinyarwanda abbreviations *Iht* (synonym) or *Ibs* (antonym) show that the words preceded by those non-typographic markers are synonyms or antonyms respectively and advise the user to look up those lemmata when necessary.

**KWISHYIRA UKIZANA** [...] Rb Iht kwigenga

**KWIGENGA** [...] Rb Iht kwishyira ukizana.

Where the address in question is a polysemic lemma and the synonymous or the antonymic relation is referring to one of its sense and not to the full meaning of the lemma, a number should be used to

ensure a rapid inner access to the specific sense in question. For example, the cross-reference to the lemma *kwishyingira* (a girl to get married unofficially or secretly) directs the user to the fifth and second sense of the lemmata *gucikira* and *kwijyana* as one can see below.

**KWISHYINGIRA** [...]. Rb Iht gucikira 5; kwijyana 2.  
**GUCIKIRA** [...] 1 [...]. 2 [...]. 3 [...]. 4 [...]. 5 Rb Iht kwishyingira.  
**KWIJYANA** [...] 1 [...]. 2 Rb Iht kwishyingira.

It should be noted that the lemmatisation of idioms as primary level treatment units can be implemented in general dictionaries as well as in dictionaries with idioms as sole subject matter. In most cases, dictionaries of idioms are cognition-oriented. When idioms are presented in a cognitive dictionary, the provision of the meaning, the origin and the cultural representation of the idiom is enough. This comment should contain encyclopaedic information of cultural, historical, linguistic, rhetorical, pragmatic and literary nature. The lexicographer should explain the origin of the expression or the mechanism underlying its formation and its meaning and conclude by indicating the meaning resulting from the explained process.

#### 4.3.3.3 Challenges and special issues in the straight lemmatisation of idioms

- **Alphabetical ordering and white space**

The straight lemmatisation of word combinations does not go without problems. One of them is related to the implication of the empty space between the components of an idiom in the alphabetical arrangement. One could be tempted to consider that *kwishyiraho* should appear before *kwishyira ukizana* merely because the letter *h* precedes the letter *u* in the general alphabet. However, it could be noted that the white space serves as a symbol and is shown before visible letters, the reason why the latter idiom should come first. Following the recommendation of Nielsen (1994) reiterated by Nkomo (2008), it is the duty of lexicographers to mitigate the inadequacies of the general alphabet of a given language by devising a lexicographic access alphabet which also determines the order in which its different signs are arranged. In addition, this issue should be explained to the users in the front matter. This applies not only to idioms, but also to other categories of lemmata such as compounds and other grammatical lemmata which are written as two or more words. In Kinyarwanda, a compound consisting of less than four words is written as a single word, but those consisting of four or more words (very rare) are written with hyphens between the components.

- **Stem-based versus full word lemmatisation**

The second challenge raised by this model relates to the use of the stem-based principle in lemmatisation and the alphabetical ordering. Many compilers of Bantu dictionaries follow a linguistic

biased tradition consisting of lemmatizing only the stems of words or arranging the macrostructure according to the first letter of the stem or root. This practice referred to by De Schryver (2010: 162) as the “one-size-fits-all approach” prevails in most Kinyarwanda dictionaries. Unfortunately it is not user-friendly. Most dictionary users, from lay people to university graduates, are unable to find a Kinyarwanda word in a dictionary displaying a stem-based lemmatisation. This practice is at the heart of the matter.

The problem of the stem-based lemmatisation of words in Kinyarwanda lexicography is a serious debate in Rwanda between dictionary users and lexicographers whose adherence to linguistic considerations is to the detriment of users’ needs (Prinsloo 1994, 2011; De Schryver & Prinsloo 2001; De Schryver 2010). The issue concerns the whole macrostructure and hence goes beyond the scope of this thesis. When recording idioms as macrostructural elements in a dictionary, it would be suggested to lemmatize them in full form and to arrange them according to the first letter of the opening word instead of using the stem of the opening word. The same applies to all elements of the macrostructure. For idioms having a verb in the first position, the use of the infinitive form should be considered for lemmatisation.

The use of the infinitive form for verbal idioms is merely motivated by a natural observation: every time when a Kinyarwanda speaker asks a question about the meaning of a verb or an idiomatic verbal expression, he/she automatically words it in the infinitive form, even if it appears in a conjugated form in a text right in front of him/her. This is most likely done in order to disambiguate the question, because a conjugated form includes not only the meaning of the lemma. Due to the holophrastic and agglutinative characteristics of Kinyarwanda, coupled with a conjunctive writing system, a conjugated verb includes many other senses entailed by non-lexicalised affixes. Moreover, it is more difficult to isolate the stem of a verb than to reconstruct the infinitive form. This is probably because a verb stem is never used alone and the user may never have seen it alone. Although the use of the infinitive form has recently been successfully experimented with in the lexicography of Kinyarwanda (precisely in DIEL), it is not the sole possible alternative to the stem/root-based lemmatisation approach in the lexicography of Bantu languages; different proposals have been formulated for other African languages such as Zulu and Kiswahili (cf. De Schryver & Prinsloo 2001; De Schryver 2010).

- **Provision of items giving the part of speech**

The part of speech or grammatical category is one of the most important lexicographic data types to be provided for each lexical item. The part of speech of a lemma specifies its morphological properties

and syntactic behavior or valence, that is to say, syntactic dependencies within which it is eligible. This applies also to idioms as it has been shown above. According to Mel'čuk (2006: 33), « La partie du discours d'un lexeme L est la classe à laquelle L appartient selon sa capacité d'être syntaxiquement dépendante d'un type particulier de gouverneur et d'avoir des types particuliers de dépendants dans un arbre syntaxique de dépendance ». [The part of speech of a lexeme L is the syntactic class to which L belongs according to its ability of being syntactically depending on a particular type of governor and of having particular types of dependants in a syntactic dependency tree]. In the framework of the MTT, a striking difference is established between the Deep parts of speech and the Surface parts of speech. The former characterises the lexemes and idioms appearing in Deep Syntactic Structures while the latter characterises the lexemes appearing in Surface Syntactic Structures.

Unlike a simple lexeme, an idiom does not receive a surface part of speech because it does not exist in the Surface Syntactic Structure; it appears only in the Deep Syntactic Structure of a sentence where “it occupies a single node, but corresponds to a sub-tree in the Surface SyntS [= SSyntS] - since at this level an idiom is a regular phrase” (Mel'čuk, 2006: 29). At the level of the Surface Syntactic Structure, idioms are like any other free syntactic groups; their constituents receive their own parts of speech in their respective articles in the dictionary. For this reason, it is impossible to assign parts of speech to idioms in the sense of lexemes. However, as Mel'čuk suggests, it is possible to indicate the main syntactic class of the idiom based on the head of the idiom in surface syntax, similar to the syntactic characterisation of a free phrase. This major syntactic type of the idiom is not really a part of speech, but a '*pseudo-Surface part of speech*' (Mel'čuk 2006: 29) in the form of nominal idiom, verbal idiom, prepositional idiom, adverbial idiom, conjunctional idiom, etc. The indication of deep parts of speech of lexemes and most of the idioms has no practical value in a general public dictionary intended for users with no linguistic training. Because the deep parts of speech of lexemes overlap with surface parts of speech in most cases, the indication of the latter suffices to give information on the syntactic use of the lexical items. However, if this approach works for most of idioms, it fails to account for a particular type of idiom. To understand this issue, a clarification is needed between endocentric and exocentric idioms.

At the syntactic level, a distinction is made between “syntactically endocentric idioms” and “syntactically exocentric idioms” (Mel'čuk 2006). In the former, the major syntactic class is determined by the constituent taken as the syntactic head of the idiom. These are the normal and prototypical idioms. For these idioms, the pseudo-surface part of speech corresponds to the deep and

surface parts of speech of the syntactic head. However, as Mel'čuk pointed out, for syntactically exocentric idioms, the syntactic classes or pseudo parts of speech do not correspond with their deep part of speech (Mel'čuk 2006). The deep part of speech cannot be determined by internal linguistic rules from the pseudo-surface part of speech. They are semantically and syntactically fixed. An idiom may be characterized as a nominal or conjunctional idiom whilst playing the adverbial or adjectival role in the sentence. Therefore, the lexicographer has to provide not only the surface part of speech of the idiom which is determined by its syntactic head, but also the deep part of speech which is determined by the use or function of the idiom in the sentence. Thus, apart from its surface syntactic characterisation, an exocentric idiom should receive an indication of its deep syntactic characterisation “in the form of an expression «used X-ally», where X is a Deep PartSp” (Mel'čuk 2006: 29, 51). In Kinyarwanda, we have idioms like:

- *mu máazi abíra*: idiomatic clause, adverbial use: in a bad state, in despair
- *nk'ibití by'i Bugesera*: coordinated nominal idiom, adjectival use: very slender
- *patá na ruúgi*: coordinated nominal idiom, adverbial use: intimately united
- *mbará na mbariro*: coordinated nominal idiom, adverbial use: for long
- *mpirú na nyoni*: coordinated nominal idiom, adverbial use: definitely.
- *mutí wa mperezayo*: coordinated nominal idiom, adverbial use: for ever
- *kibúno mpá amaguru*: idiomatic clause, adverbial use: as fast as one could
- *bubí na bwízá*: coordinated nominal idiom, adverbial use: at all costs, no matter what.

In most cases, exocentric idioms are used as unique distribution modifiers in collocations. One should be tempted to refuse them a place in dictionary merely because they are treated together with the collocations in which they stand as collocates considering their lemmatisation redundant. In addition, it is in line with the general recommendation of treating the collocations in the microstructure of their keywords. However, due to the high level of idiosyncrasy of these collocates, the non-native speakers and young users may not understand them and look up their definitions in the dictionary. Because at first glance the users may be struck by the idiosyncrasy and the particular structure of these idioms, they may fail to recognise that they participate in larger expressions (collocations). Therefore, it is acceptable to lemmatize these idioms at least for decoding purposes for the user who may fail to recognise the whole collocation and to prevent some users to think that these idioms do not exist because they are not recorded in the dictionary.

#### • Provision of transitivity

In the comment on the form of verbal idioms, it is important to provide an item giving data on the transitivity of the idiom as well. Current Kinyarwanda dictionaries fail to give this data. Consequently,

the objects of some idioms are confused with their constituents. For example, the two idioms given under the first sense of the lemma *muunsi* (adverb meaning: under):

*Gucá muunsi*: to be smaller or shorter than (*lit. to go under*).

*Gucá umuuntu muunsi*: to conspire, to plot against *somebody* (*lit. to go under sb.*).

The first idiom is an intransitive idiom. Neither an object nor a modal complement is required and its constituents do not fulfil such roles. The reverse is true for the second idiom. It requires an object, *umuuntu* (somebody) which can be inserted between the two real constituents of the idiom (*gucá umuuntu muunsi*), in the verb as an infixed object substitute (*kumúca muunsi*) or appear in the last position (*gucá muunsi umuuntu*). Here, *umuuntu* stands as generic object which can also be replaced by a person's name or a noun of another well-identified referent in real discourse (e.g. *umuturanyi*, a neighbour). Thus, the object is not part of the idiom and should not be presented as part of the lemma sign and the idiom should be lemmatised without the word that is subject to change.

On the other hand, the type of transitivity should be indicated in the comment on form and the concept of words likely to stand in the object position indicated in brackets or square brackets in the beginning of the comment on semantics. In the present case, each of the two idioms could be included as separate lemma sign or separate idiom, each one having its own comment on form and on semantics without indicating whether they are homonymic or polysemic; see Tarp (2001) for a discussion on the two concepts in lexicography. Notwithstanding the foregoing, when it is evident that the surface object is part of the idiom, in that case the idiom is not transitive. For example:

- *Gukúbita ikáramú*: to steal skillfully by willingly distorting calculations (*lit. to hit with a pen*)
- *Gukúbita impyisi inkoni*: to master oneself to perform a commendable task (*lit. to beat a hyena with a stick*)
- *Gukúbita umutwé ku ijuru*: to tell a white lie/to affirm something that is impossible (*lit. To bang one's head on the firmament*)

In the surface syntax, the verb *gukúbita* in the first idiom would be seen as having one object and two objects in the second idiom (direct and indirect), but because the three nouns (*ikáramú*: a pen, *impyisi*: hyena and *inkoni*: a stick) are lexicalised as part of the idioms in which they occur, the verbal idioms are intransitive. Likewise, *gukúbita umutwé ku ijuru* is an intransitive verbal idiom although the verb *gukúbita* could have had *umutwe* as object and *ku ijuru* as adverbial phrase of place had the string been a free combination. This contradicts the claim by Ntazinda (1980: 23-24) and Coupez (1980: 507) who have mistakenly considered these constituents as idiomatic complements (complements locutionnels) while they are simply idiomatic constituents.

- **On the issue of space saving**

Coming back to idioms, the status of idioms as independent lexical items cannot be stressed often enough. However, as for their lexicographic inclusion, it could be argued that this status should not prevail over other lexicographic pitfalls, such as producing a cumbersome dictionary because of the inclusion of unnecessary items or failing to use space-saving strategies during the compilation of the dictionary. Thus, the macrostructural inclusion of idioms is straightaway excluded in printed dictionaries because it is seen as space-consuming. Among the scholars who express their scepticism against strict lemmatisation of idioms is Louw who, when discussing school dictionaries in South Africa, argues that listing idioms in the macrostructure could be inappropriate in three ways (Louw 2006: 98):

Firstly, it will not always be possible to identify the initial component of the idiom. Articles are often interchangeable or optional at the start of an idiom and other subtle variations can occur. Secondly, lemmatisation of idioms can take up more space than alternative methods. Thirdly, the current dictionary culture (perpetuated by the available dictionaries) is one in which users of school dictionaries will probably expect to find the idiom as a sublemma under the lemma corresponding to the first word in the idiom that is considered to be semantically prominent, especially a noun, verb or adjective.

However, it could be reasoned that the three arguments do not stand in the Kinyarwanda context. Firstly, the problem of identification of the initial component of idioms does not occur in Kinyarwanda idioms because they do not have particles or articles in the first position.

Secondly, opposing the lemmatisation of idioms on account of space-saving is worthy of examination. Louw makes reference to school dictionaries which have their own spatial requirements. In Kinyarwanda dictionaries, the inclusion of idioms is not the only factor increasing the bulk of a dictionary. The volume of a dictionary should be negotiated taking into consideration its functions. It has been emphasised that the scope of items to be treated in a given dictionary and the extent to which their treatment should go, depend on the function of the dictionary. For example, a general language dictionary meant for text reception will undoubtedly be less bulky than its counterpart meant for text production, because the former will leave aside many collocations and illustrative examples. Current Kinyarwanda dictionaries have been compiled irrespective of lexicographic functions. They can be regarded as poly-functional although the way different functions are accommodated in them is still unclear. This has an impact on the use of space in the dictionary. By encompassing many functions in one dictionary, lexicographers have ended up with cumbersome and less serviceable dictionaries.

One of the examples is the use of a writing system indicating tones and vowel duration when writing a

Kinyarwanda dictionary. Normally, the regulations on Kinyarwanda orthography make provision for two writing systems: the common writing system that marks supra-segmental signs and vowel length on the one side, and the academic writing system that marks only high tones and vowel length by doubling the vowel. The latter system is used too seldom in some academic circles. For those who are not familiar with Kinyarwanda phonology, two words can be semantically distinguished by a tone or the duration of a vowel on a syllable in the same paradigm. But generally, these tones and length of vowels are predictable and Kinyarwanda speakers can read accurately without them.

Most people who consult a dictionary do so in connection with communicational tasks involving the common orthography to get assistance in written text decoding or in text writing. For those consulting a dictionary for oral text production related needs or written text production in academic writing, the provision of phonological transcription in the comment on form of each lemma sign would be sufficient. Therefore, when writing dictionary articles the use of tones and double vowels is superfluous and unnecessary and has no added value because communication is possible without them. In the dictionary, they do not satisfy any identified users' need; instead, the use of tones and double vowels dishearten users from consulting dictionaries because those signs are unfamiliar to them. It seems that this lexicographic practice does not take into account the language policies and users' needs.

An evaluation that was carried out on DIMO using a randomly selected sample (733 pages general monolingual dictionary) has shown that using the common orthography could save 22% of its volume, a space which is more than enough to host idioms as primary level treatment units, i.e. as lemmata. If this finding is any indication, it can be concluded that the space-saving pretext can no longer be invoked as an argument against the vertical lemmatisation of idioms. In a dictionary having text reception as dominant function, idioms should constitute an important part of the macrostructure because they are likely to be infrequent and more difficult to decode than other simple items. Therefore, emphasis should be put on them at the expense of frequent words. Of course, this is not the sole factor influencing the volume of Kinyarwanda dictionaries; it also depends on publishing factors such as the typeface size and style, the paper density and size, the choice of the format and the quality of binding.

As quoted above, Louw (2006: 98) suggests that dictionary compilers take into account users' dictionary consultation habits when formulating lemmatisation policies for idioms. However, regarding dictionary consultation culture among Kinyarwanda speakers, one has to remember that a dictionary is still a rare object in Rwanda such that the majority of them can hardly find it. The dictionaries of

Kinyarwanda are almost exclusively available in school or academic settings. Moreover, within and outside schools, most users lack required reference and linguistic skills for successfull dictionary consultation. Not only do they fail to search idioms and other types of phrasemes within a dictionary but they also fail to find lemmata in the macrostructure due to prevailing use of the stem-based convention in the lemmatisation of lexical items. Therefore, since there is no established dictionary consultation culture in the users of general language dictionaries of Kinyarwanda, there is no reason to believe that the lemmatisation of idioms as primary level treatment units could confuse users; instead it could make the consultation process easier.

In concluding this section, it is important to emphasise that the inclusion policy favoured by the lexicographers should be influenced by the functional use of the dictionary. Therefore, when planning the dictionary structures it is important to determine the main function of the dictionary and where and how idioms will be recorded to fit into that function.

Concerning lemma treatment, the lexicographers should welcome the inputs of theoretical linguistics to dictionary making when this is justifiable. Because a dictionary is more than the application of linguistic theories, the latter are useful but not always essential to dictionary compilation. Dictionaries are first and foremost utilitarian tools meant to satisfy the needs of users and to a certain extent to anticipate these needs. The user who performs a search in a dictionary is not using any linguistic theory; he/she simply is using a lexicographic tool. Therefore, because of being a mediator between scholars and scientists on the one side, and non-specialist users on the other side, the lexicographer above all considers what is advantageous to the potential users of the dictionary. For him satisfying the user needs prevails over using theories produced in the context of other scientific disciplines. As it has been deplored, the linguistic influence that has prevailed in Kinyarwanda lexicography has led to a word bias conception of the macrostructure of dictionaries. However, as stressed in this chapter, not only is the lemmatisation of idioms lexicographically and functionally possible, but it also yields a great deal in terms of improved data distribution, description and accessibility.

Apart from the lexicographic representation of Kinyarwanda idioms, the representation of proverbs constitutes another problem in the lexicography of Kinyarwanda. This will be addressed in the following section.

#### **4.4 Lexicographic treatment of Kinyarwanda proverbs**

The proverb (*umuganí w'ümugémuurano* in Kinyarwanda) is one of the well known traditional folk

literary genres in Rwanda. Since the colonial period, Kinyarwanda proverbs have fascinated Kinyarwanda scholars in Rwandan studies and have been the subject of numerous published compilations and studies. As an example, one can refer to Hurel (1920), Pauwels (1953), Kagame (1953), Merriam and Armstrong (1954), Nkongoli and Kamanzi (1957), Bigirumwami, (1967), Crépeau and Dion (1968) and finally Crépeau and Bizimana (1979) whose compilation is the most exhaustive (4454 proverbs) and whose second edition is in print. A comprehensive review of these works goes beyond the purpose of this thesis. In the following paragraphs, neither the characteristics of the Kinyarwanda proverb nor its typology is discussed. Little attention is also paid to dictionaries having proverbs as subject matter such as those referred to above where the alphabetical arrangement according to the first word has been applied. Instead the analyses focuses on the challenge that proverbs pose in general lexicography with special reference to Kinyarwanda dictionaries. The key question being addressed is “why, how and where to place proverbs not only in the general popular dictionary but also in a specialised dictionary of proverbs?”

In Kinyarwanda general dictionaries, proverbs are recorded almost in the same way, i.e. as usage examples of a sense of the lemma. The literal and figurative meanings of the proverb follow to help the user understand, although this is not the case in all dictionaries (DIMO for example). This strategy of treating proverbs as examples will be addressed in the discussion of different procedures of lexicographic representation of proverbs. In specialised proverb collections (types of dictionaries of proverbs), the most prevailing recording procedure is the alphabetical ordering according to the first word of the proverb.

#### **4.4.1 Proverbs as lexicographic data category**

The wish to include proverbs in dictionaries dates back a very long time in the history of lexicography. Although proverbs used to be marginally treated in general dictionaries, as a category of lexicographic data, proverbs are nowadays taking shape. The lexicographic status of proverbs can be derived from the shared function of both the dictionary and the proverb. In this connection, Mabika-Mbokou (2002: 283) reminds in the following words of one of the functions of dictionaries (my own translation):

For long, dictionaries have been used as tools for language acquisition and correct use. One of their attributes is that they provide users with a wide range of information which help them deepening their knowledge. They are reference works. They have always enjoyed a certain authority because, by allowing the use of words, constructions and meanings of a language, they integrate them into the community. Thus, they legitimate [sic] them, or condemn them by rejecting their use. (...) in addition, the language disseminated by the dictionary is essentially didactic. It contributes, therefore, to instructing its users. (...) As a consequence, apart from

preserving proverbs and using them as illustrative examples, the treatment and use of proverbs in dictionaries contribute to the pedagogic aspect inherent to the function of the dictionary.

From the above comment it can be concluded that both the proverb and the dictionary have in common the fact that they communicate linguistic and cultural contents to the community. Since the proverb is an important tool in community education, it is understood why they must be recorded in the dictionary. Thus, its inclusion is in line with the educational function of a dictionary and it is important for African societies to have such valuable traditional wisdom conveyed by modern education tools. In addition, proverbs go beyond expressing the experience of the cultural group they originate from (their cognitive function); they allow to formulate truths, condemn the follies and propose the behaviours that are peculiar to man of all times and places (Crépeau & Bizimana 1979: 4). In so doing they help users communicate efficiently.

Since the role of proverbs in communicating the community's wisdom cannot be underestimated, recording them in a dictionary is a way of putting that wisdom at the disposal of the greater humanity while at the same time, enhancing the communicative function of the dictionary. Of course, the number of proverbs to be included in a general dictionary, and subsequently their treatment, depends largely on a number of other factors in addition to their role vis-à-vis the lexicographic functions in which they are involved. Those factors are among other things, the frequency of use of a proverb, the available space and the inclusion strategy adopted by the lexicographer. Another reason that justifies the interest in proverbs for Kinyarwanda lexicography is that due to scarcity of lexicographic corpora in Kinyarwanda from which to extract illustrative materials, proverbs are frequently used as illustrative items.

#### **4.4.2 Lexicographic treatment of proverbs in general dictionaries**

As far as the recording of proverbs in the general dictionary is concerned, the lexicographers can envisage one of the following options: treatment of proverbs in the central list or in outer texts (Mabika-Mbokou 2002: 283). What follows is an examination of the lexicographic potential and drawbacks of the two approaches from a functional perspective, with reference to Kinyarwanda dictionaries.

##### **4.4.2.1 Lemmatizing proverbs as macrostructural primary level treatment units**

In the present state of knowledge, the lemmatisation of proverbs as main lemmata in the central list has never been applied, at least not in general language dictionaries. Where proverbs and pithy sayings

have been given a lemma status, they have been moved to the back matter in order not to disturb the general aesthetics of the dictionary. This was the case in *Duramazwi Guru ReChishona* (Chimhundu 2001). Chabata (2002: 111) explains that proverbs and pithy sayings were placed in the second section of the macrostructure (i.e. the back matter) as they have to appear in the dictionary as full or complete sentences that are in most cases too lengthy, thus, their inclusion in the central list, along with other lemmata, would have yielded an unattractive presentation of the dictionary pages, particularly given they would not be evenly distributed throughout the dictionary. Mixing proverbs with other macrostructure elements would also be problematic because dozens of proverbs may share the same opening word, as will be shown later.

#### **4.4.2.2 Treatment of proverbs within the microstructure**

Different types of microstructures can be envisaged when coping with the treatment of proverbial phrases. Lexicographers can for example list some proverbs within the microstructure of the article of a lemma representing one of the lexical items constituting the proverb. They can also use a proverb as an illustrative sentence of one of the senses of the lemma sign or as a polysemous sense or they can devise a proverb data category in the article structure of the lemma.

The option of using proverbs as examples that are preceded by a specific label as introduced in the front matter is widespread in Kinyarwanda dictionaries as illustrated by the following part of an article taken from the bilingual dictionary DIFO. (English translations are given in brackets in order to make the article understandable to the Anglophone readers).

**-CUTÍ 9,10** inshu A. **Umuuntu wuuzúra n'úundi, bafitánye umubáano unóze. Ami [a friend].** Napfúushije - - ntatezé kwiibagirwa na rímwe. Narárwaaraga ikaaza kuunsuura, naabá nshóonje ikaampa iyáabo. *J'ai perdu un ami que je n'oublierai jamais. Quand je tombais malade, il me rendait visite. Quand j'avais faim, il me donnait à manger.* **[I lost a friend that I will never forget. When I was ill, he/she visited me. When I was hungry, he/she gave me some food]** Pv Inshutí uyibona mu byáago. *L'ami, tu le vois dans les malheurs* **[Literally : A friend, you see him during calamity].** Compr: c'est dans les mauvais jours qu'on reconnaît ses vrais amis. **[A friend in need is a friend indeed].**

The marker *Pv* means proverb while *Compr* (fr. comprendre = understand) indicates the real meaning of the proverb, after the word-for-word translation. In this section of the article, the proverb is used as a second illustrative sentence of the sense “friend”. The sense of the lemma *-cutí* (*inshutí* in full form) which is illustrated here is the same in the proverb. However, many proverbs are not as transparent, and it can happen that the sense of the word used in a proverb has nothing to do with the sense being illustrated. Consider the following article. The number following the abbreviation *Pv* is the number of

the proverb in the publication from which it was taken, namely Crépeau and Bizimana (1979).

**-AÁGE 14 ubwaá A. Ubukené bwó kubúra ibigútunga bihaagíje.** *Pénurie de vivres [lack of food]* Pv. 2.894 Ubwaáge bukoombesha umugabo igikómá. *Le dénuement réduit un homme à lécher la bouillie [Literally: The lack of food compels the man to lick the porridge.]* Compr nécessité fait loi [Beggars cannot be choosers/Necessity knows no law].

In this article, the meaning of the lemma **-AÁGE** (ubwAÁGE) meaning (*shortage of food*) is semantically more restricted than the sense in the proverb, which refers to ‘the concept of situation of need in a broader sense, necessity’. In similar contexts, the use of proverbs as illustrative sentences can be misleading or useless to the user. This is worse when the dictionary does not indicate the meaning of the proverb in order to help the user understand the particular analogical use of the lemma. The example below is an extract from an article of the entry *UmuGREENZI* as represented in a Kinyarwanda monolingual dictionary (DIMO). The English translations are mine.

***umuGREENZI 1 Umuuntu urî mu nzira ugêenda*** [a traveller, a passer-by]. Yabâaye indoondogozi asigaye ayâagira ibyâago byê abahîsi n'âbageenzi [He has become a driveller, he tells his misfortune to everybody he meets]. ***Umg*** Inzira ntîibwiirâ umugeenzi [Literally: the road does not tell the traveller, real meaning: It is regrettable to know something too late that could have been to one's advantage].

In the DIMO, the proverb is indicated by the abbreviation **Umg** (from the Kinyarwanda term umuganí/proverb). But the dictionary does not provide the real meaning of the proverb. As far as the latter is concerned, it does not clearly denote the meaning of the traveller. If the user would envisage the sentence in the real interpretation of the proverb, it will require a long and hard interpretation exercise in order to be able to match the lemma with its representation in the proverb. If he/she views things from the perspective of the literal meaning of the proverb, such representation of the lemma may be out of context and consequently misleading. For example, the sentence “*Inzira ntîibwiirâ umugeenzî*” (It is regrettable to know something too late that could have been to one's advantage) literally meaning “*the road does not tell the traveller*”, is semantically an abnormal sentence, because a road cannot speak. For the same reason, it is not a good example. Not only can the use of a proverb as example sentence be misleading or useless, but it also makes it difficult to access paremiological data in the dictionary, because a proverb can be recorded under the lemma of any one of its lexical constituents. In this context, the user has no criterion helping him to know which one of the proverb's constituents is taken as the guiding element of the proverb. Even though he/she manages to find it, the treatment given to proverbs will be limited and less helpful for both decoding and encoding activities.

Although the use of proverbs as illustrative sentences gives an impression that proverbs are accounted for in the dictionary, that abstract words are illustrated and that it spares the lexicographers from a detailed description of proverbs, it is not, in the final analysis, a profitable lexicographic practice because it is likely to bring about semantic confusion, which is, according to Prinsloo and Gouws (2000:145), one of the characteristics of a bad example. However, nothing prevents the lexicographer from using a transparent proverb as illustration provided it does not create any of the problems mentioned above and that it is the only possible phrase. This being said, this practice should not be looked at as a policy for the inclusion of proverbs, but should be done merely for the sake of illustration.

A very uncommon option is the treatment of proverbs as polysemous senses of the lemma sign, instead of giving it its own paraphrase of meaning in a specific zone designed for paremiological data within the article, as have been suggested for idioms. The lexicographic potential of recording proverbs as polysemes is limited because it concerns a small number of lexical items. The shortcomings of the approaches discussed in the foregoing section, compel this study to adopt a transformative view and to explore other alternatives that should be looked at in the compilation of new Kinyarwanda dictionaries or the revision of the existing ones. Some of the alternatives are the use of additional alphabetical lists of treated proverbs, the provision of proverbs in a phraseological devised search field attached to the article of the lemma sign taken as the headword of the proverb (generally the first name) (Mabika-Mbokou 2002, 2006), or a pure and simple lemmatisation of proverbs in the macrostructure of the dictionary (Mpofu 2007). The first approach entails the use of outer texts while the latter two approaches remain within the scope of the central list.

Still in a microstructural treatment, the access-related problems pointed out above can be lessened if proverbs are made poly-accessible by means of an additional alphabetical list of treated proverbs. The back matter could be used to accommodate such a list that will indicate after each proverb, the lemma in the article of which it is listed and treated, and if applicable, the number indicating the sub-comment in which the proverb appears.

Regarding the inclusion of proverbs as treatment units in a special zone attached to the end of the article of the lemma representing the selected headwords, Mabika-Mbokou (2006: 170-171) proposes that (idioms and) proverbs should be listed alphabetically according to the words following the first noun of the idiom/proverb. Although the proposal of Mabika-Mbokou is an excellent idea as it shields users from guessing the lemmata under which a specific proverb is treated, the mixing of proverbs and

idioms in one data category it is not regarded as user-friendly in Kinyarwanda. This is mostly because users could expect to find idioms and proverbs in different places as they are not considered to be the same. Of course, in a dictionary where idioms are lemmatised as vertical treatment units as has been suggested in this chapter, this idea of Mabika-Mbokou would not be considered.

Mabika-Mbokou (2006: 170-171) further states that any other pattern of part of speech should be chosen provided it is clearly indicated in the front matter and consistently adhered to throughout the central list. The proverbs thus listed can be semantically described, something that would have been impossible if they were being used in non-lemmatic addressing such as illustrations, or recorded as part of polysemic lemma treatment. In addition, the lexicographers can list as many proverbs as the function and the type of the dictionary allows. Yet again, it will not be easy to decide under which headword a given proverb should be placed as unlike idioms where the verb or noun in the first position serve as a guiding element, more than one word qualifies as guiding elements in a proverb. It could be tempting to use the word in the first position as guiding element, but the fact that many proverbs start with the same word or with a non-content word, needs to be seriously taken into consideration. The same applies to the scope of semantic data that the lexicographer can provide when a proverb is listed in a semi-integrated microstructure or in a nested sub-article.

In all respects, as far as alphabetical arrangement is concerned, listing the proverbs according to the word in the first position is the best option. The lexicographers should consider listing the opening phrase for long proverbs displaying a binary structure. Moreover, in order to make it easier to access proverbs from the back matter list, each proverb in the text block dedicated to proverbs should be numbered. Once again, the inclusion in the back matter of an alphabetical list can enhance the accessibility of paremiological data. In the list, each proverb should be provided with an indication of the lemma under which it is recorded followed by the corresponding number indicating the order of the proverb in the proverb search zone. Proverbs in the list should be arranged alphabetically according to the first word while in the search field of the proverb data category, the alphabetical ordering should be combined with a numerical ordering.

If well implemented, this model can be helpful especially in a text reception oriented dictionary. However, as far as mother-tongue speaker needs are considered it could be argued that proverb searches would be more frequently performed in text production than in decoding-oriented tasks where context can assist the user to understand the meaning. According to this perspective, the listing of proverbs in outer texts, to be specific in the back matter, seems to open a promising way.

#### **4.4.2.3 Treatment of proverbs in the outer texts**

Treating proverbs in a communication-oriented outer text is the most compelling alternative to the treatment in the central list. In this framework, the lexicographers can record proverbs in the outer text of the dictionary, specifically in the back matter, as a separate section. At this level, it should be mentioned that dictionaries contain two main types of texts, the central list and the outer texts. The outer texts may either be in the front matter or in the back matter section. Referring to Kammerer and Wiegand, Gouws (2004a: 69) writes that both the back matter and the front matter texts together with the central list form the frame structure of the dictionary. Gouws (2004a) refers to the notion of extended outer texts to account for an outer text that has its own outer texts. The front matter or the back matter can display an inner frame structure functioning as a secondary frame structure. This led Gouws to distinguish two levels of frame structure: primary frame and secondary frame (Gouws, 2004a: 71).

The secondary frame structure applies to a dictionary in which the front matter or the back matter displays an inner frame structure, i.e. having the inner front matter and/or inner back matter in addition to the inner central list within it. The same applies to a bilingual dictionary when each of the two word lists has its own outer texts besides the outer texts addressing the whole dictionary (Gouws 2004a: 71-72). In addition, Gouws (2004a: 68-69) indicates that

(...) any text occurring in the front or back matter can be extended and this extension can be complete or partial. A partial extension leads to a given text being complemented by either front or back matter texts whereas a complete extension results in a text complemented by both front and back matter texts.

This applies to the proposed model because a presentation of proverbs in outer-texts will necessarily occur in an extended back matter text. The back matter allows a comprehensive treatment of proverbs and is in my opinion the best venue for their recording. Apart from the meaning of the proverb, the lexicographers can indicate in the comment on semantics the context in which it is used. Their listing can be implemented in two different ways with respect to the function that prevails in the dictionary. The thematic lemmatisation could fit users' needs in a dictionary designed for text production while the alphabetical listing could be enough in a dictionary meant for assisting users in text reception or translation. Yet again, the decision on the presentation of proverbs in outer texts must be done with due consideration to the needs and situations of the potential users and the dominant lexicographic function.

- **Alphabetical listing in the back matter**

A short discussion on the alphabetical listing of proverbs in the back matter follows. The selection of the guiding element has been discussed above. What has not been said is the treatment that a proverb should receive in a text production oriented list. One has to admit that unlike idioms which are integrated in sentences and whose lexicographic description includes items giving grammatical and syntactic characterisation, proverbs do not require such data categories. From a text decoding perspective, the only compulsory item is the comment giving the real meaning of the proverb. Naturally, the meanings of some proverbs, especially those referring to natural phenomena, are quite transparent and easy to interpret thanks to the symmetry between the literal meaning and the figurative meaning, while others, particularly proverbs linked to socio-cultural relations require a skilful analysis. When selecting the proverbs to be lemmatised, the lexicographer should focus on the latter category. However, the lexicographers must not allow themselves to get lost in the maze of conceptual and rhetorical analysis of proverbs. When looking up a proverb in a dictionary for assistance in text reception, a user wants the exact meaning of the proverb and nothing else. However, if the function of the dictionary or the proverb outer text is cognitive, the lexicographer should expand the comment to include stylistic analyses and the socio-cultural and historical value of proverbs. In a text production dictionary, the explanation of the symbolism of a proverb should be accompanied by a second sub-comment indicating the typical context in which the proverb is used. This applies especially to text production dictionaries where proverbs should be arranged according to the thematic principle.

In addition and along with the introductory user guide in the back matter text in question, a secondary level front matter text should be used to raise the users' encyclopaedic knowledge about the proverb as a literary genre for example.

- **Thematic listing in the back matter**

The thematic classification of proverbs can be done during the phases of data acquisition, preparation and processing. In thematic or conceptual lemmatisation, the lexicographers should list the selected proverbs in an extended back matter text complemented by an introduction and a table of contents followed by the main text in which proverbs are distributed in different themes. The main list should be complemented by a back matter section containing the alphabetical index of all treated proverbs. For each listed proverb in the alphabetical list, a dual number indicating the number of the theme and the order number of the proverb should be provided so to make the proverbs poly-accessible. Each concept should be divided into different sub-themes in which proverbs should be distributed according to their meanings. The themes should also be numbered continuously and highlighted with a font style and

shaded colour. The numbering of sub-themes is also important because their numbers may later play a role in the consultation process. The format of the number should include the number of the main theme plus the order number of sub-themes in the respective theme or concept. On the contrary, the order number of the proverbs should be provided, irrespective of the concepts and sub-theme boundaries, so as to make their indexing easier.

The outer access to proverbs should be enhanced by the table of contents in which the boundaries of each concept and sub-concept within it are indicated. In addition, the lexicographer should in the introduction of that text explain its content and access structure. The outer access structure should include the introduction, the table of contents, the titles of the concepts and sub-concepts, the alphabetical list and the numbers of concepts and proverbs. It should be supplemented by two types of running heads. The first running head containing the word *Imigani* (proverbs) as structural indicator, followed by a serial number, e.g. *Imigani 1–15* indicating the scope of the page content, should be put at the left header of the recto page. The second running head on the upper right side of the same page should contain the number and the title of the concept starting from that page. If that concept occupies many pages, it should be continuously provided as running head until a page containing a new concept is reached. On the back page, the lexicographer should interchange the positions of the two running heads. When arranging the running heads, the lexicographer should use them in an innovative way so that they do not create interference with other data presented in the header such as page numbers.

In this back matter text, the user may access a proverb via two roads depending on the situation and nature of the needs. If the user wants to ascertain the meaning or the context of use of a specific proverb (text reception/production) it would suffice to locate the proverb in question in the alphabetical list, take note of the number and come back to the main list to search the proverb carrying the identified number. Using the running head will undoubtedly speed up the consultation process. On the other hand, if the user wants to express a given meaning by a proverb but does not know the appropriate proverb, he/she can use the table of contents showing the covered concepts and sub-concepts and identify the concept and sub-theme(s) likely to contain the proverb expressing the idea. Once the theme and one or more sub-themes are identified, the user can reach them in the main list by means of running heads and titles of concepts and sub-concepts, after which he/she can navigate through the listed proverbs to see whether one of them matches the needs. This process can take time, but it is not possible to leapfrog these steps when the user searches for a proverb never seen before.

Although the conceptual lemmatisation of proverbs functions mostly in the perspective of text

production, it also plays a role in cognitive-oriented dictionaries because it allows users to study proverbs according to concepts. Therefore, by expanding the encyclopaedic scope of the explanations provided to proverbs, the lexicographer can kill two birds with one stone. Thus, if one sticks to user-friendliness and to the communicational function of the inclusion of proverbs, the communication role of such an approach is undisputable in that it accounts for both the production and the cognitive needs.

However, the conceptual lemmatisation as it is specifically referred to in this chapter should not be confused with listing a proverb under a headword merely taken from its wording because this headword may have nothing to do with the meaning of the proverb. If one follows this perspective for instance, the proverb below may be related to one of the following three areas: cattle (*inká*: cow), animal (*igikerí*: frog) or animal growing/husbandry (image of a cow trampling on the frog while grazing). English translations of both the literal and real meaning are provided in square brackets. The diacritics used in Crépeau and Bizimana (1979) are not modified.

*1009. Igikerí gikandagiwe n'inka kiti ukó zivuzé nyamáhembé.*

Le crapaud, piétiné par les vaches, dit : chaque fois que les (tambours) battent, ô Altesses cornues. [Literally: *The frog, being trampled on by the cows says: as you want, oh! horned prince*]

Les faibles doivent se résigner à subir les offenses des puissants. [Interpretation: *The weak have to resign themselves to oppression from men in high places*]

It is clear that the meaning of this proverb has nothing to do with the domains stated above. Similarly, although it talks about *goats* and *sheep*, the proverb below does refer to something radically different.

*Ihené ntíjyá mù ntàma.*

La chèvre ne va pas parmi les moutons. [Literally: *The goat does not go to sheep*]

Les gens de condition sociale différente ne se fréquentent pas. [Interpretation: *People from different social classes do not frequent one another*]

Both the first and second proverb may be properly recorded under the theme of social relations. In the proposed thematic classification, one major theme can be sub-divided into other fine sub-themes if necessary. Where justifiable, one proverb can belong to more than one theme.

Moreover, although this model is designed for the treatment of proverbs in general dictionaries which constitute the subject matter of this thesis, it can efficiently replace the widely used alphabetical listing in books of proverb collections, that are a type of proverb dictionary, when text production or cognitive functions are at stake. This is used in almost all compilations of Kinyarwanda proverbs. However, the

traditional alphabetical listing of proverbs has a disadvantage of complicating the search and access to the data when many proverbs are recorded under the same headword. This is the case in Crépeau and Bizimana (1979) for example, where the lexemes *umugoré* (a woman) and *umugabo* (a man) are chosen as the leading headwords of 30 proverbs starting with *umugoré* (N°3363-3392: 6 pages) and 52 proverbs starting with *umugabo* (N° 3287-3338: 7 pages).

## 4.5 Conclusion

The lexicographic listing of idioms and proverbs was discussed in this chapter. The main contribution has been on the functional model for inclusion and treatment of idioms and proverbs in Kinyarwanda lexicography. This chapter has emphasised the compelling role theoretical considerations play when determining a lexicographic inclusion policy for both idioms and proverbs. In this respect, the lack of sound principles for the lexicographic placement of these two types of expressions was highlighted. After a linguistic description of the Kinyarwanda idiomatic category, user-driven lexicographic inclusion and treatment models of Kinyarwanda idioms and proverbs reposing mainly on a transformative approach were proposed.

It has been found that while Kinyarwanda idioms are usually listed in the articles of the claimed ‘heavy/restricted’ semantic constituents, this recording model thwarts a swift access to idiomatic data in the dictionary, not only because it requires the user to make pre-consultation semantic judgments in order to guess the location of the idiom, but also because it hinders a horizontal alphabetical arrangement of data. The inclusion model developed in this chapter does not take into account the semantic intuition of both the lexicographer and the user as the guiding criterion. Rather, it proposes two alternatives for the inclusion of idioms, one that views idioms as second level treatment units and consists in the horizontal alphabetical ordering of idioms within text blocks attached to the article of the main lemma representing their first constituents. The part of speech orientation of this model is believed to facilitate further lexicographic descriptions of idioms with regard to their use and ensure an unimpeded data access structure. The second account and the most compelling one in my opinion is the straight lemmatisation of idioms. This model has been proved feasible and its implementation has been illustrated in this chapter. It follows that, contrary to the prevailing word-biased approach, idioms can be lemmatised as first level treatment units in the macrostructure of Kinyarwanda dictionaries. This is the most user-friendly way to deal with idiomatic expressions.

The lexicographic representation of proverbs was also explored in order to challenge the prevailing

practice of treating proverbs as part of the microstructure as usage examples. Regarding the inclusion and treatment of proverbs, this chapter has claimed that the semantically driven thematic recording in the back matter represents the best strategy as far as both communicational and cognitive needs are concerned.

Undoubtedly, the treatment of idioms and proverbs constitutes one of the complex issues in lexicography. In this chapter many scholarly accounts about their inclusion and treatment in various types of dictionaries were reviewed. It was underscored that to include idioms and proverbs in a dictionary should not be done for its own sake; the inclusion is aimed at catering for communicational and cognitive needs of potential users. Thus, for the user to fully benefit from the lexicographic treatment of idioms and proverbs, Kinyarwanda lexicographers should make sure the inclusion model they apply reflects the functional needs of users. In this regard, the contribution of the content of this chapter should not be undervalued during the compilation of new dictionaries or the revision of current dictionaries.

## **CHAPTER 5: A FUNCTION-DRIVEN MODEL FOR THE TREATMENT OF COLLOCATIONS IN KINYARWANDA LEXICOGRAPHY**

### **5.1 Introduction**

In the foregoing chapter it was emphasised that idioms are lexical items in their own right and a recommendation was made to treat them as either first or second level treatment units depending on the lexicographic function prevailing in the dictionary. Unlike idioms, collocations are not lexical items though they are phraseological data, and they do not qualify for a lemmatic treatment. In this chapter the issue of the lexicographic representation of collocations is addressed and a model for the lexicographic treatment of different types of collocations in Kinyarwanda general language dictionaries with respect to various situations of dictionary consultation and lexicographic functions is proposed. The focus is on text production and text translation due to the prevailing role of collocations in these functions and only a brief discussion on the other two functions is provided. In order to achieve this objective, the arguments presented in this chapter are organised in four main sections. In Section 5.2 the definition of the concept of collocation is elaborated on. In Section 5.3 the lexicographic representation of collocations in a dictionary meant for satisfying text production and translation related needs is discussed. In Section 5.4 proposals for handling collocations in text reception dictionaries are formulated while the focus is on cognitive perspectives in Section 5.5. In conclusion the major recommendations are summarised in Section 5.6.

### **5.2 The concept of collocation**

#### **5.2.1 Two approaches to the concept of collocations**

The concept of collocation is envisaged from two opposing perspectives in phraseology. The first approach is based on the frequency of occurrence of words appearing together in texts. In this sense, any sequence of words tending to appear together in more or less closed contexts will be qualified as a collocation. In the field of lexical statistics, collocation means merely a sequence or pair of words appearing together, and it is an operating concept particularly in corpus linguistics, discourse analysis and language teaching. This broad conception of the notion of collocation goes back to Firth (1957) and was taken up by British functionalists like Halliday and Hasan (1976) and Sinclair (1991). According to the former, “collocations” (in that sense) contribute to text cohesion. For example, in a medical text, it is normal to find associations such as *hospital...physician*, *ill....doctor*, *patient.....doctor*, *medical...consultation*, and so on. For Halliday and Hasan (1976), the cohesive effect of such pairs of words depends not so much on any systematic relationship as to their tendency to share the same lexical environment, to occur in “collocation” with another. In general, any two lexical items

having similar patterns of collocation, i.e. frequently appearing in similar contexts will bring a cohesive force if they occur in adjacent sentences. This “corpus-driven” or “knowledge-free” paradigm excludes the use of any prerequisite linguistic knowledge in extraction and analysis of collocations. The statistic data are seen as the point of departure for the building of knowledge or theories, not the reverse (Heid & Weller 2009: 30-31).

However, in recent decades, a new understanding of the concept of collocation has emerged. It is rooted in a theoretical perspective regarding the collocation as restricted lexical co-occurrence of at least two linguistic elements in syntactic and semantic relationships. This definition is opposed to the first, which is purely based on statistical distribution and assessment of the frequency of occurrence in corpora. In this direction, statistical approaches can help in identifying collocations, but alone do not provide definitional criteria and do not explain why and how collocations are formed. If one wants to extract typical phraseological collocations primarily by means of statistical methods, the extracted data must be manually refined; the use of linguistic criteria besides the researcher’s introspection is therefore crucial to avoid redundancy and noise inherent to the methodology itself. However, both free and non-free frequent word combinations are considered contextual illustrative data in corpus lexicography.

Although well represented in the corpus, the lexicographic interest of collocations is higher due to the constraints they pose on the use of and access to the lexicon resources. A collocation is lexical knowledge and a property of the lexicon. Any native speaker of a language is able to recognise it. The definition of the concept of collocation does not refer to the frequency of occurrence but to the way in which it is formed, the way the speaker accesses it and its phraseological nature. Unlike in knowledge-free approaches, the researcher first applies pre-determined symbolic patterns to get homogeneous data, which are then to be reorganised in significant associations, using statistic detection methods. This way of proceeding is followed by Heid and Weller (2009). This study is interested in the lexicographic representation of phraseological collocations and the notion of collocation in this thesis is considered in terms of the second conception and has a concrete formalisation in the Meaning-Text Theory of Mel’čuk. This theory has permitted the establishment of a sharp division between collocations and idioms, two types of phrasemes which are being confused in Kinyarwanda lexicography. In this chapter, collocations are dealt with. It is therefore of paramount importance to remember the mechanism leading to the production of and access to collocations in natural languages.

### 5.2.2. On the definition of the concept of collocation in the Meaning-Text approach

There is such an immense literature on the concept of collocation in the framework of the MTT and in various fields of research that it is irrelevant and impossible to review here, as this section aims to define the phraseological status of collocations<sup>3</sup>. The following discussion will therefore only refer to publications reflecting the phraseological status of collocations. In Section 3.2 it has been argued that a collocation is a compositional phraseme. Following Mel'čuk (Mel'čuk: In press), it has also been emphasised that a lexical phraseme is a phraseme whose meaning ' $\sigma$ ' is unrestrictedly constructed by the speaker for any situation to be verbalized or referred to, but the selection of lexemes for the expression of ' $\sigma$ ' being restrictedly made. Phraseological constraints operate between semantic representation (SemR) and syntactic representation (SyntR), and target the lexical expression of (signifier) ' $\sigma$ '. This means that for the given ConceptR, any signified obtainable by general selection rules is possible, but for a selected signified 'A+B', the corresponding signifier /X/ is restrictedly constructed: if ' $X = A+B$ ', then  $/X \neq /A+B/$ . Further to the apparent similarities between idioms and collocations, it has been difficult for a long time to separate the two lexical classes. Thus, it is important to indicate at this level of discussion how collocations and idioms are different. On the one hand, the restricted co-occurrence helps to define a collocation in a very strict way and distinguishes collocations from idioms on a syntagmatic basis.

For example the phraseme *gukúura ubwaátsi* meaning 'X who has received W as a gift K from Y goes to say thanks to Y normally bringing Y a gift Z' (literally *to pull up a blade of grass*<sup>4</sup>: X pulls up Z). In this phraseme, it is obvious that the signified is richer than the signifier. This is exemplified by a relatively high number of semantic actants in the signified against two actants in the signifier. The opposite is also possible. In *guhenera umugina* meaning to die or 'X dies' as SemR (literally to bend down and lift up one's bottom against + a mound<sup>5</sup>: X bends down and X lifts up the behind W of X against + Z). One can easily see that in terms of semantic actants, the signifier is richer than the signified. In these two phrasemes, lexical restriction is total. They are idioms.

<sup>3</sup> For a comprehensive account of the concept of collocation in phraseology, see Grossmann and Tutin (2003) and Tutin (2010).

<sup>4</sup> In this traditional ceremony, it was a custom to give to the master a bunch of fresh grass as a symbol of prosperity. The rite was performed by somebody who has received a cow from his chief, therefore the image, given the importance of pasture in a pastoral community. That way he expressed his wish to see the cattle of his master increase endlessly. But note that the grass was also used as cattle's litter or bedding.

<sup>5</sup> It is instructive to map the death-mound symbolic association in other languages: Fr tertre funéraire and Eng burial mound.

However, there are other lexical phrasemes, in which lexical restriction is partial. Thus, consider the phrase *gutéera indíriimbo* (to break into song, to start singing) with the SemR X starts singing W such that Y hears W of X). If the phrase were freely constructed, a corresponding phrase would be a paraphrase such: *gutáangira kuríriimba indíriimbo ikúumviikana* (to start singing aloud). But we have *gutéera indíriimbo* (to strike up a song). If the lexical item *indíriimbo* (B) is freely selected for its meaning ‘A’ in Kinyarwanda, it is not the same with the lexical item *gutéera* which is a semantically empty verb. Rather, *gutéera* (C) is selected in function of ‘A’ to express a meaning (M) in co-occurrence with the lexeme B (*indíriimbo*). In this instance, *gutéera* is selected as a light verb to support the noun *indíriimbo* which is a semantic predicate (song by X) with its verb properties (time, mode, aspect, and flections) in order to enable the noun *indíriimbo* to function as a predicate in its own right. This way, *gutéera indíriimbo* is lexically half-restricted; only one of its two components is selected restrictedly, unlike in idioms where all the components are restrictedly selected. This expression is a collocation. A collocation is a lexically half-constrained phraseme. That is to say, one component is freely selected by the speaker, just for its meaning while the other constituent is chosen according to the meaning to express in co-occurrence with the first component. The first component is called the base (B) or keyword of the collocation, and the other one is the collocate (C) (Mel'čuk: In press). This property of collocations has been referred to as restricted co-occurrence by scholars studying collocations in production perspective like Mel'čuk (In press, 1998), Hausmann (1989) and Tutin (2010).

According to Tutin (Tutin & Grossman 2002: 11), the restricted co-occurrence implies that the collocation is being considered in the framework of text synthesis or production. It means that once the speaker wants to produce a combination of two components, the choice of one of the two components is not free, but imposed. The co-occurrence is not free, but restricted, e.g. *guseesa amaseezerano*: to cancel a contract. The speaker first raises the question: what do people do with a contract when they want to stop it? Then he/she will choose as collocator the lexeme *guseesa* if he/she has it in his/her lexical repertoire, to express the sense of "to make a contract to stop being in effect."

Apart from restricted co-occurrence, semantic compositionality is another important defining element of the collocation. For the role and mechanisms behind the linguistic compositionality, cf. Section 3.2 where this aspect is also discussed. Semantic compositionality, in the case of collocations, implies that the explicit components of meaning found in the definition of a given collocation can also be found in the definitions of individual constituents of the collocation in question (Mel'čuk: In press; Polguère

2003). On the contrary, idioms are non-compositional lexical phrasemes.

Indeed, both restricted co-occurrence and semantic compositionality properties bring a sharp distinction between collocations and idioms. On the one hand, the restricted co-occurrence helps to define a collocation in a very strict way and distinguishes collocations from idioms on syntagmatic basis. On the other hand, semantic compositionality differentiates them on the paradigmatic axis.

### **5.2.3 Types of collocations in Kinyarwanda**

As indicated above, restricted co-occurrence and semantic compositionality are the criteria that help to discriminate collocations from other types of phrasemes, especially idioms. However, there are other important properties that serve as internal defining criteria of collocations. The typology presented here is an adaptation of a classification firstly put forth by Tutin and Grossmann (2002) and Mel'čuk (2004). In the following paragraphs I briefly discuss four major classification criteria (decodability, predictability, binary aspect and dissymmetry of components) that were made to classify Kinyarwanda collocations.

The four criteria described below give different categorisations of collocations according to the viewpoint the research adopts. For example, the first criterion, decodability, refers to the semantic dimension of collocations as it is closely linked to the semantic compositionality of collocations. The second criterion, predictability, evokes the cognitive aspects of collocations, i.e. the property of a collocation of being produced by the speaker by means of normal semantic and syntactic rules of the language, without having been memorised beforehand as a block. On the other hand, binary and dissymmetric aspects remind of the restricted selection nature of the collocation components, and allow classifying collocations based on the nature of syntactic relations between both components of the collocation. The semantic, cognitive and syntactic orientations of these criteria imply a tri-fold classification. In addition, this classification arguably displays an interesting parallelism with the three main lexicographic functions discussed in this thesis, namely text reception, text production and knowledge related functions.

#### **5.2.3.1 Decodability**

A collocation is decodable in the sense that the speaker can interpret its meaning even though its form is not predictable (Tutin & Grossmann 2002: 10). For example, a non-native speaker of Kinyarwanda can decode the collocation *gutáangiza ináama* (to open a meeting/conference) although he/she could not produce it had he/she not memorized the collocation beforehand. However, transparency is a matter

of degree. Thus, transparency is less obvious in *umujinyá w'úmuráanduranzúzi* (terrible rage), where the collocator is more opaque than in the former, probably due to its high individuality. Its occurrence in single distribution in this collocation makes it infrequent and therefore less likely to be transparent. According to the decodability criterion, Kinyarwanda collocations are distributed into two types: opaque and transparent collocations.

- **Opaque collocations**

The meaning the collocator has in the collocation is different from the meaning that it bears outside the collocation (lexicographic meaning); it does not have that meaning in any other combination. These combinations are arbitrary and non-predictable and in addition, they are not easily decodable. You do not perceive a semantic or stylistic relatedness that explains the motivation behind the construction. The association of the base and collocate is idiosyncratic. They are culture-bound collocations, motivated by euphemism or the avoidance of tabooed words in Rwandan society (cf. Bizimana 2002: 114). The following collocations fall in this category:

*Kubogora amatá:* (milk) to pour accidentally

*Umwaámi kwíibaambuura:* (king) to wake up

- **Transparent collocations**

Although non-predictable, these collocations are decodable due to the transparency of the collocator. The latter is empty or fuzzy in the case of support/light or achievement verbs or keeps up its original meaning, e.g. *gutééra imbabázi* (to arouse a surge of tenderness), *kuvúza indíurú* (to shout for help), *gusinya sheéki* (to issue/sign a cheque).

### 5.2.3.2 Predictability

Transparency does not prevent collocations from being non-predictable. The non-predictability or arbitrariness in collocations explains why you can say in French *gravement malade* (seriously ill), *grièvement blessé* (seriously injured/wounded) in the physical sense, but not #*gravement blessé* (except in emotional utterances: seriously offended/wounded). In Kinyarwanda collocations *umujinyá w'úmuráanduranzúzi* (a terrible rage) and *amáaso y'tinyána* (doe eyes: lit. calf eyes), the association of keywords with modifying collocators is not naturally predictable; it is arbitrary and idiosyncratic. According to the property of being predictable or not, collocations split up into two types.

- **Regular or predictable collocations**

These collocations are decodable and to a certain extent predictable. In most cases, the collocator is always and only used with a single base and its meaning includes a component of the meaning of the base. In other words, the collocator is defined, based on the base with which it combines. In most cases, the collocator expresses a property of the basis in such a way that when you know the collocator you also know the basis and vice versa. In Kinyarwanda and in some other languages, this is not only mutually semantic but it is also ideophonic to the extent that an etymological reconstruction could link both the basis and the collocator to the same etymon. Its lexicographic treatment refers to the basis and to the meaning of the collocation in which it appears.

Regular collocations do not need an explicit lexicographic treatment neither in a cognitive directed dictionary, nor in any other type of dictionary, besides the treatment given to its constituents, because they are generated from the knowledge the user has of both the collocator and the basis. What is needed is the inclusion in the dictionary of the collocator and the basis only. The derived collocations are implicitly treated as example sentences. The linkage of the collocator to the basis is natural and motivated. Only the highest singularity binding the basis and the collocator and making any synonym of the collocator unacceptable distinguishes these collocations from free expressions. Kinyarwanda has a number of these single distribution collocators that are exclusively encountered in collocations where they always appear in combinations with unique bases. For instance, we can mention the paradigm of animal sounds and other natural phenomena related collocations, e.g. *umuuntu kuvúga*: (human being) to speak, *inyoni kuríriimba*: (birds) to sing, *imbwá kumoka*: (dog) to bark, *umuyaga guhuuha*: (wind) to rustle, *motéeri gutogota*: (engine) to throb/roar/hum/drone, *inkubá gukúbita*: (lightning/thunderbolt) to thunder, *injáangwé guhiriita*: (cat) to purr, etc.

- **Irregular or non-predictable collocations**

This class contains collocations that do not fulfil the characteristics discussed above. They may be either opaque or transparent; what characterises them is that a speaker having the knowledge of both the basis and the collocator cannot produce the collocation if he/she has not learned it as phraseological knowledge. Even though some non-predictable collocations may contain single distribution collocators such as predictable collocations, their distinctness lies in the fact that the basis can accept synonyms of the collocators to express similar meanings with differing emphases. They are unpredictable, e.g. *kuvúgira kurí raadiyó* (to speak on the radio), *kugeendera kuu mbago* (to walk on crutches), *kugaba igitéero* (to launch an attack), *gutáanga igitéekerezo* (to make a suggestion, to give an idea), etc.

### 5.2.3.3 Binary and dissymmetric aspects

The binary property means that a collocation is composed of two words, or groups of words. Thus, it is more correct to talk about constituents or components than about lexemes when we refer to the pair of collocation constituents. For instance, apart from collocations such as *kurera + debedebé* (to spoil a child) we have more complex collocations such as *kugera + mu máazi abíra* (to be in an awkward situation) in which the collocator *mu máazi abíra* is an idiom, and *Gukóra + ubwúicanyi bw'iindéengakámeré* (to commit an indescribable slaughter) in which we have a combination of two collocations *Gukóra ubwúicanyi* (to commit a slaughter) and *ubwiicanyi bw'iindéengakámeré* (indescribable killing). However, the merging of two collocations does not happen often.

The binary aspect leads to another characteristic of collocations, which is the dissymmetric character of the components of collocations. In general, a collocation is composed of two elements that play different roles: a base, a keyword or headword (whatever it is called) with a normal meaning and a collocator (C) which is chosen contingent on the former.

*Inshutí (B) magará (C): close friend*

*Amáazi (B) y'úrubogóbogo (C): clear/limpid water*

To sum up, based on a binary nature and the dissymmetry of constituents, a collocation is a pair of two constituent parts being in syntactic-semantic relations to one another, in such a way that the collocate syntactically governs or semantically modifies the base. The dissymmetric configuration of the constituents of the collocation brings about two important combinatorial relations: the modification and verbalization. On the one side, we have collocations where collocators are the semantic modifiers (adjective, adverb) of the bases and on the other side, collocations in which collocators syntactically govern the base (support and achievement verbs).

- **Collocation with modifiers**

In a general way, this means that if B is the base or keyword and C the collocator, then the meaning of BC is the original meaning of B modified by C. In the following examples, the bases are in uppercases while collocators are in lowercases:

*KWÍRUKA amasígamáana* (to run as fast as your legs could carry you)

*UMUKÉECURU rukukuri* (very old woman)

In these expressions, the keywords bear their normal meaning which is intensified by the adverbial/adjectival expansions (collocates). Thus in *KWÍRUKAamasígamáana*, the meaning of

*kwíiruka* (to run) does not change significantly, apart from the intensification of the meaning (very fast) carried by the collocate *amasígamáana*. Similarly, *UMUKÉECURU rukukuri* refers to *umukéecuru* (old woman) while the collocator brings with it an additional intensification sense indicating the highest degree of aging, when referring to women.

- **Collocations with support and achievement verbs**

In these collocations, the collocator is either an empty support verb that assures the verbalisation of the nominal predicates or quasi-predicates, e.g. *gukóra icyáaha* (to commit a crime), *kíuzuza ibisabwa* (to meet/fulfil the conditions or requirements) or an achievement verb whose meaning is approximately “to do with the referent of the noun in keyword what one is supposed to do with it” (Mel’čuk 2004), e.g. *guhaaza icyífuuzo* (to satisfy a desire), *gutáanga iseezerano* (to give a promise), *indwáara guhitana* (an illness to carry off).

## 5.3 Collocations in a text production and translation-oriented dictionary

### 5.3.1 The role of collocations in text production and translation

Text production and text translation are two related activities because they converge on producing a text, the sole difference is that in a text production situation the user produces a new text whereas in text translation he/she produces or rewrite a pre-existing text in a different language. Both activities involve writing skills and a satisfactory level of proficiency in the language of production. Lexicographic needs are also quite similar in the two situations.

While recognising the importance of collocations in text production, Tarp (2004a: 315-316) cautions against the separation of idioms from collocations in lexicography, arguing that although linguists have established differences between idioms and collocations, dictionary users are not able to establish a distinction between the two aspects of phraseology. However, in the production perspective, there is strong evidence in favour of users' awareness of differences between idioms and collocations on the one side, and collocations and free combinations on the other side. This is seen in the pre-consultation search route taken by users vis-à-vis collocations and idioms, the unpredictable status of collocations and the non-compositionality of idioms. From the point of view of text reception, collocations pose less comprehension problems unless they display a higher level of opacity (idioms and single distribution collocators); therefore, they do not require a comprehensive lexicographic coverage for text reception purposes.

Many problems arise when collocations are at stake in text production activities and in translation.

Further to the idiosyncratic nature of many collocations, the need for comprehensive lexicographic presentation of collocations becomes more urgent when the needs of non-mother-tongue speakers are taken into consideration or when a translation situation is considered. Hausmann (1997: 283) reminds that a collocation is made of a base or keyword which is freely chosen and a collocator which is idiomatic. However, the idiomaticity of the collocator does not pose many problems in text reception for adult users. Quite similarly, it raises fewer concerns in text production in the mother tongue when the user has acquired a high command of the native language. Rather, it poses a challenge when envisaged from a text encoding perspective by a non-native speaker, a low language proficiency user and by a translator who is tasked to produce a given source language text into the target language. In these contexts, the collocator is idiomatic because it cannot be selected freely; it is restricted. For example, a user with a low command of the mother tongue, a non-native speaker and a translator knowing the word *amaseezerano* (agreement, contract) who wishes to find the Kinyarwanda verb used alongside *amaseezerano* to express an idea he/she knows will “cause the end of a contract” will inevitably seek assistance from the dictionary if he/she has never seen the collocation *guseesa amaseezerano* (to break/terminate an agreement). This example raises another question: If the needs of the user must be satisfied, where should this collocation be listed?

### **5.3.2 An onomasiological approach to the lexicographic inclusion of Kinyarwanda collocations**

The aim of an onomasiological representation of dictionary data is to assist the users who either ignore or do not remember the word or expression that expresses the meaning or concept they have in mind to find the required word (Sierra & Hernández 2011: 215). It goes without saying that listing collocations in the article of the collocator could be useless in this situation. Rather, they should be listed using an onomasiological approach, i.e. in the article of the base which is the point of departure of dictionary consultation in a text production situation.

#### **5.3.2.1 Theoretical foundations for an onomasiological presentation of collocations**

With reference to the inclusion of collocations in the article of the base or of the collocator, Mabika-Mbokou (2006: 213) suggested a frequency-based policy. However this is not regarded as a viable option because of two assumptions, namely that all users are at the same level of acquaintance with different components of collocations, and that frequency must be verified in an empirical way in all cases of collocations in Kinyarwanda prior to using it as a determiner of the inclusion policy. In addition, there is no direct correlation between the frequency of use and all usage situations. For all

these reasons, I opt for a function-driven approach wherein the issue of frequency is only indirectly addressed.

The examples provided by Tarp (2004a: 315) suffice to illustrate the supremacy of the base over the collocator in collocations. He points out that the English and Danish equivalents of the Spanish *lavar los dientes* (literally: to wash the teeth) are *brush one's teeth* and *børste tænder* respectively. In Kinyarwanda, the equivalent of the Spanish form is also symmetric (*kóoza améenyo*: lit. to wash the teeth). Taking collocations as idioms would lead to attaching much more importance to the verbs *lavar*, *brush*, *børste* and *kóoza*, to the detriment of the keywords *dientes*, *teeth*, *tænder* and *améenyo*. The fact that the collocators (verbs) vary from one language to another while the keywords remain symmetric is a strong indication of the importance of the base in the acquisition of and access to collocations.

In fact, when a user wants to verbalise a given conceptual representation (about cleaning teeth), the question he/she has in mind is “*what exact verb is used alongside the word teeth to express the idea of cleaning teeth, what exact verb is used with teeth coming through or falling out, etc.*”? In a L1-L2 translational setting, for a user to translate a collocation C of L1 into a collocation C in L2, it suffices first to translate the keyword and then, with the aid of a dictionary, identify the appropriate collocator in the article of the keyword. Since the meaning of the keyword at the heart of the collocation is known, the collocation should preferably be described under the lemma sign representing the keyword or base of the collocation. This is a general principle. For example, the Kinyarwanda lexical item *amaseezerano* (agreement, contract) may control more than 27 collocations denoting what happens to or is done with/for an agreement. These collocations are syntactically controlled by 27 different verbs and some of these verbs are not included as such in the Kinyarwanda dictionary because they are predictable forms (passivation, flexion) of typical verbs that are treated. Bringing together these collocations around their semantic base helps users to understand various nuances between them and to choose the appropriate one. If the base was not used to list collocations, not only would it be impossible to access most of them (except in text decoding), but close collocations would also be scattered in different unrelated articles of the dictionary.

However, as will be argued later, and for the sake of making the dictionary a polyfunctional one, it is advisable to also include the collocations in the articles of the collocates but only with cross-references to the lemmata representing the bases for treatment, especially when the dictionary targets users with a low level of language proficiency and reference skills. In addition, collocations that are likely to pose problems in text decoding should also be recorded and treated in semasiological perspective, i.e. in the

articles of collocates, (e.g. opaque collocations), and collocations containing single distribution collocates. This is important particularly in a dictionary having text reception as one of its functions. In a polyfunctional dictionary with text production as one of the accommodated functions, such collocations should also be treated in the onomasiological presentation, thus that dictionary would be receiving a double treatment.

As far as the inclusion of and access to idioms are concerned, it is contrived to imagine a user searching in a dictionary for an idiom he/she has never seen before. Although idioms represent a valuable asset in text encoding for those who already have a good command of the language, their non-compositionality rules out their access from zero in the production viewpoint. It is difficult to imagine any assistance that a user may get from a dictionary with respect to an idiom he/she has never encountered before, unless he/she reads the whole dictionary. Idioms should thus be treated as fully-fledged lexical items in dictionaries and should also be included in the central list as such. However, although idioms per se are fully-fledged lexical items and qualify as lemma candidates they still pose a number of problems in lemmatisation. Consequently, idioms are often included as sublemmata in the articles of words that represent guiding elements of idioms. This affects the way idioms are represented in dictionaries. It is not the same with collocations, which ought to be listed within the articles of their keywords.

Another difference between idioms and collocations is the importance of the two types of phrasemes in text production. Certainly, both idioms and collocations are useful ways of expressing ideas in both oral and written communications. However, despite the fact that idioms give a higher stylistic colour to the text thanks to metaphoric and euphemistic expressions, and that their use indicates a high command of language, they are not inevitable (Hausmann 2004: 312; Vrbinc & Vrbinc 2011: 250). Hill (2004: 2) argues that “the usage of idioms plays a role in the social positioning of conversational partners and to consolidate a social hierarchy”, and adds that an “idiom is more informative than its simple lexical counterpart.” Therefore, the high level of language proficiency required to use idioms explains why they are relied on in text production to a lesser degree if compared to collocations, especially for non-native speakers. Further, the text producer in most cases limits the use of idiomatic expressions so to make the text clear and easily decodable depending on the intended audience. If a user happens to look up an idiom in a dictionary for text production purposes, the ultimate goal of the dictionary consultation is often to ascertain the right use of that idiom.

On the contrary, collocations form an essential and active part of the lexical knowledge of both beginners and advanced language users whether they are native or non-native speakers, and they

constitute an important skill in language acquisition (Ferraro, Nazar & Wanner 2011; Spohr 2012). In addition, many semantic values in languages are universally exclusively expressed through collocations. Therefore, collocations are inevitable in text production, irrespective of the level of language proficiency of the user.

Referring to Gouws (1991: 82), Mphahlele (2003: 163-167) questions the practice of listing collocations in the microstructure of their bases:

The implied semantic relationship between multilexical elements and lexical elements should not compel the lexicographer to include multilexical items in the microstructure of lexical lemmata. This approach which is word-biased is detrimental to the effective retrieval of the semantic information of multilexical items. That is, if multiword lexical items such as compounds, collocations and fixed expressions are included in the microstructure or semantic comment of other lexical items, dictionary users would not be able to retrieve their meanings because they would regard any information presented after the definition of a lexical lemma as extralinguistic.

However, contrary to what Mphahlele says (2003:164), it is impossible for a lexical collocation not to have a semantic relationship with its base. Moreover, to presume that when treating a collocation as a sub-entry of the keyword it would handicap the retrieval of information regarding the collocation (because the users could consider any data presented after the paraphrase of meaning of a lexical item represented by the lemma sign as extralinguistic), is an abusive generalisation if applied to all situations of dictionary consultation; it can only be true in a text reception situation where the user consults a dictionary mainly for decoding purpose. In a dictionary designed to assist users in text production related tasks, the comment on semantics is an ideal venue for many data categories other than the paraphrase of meaning, e.g. illustrative sentences, collocations, etc. These items are essential for text production. In addition, no log file experiment has proven that dictionary users tend to confine their look up to paraphrases of meaning whatever the situation in which they decide to consult a dictionary. The inner access structure implemented in the dictionary is what mostly determines the easy accessibility to microstructural data.

In addition, when saying that “Treating multiword lexical items [term indifferently applied by Mphahlele to both the idioms and the collocations] in the entries of lexical lemmas diminishes the linguistic soundness of a dictionary”, Gouws (1991: 82) refers seemingly to idioms and compounds but not to collocations that are not considered lexical items (Gouws 1996: 60). Discussing the lexicographic treatment of collocations, Gouws (1996: 60) states that unlike the idiom which is regarded as a single lexical item,

.... a collocation is more than one lexical item frequently occurring in the specific combination. Collocations are included in a dictionary as part of the treatment of the lemma and they represent a typical contextual occurrence of the lemma. Idioms may not be included in dictionaries as part of the treatment of the lemma because they do not constitute a typical contextual occurrence of the lemma but they are independent lexical items.

From a text production perspective, a collocation represents a typical contextual occurrence of the lemma representing its base, not its collocator. The situation of use when a dictionary user deals with collocations in text production activities is the following: the user knows the meaning of the collocations he/she wants to use and the basis of the collocation, but does not know the full wording of the collocations; what he/she wants is the collocator that appears together with the basis to form the finite collocation (Jousse, L'Homme, Leroyer & Robichaud 2011: 135,137; Tutin 2005: 42). Therefore, the part of the collocation the user looks up for typical micro-syntactic configuration is generally the basis, not the collocator. If this observation is accepted, the purpose of including collocations in a dictionary should go beyond the function of illustrating the use of the lemma in a sentence. Above all, it imparts decisive semantico-syntactic information without which a good number of concepts may not be expressed. A collocation is lexical knowledge even if it is not considered a lexical item. Thus, the inclusion of collocations in text production dictionaries serves onomasiological or conceptual purposes. But they also illustrate typical syntactic combinations in which the lemma can prevail.

In the unlikely event where the user may need a certain type of collocation and knows the collocator but has forgotten or does not know the base, he/she may go to the lemma representing the collocate hoping to find in that article an item giving the collocation. Should that arise, the additional recording of the bases in the articles of lemmata representing collocates with cross-references to the articles of the bases where the treatment of collocations is detailed will allow the user to find the required collocation through a search road starting from the article of the collocator. This semasiological representation of collocations in dictionaries is also helpful from a language acquisition perspective (cf. Tutin 2005: 47).

In a text reception situation, the user looks up an idiom in a dictionary in order to find the meaning of the idiom which is often obscure while in a text production perspective, the need may be to ascertain its meaning, its default structure and possible structural transformations, its grammatical and syntactic properties or one of them. The user wants to know how the idiom is properly used in a broader sentence. In either case, the user must know the idiom to be able to locate it in the dictionary.

On the contrary, collocations are on the front burner in text production and translation orientations, where the user who already is familiar with the deep syntactic and semantic information he/she wants

to communicate, seeks the proper collocator (verb or modifier) to anchor the keyword in the required semantico-syntactic configuration. When dealing with text revision or correction, the user may also need to check the appropriateness of collocations and the accuracy of collocators for specific bases. Finally, the user can also consult a monolingual dictionary when translating a text from L2 into L1. This can happen when the translator manages to translate the basis at the word-by-word basis but cannot translate the collocator in the same way and decide to consult the L1 text production dictionary to select the collocator with which the base expresses the collocational meaning in question. What the user seeks in these three situations is the syntactic configuration in which a given base expresses a given meaning. The ultimate result of the search is neither the meaning nor the use of the collocation in a broader phrase, but the use of the base in a semantico-syntactic configuration which is the fully formed and well-built collocation (Heid & Weller 2009: 27). It is apparent that a collocation is nothing else than a semantic-syntactic use of the base along with a word with a higher degree of individuality, i.e. the collocator.

If this scenario is accepted, the base should be the default point of departure of dictionary consultation in text production activities because it is the only part of the collocation that is most likely to be known by the user before he/she undertakes the consultation process. Such a consultation process is successful when collocations are recorded in the article of the lemma representing the base (onomasiological representation). This is without reminding of the linguistic dichotomy theme-rheme or topic-comment/focus where the theme represents what is being talked about and the rheme what is being said about the theme (N'diaye-Corréard 2008). Because the base is the most likely known component of the collocation, recording the collocation as a co-text item in the integrated microstructure of the base poses no problem with respect to the inner access route; rather, it helps the user to get what he/she wants as quickly as possible and with little effort.

### **5.3.2.2 A comparison of semasiological and onomasiological presentations of collocations**

The following study compares the listing of collocations following the collocate (semasiological approach) on the one side and the presentation of collocations in the article of the base (onomasiological approach) on the other side. With the verb *gutéguura* acting as collocate, we can list the following 10 collocations.

#### **guTÉGUURa**

I (to organise)

*Gutéguura igitéero* (prepare/organise an attack)

*Gutéguura isomó (prepare a lesson)*

*Gutéguura ubukwé (organise a wedding ceremony)*

*Gutéguura imyúigaragaambyo (to organise a demonstration)*

*Gutéguura ináama (organise a meeting)*

II

*Gutéguura imibónano (to prepare a partner for sexual intercourse)*

III

*Gutéguura icyuúmba (to set a room)*

*Gutéguura améezá (to set the table)*

IV

*Gutéguura inyamáaswa (to untie an animal)*

*Gutéguura umutégo (to undo a trap)*

*Gutéguura igisaásu (to demine)*

The first group of collocations have the core sense: to organise, to prepare. The second group containing a single collocation refers to sexual preludes; the third to decorating and setting; the fourth to undoing/relieving of a trap. One can see that the four groups of collocations are far from being close; there is also a problem of how to separate collocations according to different homonyms of *gutéguura* (IV is a homonym of I, II, and III) or how to separate collocations derived from different polysemes of *gutéguura* (I, II, III). Undoubtedly, this practice does not provide a rational structuring of dictionary information. On the contrary, the onomasiological approach helps to model support or achievement verb collocations in a well-organised manner. Consider for example collocations that are built around the noun *igitéero*, the base of the first collocation in the foregoing list.

### **IgiTÉERO** (attack, assault)

*Gutéguura igitéero (prepare an attack)*

*Kugaba igitéero (to launch an attack)*

*Kúubura ibitéero (to resume an attack)*

*Kuyobora igitéero (to command an attack)*

*Kuraangaaza imbere igitéero (to command an attack)*

*Gutsínda igitéero (to win an attack)*

*Kubá mu gitéero (to be in attack)*

*Gukúmiira igitéero (to prevent an attack, to counterattack)*

*Kubúrizamó igitéero (to thwart an attack)*

*Gusáanza igitéero (to scatter an attack)*

*Gushwīragiza igitéero* (to scatter an attack)  
*Gusúbiza inyuma igitéero* (to repel an attack)  
*Gukúbita inshúro igitéero* (to crush an attack)  
*Gusakirana n'igitéero* (to meet an attack)  
*Kwiigaamba igitéero* (to claim responsibility for an attack)  
*Kugabwahó igitéero* (to be under attack)  
*Igitéero kwíibasira* (an attack to target)  
*Igitéero guhítana* (an attack to kill)  
*Igitéero cy'ubwíiyahuzi* (a terrorist attack)  
*Igitéero shuma* (small attack, skirmish)  
*Igitéero ruraangiza* (ultimate attack)  
*Igitéero siimusíga* (ultimate attack)  
*Igitéero cyo kwíihorera* (revengeful attack)

The listing of collocations in the microstructure of the base could help bring together the 23 collocations listed above that could enhance the pedagogic dimension of the article and let a wide range of options open to the user who seeks an appropriate collocation for translation or text production purposes. On a cognitive and psycholinguistic level there is good reason to view these collocations as a network paired around the node word *igitéero* such that its mere mention calls in mind the whole network. Otherwise, all these collocations would be scattered in at least 19 different articles of a given dictionary, with risk of losing or forgetting some of them.

### 5.3.2.3 Implementation of the proposed onomasiological model

Using articles retrieved from an on-going text-production hybrid dictionary Kinyarwanda – Kinyarwanda – English (KED) intended for translators and learners, it is shown how the recognition of collocation status could contribute to making the access to lexicographic data in Kinyarwanda dictionaries easier. I respect the notation system used in KED which does not show high tone and vowel length.

**KWICA /kwíica/ sh Cy 1**

**1 Kwambura umuntu cg inyamaswa ubuzima** (*person, animal*) **to kill.** Yahuye n'imbogo igiye kumwica ayirasa umwambi **Prov Ujya kwica ubukombe arabwagaza**

**2 Gukuraho icyariho, gutuma gisibangana** (*person*) **to cancel, to suppress (path).** Iyi nzira iri hafi y'urugo uzayice.

**3. Gupfunda cg gutera ihembe kw'inyamaswa** (*animal*) (*with a horn*), **to gore, to hit with the horn/head, to attack, to charge (at).** Uramenye ntunyure imbere y'iyo nka irica. Nyuza uwo mwana hirya iyo nyana itamwica.

**4 Kubuza ikintu kuba** to prevent from happening; to cause to fail. Urugendo nari mfite narwishe kuko mfite abashyitsi. **IsbKwica inyota:** *Kunywa (one's thirst) to quench, to slake, to drink.* **IsbKwica itegeko:** *Kurenga ku cyo wabujijwe (Jur) to infringe, to contravene, to transgress (the law, rule).* **Isb Kwica ubukwe:** *Kubuza ubukwe gutaha kubera impamvu iyo ari yo yose to cause the marriage/engagement to fail.* **IsbKwica umubyizi:** *Gusiba gukora ku munsi wari usanzwe ukorwaho to be out of work (on a working day).* **IsbKwica umuzungu:** *Kugira amahirwe to be (extremely) lucky.* **IsbKwica urubozo:** *Kugirira umuntu nabi bimugaragura to torture, to kill sb slowly.*

**5 Gutuma ikintu cg icyo umuntu akora kimera nabi** to damage, to corrupt; (air) to pollute, (*language*) to debase. Kwica inzoga. Kwica umurongo. Kwica isuka. Kwica igifaransa.

**6 Gutuma umuntu ababara** to harm, to make sb suffer. Ihwa riranyishe Uravuga ko utanyishe kandi unteye ibuye mu jisho

**7 Kumerera umuntu nabi kw'ikintu** to bother, to make sb unwell. Unyoye inzoga nyinshi iramwica. Niriwe ubusa none inzara iranyishe.

**8 Kurandura igihumyo cg ikiyege** to gather (eg: mushroom). Nagiye kugura inyama nsanga aho ibihumyo byapfuye ndabyica inyama ndazireka.

**9 Kurenganya umuntu** to be unfair against sb. Gukoresha abantu ntubahembe ni ukubica.

**10 Gukora ikintu nabi (to do sth badly)** to bungle, to botch, to mess up. Mu mihamirizo itanu bamubajije, ine yose yayishe. Navuze igisigo ngejeje hagati nsanga nabyishe nsubiramo Antgutunganya.

**11 Guca ku masezerano wagiranye n'umuntu** to contravene, to break (a contract). Twari twemeranijwe kuzahurira i Kigali nyuma y'iminsi ibiri unyicira amasezerano ntiwaza **We were committed to meet in Kigali, but after two days you broke our commitment and did not come.**

**12 Gufindura igisakuzo (a riddle)** to answer/guess.

**13 Shy Mu mikino y'umupira, gukinira undi nabi umuhutaza cg umukorera ibindi bibi bimeze nk'ibyo Neol (football, basketball)** to knock sb (usually willingly).

*Article KWICA from KED: Inkoranya y'Ikinyarwanda mu Kinyarwanda no mu Cyongereza (Kinyarwanda-English Dictionary) (draft of 5 June 2012)*

If we resolutely look at things from a production point of view, some observations stand out in the analysis of this article. At first, the microstructure of the article *KWICA* is a mixture of collocational and non-collocational meanings. Only senses 1, 6, 9 and 13 are to be seen as basic senses of the lexical item *Kwica*. In addition, sense 6 and sense 9 can be seen as different uses of one basic sense, this suggests grouping them in one subcomment on semantics. These uses are specified by corresponding illustrative sentences and separated by semi-colons. Special field labels, e.g. (football, basketball) and semantic labels e.g. (emotionally) are also used to distinguish different uses representing the same basic sense, while Arabic numbers could be used to mark different senses of the lexical item. The rest (senses 2, 3, 4, 5, 7, 8, 10, 11, 12) are collocational senses, which ought to be described in detail in the articles of their bases. A compromise would be to give a minimal definition conveying the core collocational meaning in the article of the verb, and using an appropriate mediostructural device (below the right-turned triangle is used to this end), to direct the user to the lemma representing the basis of the

collocation, under which the collocation is furnished and treated. Therefore, the above-mentioned article could be reformulated as follows:

### **KWICA /kwíca/ sh Cy1**

**1 Kwambura umuntu cg inyamaswa ubuzima** (*person, animal*) to kill. Yahuye n'imbogo igiye kumwica ayirasa umwambi **Umg** Ujya kwica ubukombe arabwagaza.

**Inm sh. ◊ Kwica urubozo:** *Kugirira umuntu nabi bimugaragura* to torture, to kill sb slowly.

**2 Gutuma umuntu ababara, amererwa nabi mu mubiri cg mu byiyumvo** (*physically*) to harm, to make sb suffer/unwell Ihwa riranyishe. Uravuga ko utanyishe kandi unteye ibuye mu jisho?; (*emotionally, morally*) to be unfair against sb Gukoresha abantu ntubahembe ni ukubica.

► umukinnyi (player), ihwa (thorn), inzoga (beer), inzara (hunger), inyota (thirst), urukumbuzi (nostalgia).

**3 Shy Mu mikino y'umupira, gukinira undi nabi umuhutaza cg umukorera ibindi bibi bimeze nk'ibyo Neol** (*football, basketball*) to knock sb (usually willingly).

**4 Gutuma ikintu kimera nabi** to damage, to corrupt; (*air*) to pollute, (*language*) to debase.

► inzoga (beer), umurongo (line), isuka (hoe), ururimi (language).

**5 Gukora ikintu nabi** (*to do sth badly*) to bungle, to botch, to mess up, to fail at. Mu mihamirizo itanu bamubajije, ine yose yayishe. Navuze umuvugo ngejeje hagati nsanga nabyishe nsubiramo. ► ikibazo (examination question, item). **Ibs** gutunganya (to do perfectly).

**6 Gukuraho icyariho, gutuma gisibanga cg kuba icyendaga kuba** to cancel, to suppress, to prevent from happening; to cause to fail. ► inzira (path), umwenge (hole), umwobo (hole), umuryango (door), idirishya (window), ubukwe (marriage), inyota (thirst), itegeko (law), isezerano1 (contract), umubyizi (working day), igisakuzo (riddle).

**7 Gupfunda cg gutera ihembe** (*animal*) (with a horn) to gore, to hit with the horn/head, to attack, to charge (at). ► inka (cow), imbogo (buffalo).

**8 Gusarura ibintu byo mu bwoko bw'ibihumyo** to gather (e.g. mushroom). Nagiye kugura inyama nsanga aho ibihumyo byapfuye ndabyica inyama ndazireka. ► igihumyo, icyoba, imegeri, inopfu, indenganzira.

### **KWICA UMUZUNGU /kwíca umuzuúngu/: Isb sh Cy0**

**Kugira amahirwe** to be (*extremely and inexplicably*) rich or lucky.

As one can see from this article, some meanings have been regrouped in an effort to direct the user to the noun with which the verb combines to form collocations (by means of a right-turned triangle indicator ►). By implication, those collocations have to be treated under the indicated lemma, within an integrated microstructure. This resulted in saving space and a more concise article. One verbal collocation (**Inm sh**) with intensifier having the verb KWICA as base *kwica urubozo* is identified and listed as a co-text in a collocation data category zone after the relevant meaning and other co-text data. There was also a true idiom *kwica umuzungu* (to be extremely and inexplicably rich or lucky) mixed with collocations in definition 4. This idiom has been phased out to the straight macrostructural level following the suggestion made in the previous chapter to present idioms as primary level treatment units in a text production dictionary. For collocational meaning, instead of treating them in the

microstructure of the collocator, cross-references are made to the lemmatic addresses of the respective bases under which collocations are described. Of course, the cross-reference addresses provided may be of no use if they are incomplete. It is necessary to explicate the specific sense under which the collocation is found by indicating the number of the sense if applicable, e.g. *isezerano1* means the collocation having *KWICA* as collocator and *isezerano* as base is found in the first sub-comment on semantics of the lemma sign *isezerano*. Let us illustrate this with the lemma sign *isezerano* (pl. *amasezerano*: agreement, contract) referred to in subcomment 6. The current outline of the article of ISEZERANO is given below (as of draft of 5 June 2012).

### **ISEZERANO /isezerano/ zn 5, 6**

**1 Ubwumvikane bw'abantu bashaka kugera ku kintu.** *agreement, mutual promise.* Mbere yo kukurongora wari wambwiye ko uzajya unyumvira noone wishe amasezerano ntugikora icyo nkubwiye. *Before marrying you, you promised me full obedience. Now, you have broken the alliance and you no longer obey to me.*

**2 Icyo uwihaye Imana yiyemeza gukora vow.** Isb *Isezerano rishya*: *Ibitabo bya Bibiliya byanditswe mbere y'ivuka rya Yezu. the New Testament.* Isb *Isezerano rya kera*: *Ibitabo bya Bibiliya byanditswe nyuma y'ivuka rya Yezu. the Old Testament* Isb *Umugore w'isezerano*: *Umugore mwashakanye mu buryo bwemewe n'amategeko y'idini. (religion) lawful wife* Isb *Umugabo w'isezerano*: *Umugabo mwashakanye mu buryo bwemewe n'amategeko y'idini. (religion) lawful husband.*

As can be seen, there is no mention of the collocation *kwica amasezerano* (to break a contract, to infringe on a contract). Logically, such a collocation would be linked to the first meaning (sense 1). This omission is applicable to many other collocations that are semantically controlled by the lemma *isezerano* and illustrates a poorly furnished article. To worsen the matter even further, the collocation *kwica amasezerano* is not described elsewhere in the dictionary, with the exception of being mentioned in the example sentence of sense 11 in the initial article *KWICA*. Thus, in order to rectify this omission, the article of the lemma sign *isezerano* has been reworked and supplemented with more collocations, including *kwica amasezerano*. A collocational zone was created after the relevant sense and marked by the structural indicator *Inm* (Inyume, collocation, in Kinyarwanda). For the sake of clarity, nominal collocations (**Inm zn**) and verbal collocations (**Inm sh**) were separated and the vertical blank diamond (◊) was used as the structural indicator preceding each collocation. The bold italic is also used as structural indicator for Kinyarwanda collocations while the normal italic is used for the English equivalent. Some alterations were also made on the phraseological expressions appearing under the sub-comment 2. The two idioms *isezerano rya kera* and *isezerano rishya* are raised to the status of macrostructure elements, while the last couple of expressions *umugore w'isezerano* and *umugabo w'isezerano* are collocational expressions related to the lemmata *umugore* and *umugabo* and are not

meant to be represented in this article and are to be treated within the microstructures of the articles of lemmata representing their bases. Gouws, cited by Louw, emphasises in the following words the need for a marked separation between the treatment zones of idioms and collocations within the article microstructure (Gouws 1996: 5).

Belonging to separate information categories the collocations and idioms (...) should be accommodated in different article positions which will leave the user with different search areas allocated to each information category. By using different typefaces or structural markers the user could be lead to a clear distinction between these two information categories.

Louw goes further arguing that whereas idioms should be treated as sublemmata attached to the article of the lemma taken as the guiding element, collocations “will fit into that part of the comment or subcomment on semantics reserved for examples” with the possibility of listing less transparent collocations as treatment units (Louw 2006: 99). In my opinion, nothing prevents the lexicographer from using a collocation for illustrative purposes if he/she deems it quite transparent and clear to fulfil the purpose and requirements of a good example (Prinsloo & Gouws 2000: 144-145). However, the illustrative information being limited to the purpose of illustration, confining the treatment of those collocations to an illustrating role may preclude a comprehensive presentation or coverage of collocations. In addition, the separation of more transparent from less transparent collocations makes sense only when envisaged in a text reception perspective. Furthermore, as shown in Section 5.2.3, collocations do split up into transparent and opaque collocations when seen from the point of view of decodability. But this is only one of the three criteria used to classify collocations, a criterion which functions in a text reception perspective. Therefore, I recommend a uniform and consistent presentation of collocations without any distinction between less and more transparent ones in a text production dictionary.

In this kind of dictionary, except for predictable collocations occurring as illustrative items, all collocations should appear in a salient position in the microstructure, within a special search zone attached to the relevant (sub-)comment on semantics. Each collocation should be treated under the relevant sense of the lemma standing as its keyword. Thus, Louw (2006: 99-100) recommends including “the collocation at the specific sense or syntactic function it corresponds to” instead of placing it at the end of the entry, so to spare the user remote addressing related problems and to provide him/her with “microsyntactic empowerment”. In this way, collocations are recorded as “co-text entries” illustrating what Gouws (1989:227, quoted in Louw 2006: 99) calls the “typical microsyntactic context of the lemma.” However, although collocations are part of the co-text entries where one also finds free

example sentences, the collocation search zone should be separate and clearly identifiable because, it has been argued, collocations are more than a mere side by side occurrence or a syntactic use of a given word. Therefore, the presentation of relevant collocations in a separate data category will help to avoid confusion between the occurrence of the lemma in question in free sentences or combinations from its occurrence in typical semantic-syntactic combinations or collocations. It will also ensure a comprehensive coverage of collocations in the dictionary.

For cultural-bound collocations (the concept is introduced by Gouws 2006: 28), it should be necessary to give a comprehensive treatment. For example, while most collocations do not require a paraphrase of meaning because they are transparent, culture-bound collocations are often opaque and if they are not explained, they cannot be of any use in text production for the users who ignore the meaning of the collocators. It should be appropriate to indicate that the collocator cannot be replaced by its synonym, indicating what is said but also what is not said. For instance, whilst Kinyarwanda uses the general verb *gupfá* (to die) in free combinations for anybody but the king (*umwaámi*), for the latter the same concept is expressed via a collocation whose collocator is the verb *gutáanga* (lit. to give: he gives the kingdom to his successor, his son). In the article of *umwaámi* (king), not only should the lexicographer indicate which verbal collocation is used to convey the meaning “(king) to die”; but he/she should also indicate that the ordinary verb *gupfá* is never used in this situation. For this to be clearly visible the lexicographer should use integrated inner-texts or text blocks (Gouws & Prinsloo 2010: 502) to provide these cultural remarks to warn the user against erroneous or out of context use of words. Following a suggestion made by Gouws (2006: 34) about the treatment of cultural phrases in the microstructure, a specific structural marker (**Umc** for Kinyarwanda word *umuco*: culture) should be used for these texts blocks. This will incontestably empower the user seeking lexicographic assistance in text production activities.

In the stretched article of the lemma *isezerano*, collocations are recorded in two separate text blocks depending on whether they are verbal or nominal syntagms, in order to help the user grasp the syntactic function of the collocations. However, horizontal alphabetical listing is strictly followed within each clustered text block, but the particular syntactic structure (V+N) of verbal collocations leads to the violation and continuation of an even alphabetical arrangement between the nominal collocations and the verbal collocations condensed text blocks. The resulting articles looks like this:

### **ISEZERANO /isezerano/ zn 5, 6**

**1Ubwumvikane bw'abantu bashaka kugera ku kintu** agreement, mutual promise. Mbere yô kukurongora warî wambwiye kô uzajya unyûmvira nôone wiñshe amasezerano ntûugikora icyô nkubwiÿe. *Before marrying you, you promised me full obedience. Now, you have broken the alliance and no longer obey me.*

**Inm zn.** ◊ amasezerano magirirane bilateral contract ◊ amasezerano rusange collective agreement ◊ amasezerano y'inguzanyo loan agreement ◊ amasezerano y'akazi employment contract ◊ amasezerano y'amahoro peace agreement ◊ amasezerano y'ishyingiranwa marriage alliance ◊ amasezerano y'ubugure sale contract ◊ amasezerano y'ubugwate security,repurchase contract ◊ amasezerano y'ubukode lease contract

**Inm sh.** ◊ amasezerano afite agaciro valid agreement ◊ amasezerano gushyirwa mu bikorwa agreement to be implemented ◊ amasezerano gutangira gukurikizwa agreement to enter into force ◊ amasezerano guteganya agreement to provide for ◊ amasezerano kurangira agreement to end ◊ gukurikiza amasezerano to respect a contract/agreement ◊ gusesa amasezerano to terminate an agreement or contract ◊ gushyira mu bikorwa amasezerano to impliment anagreement ◊ gushyira mu kaga amasezerano to undermine/endanger an agreement ◊ gusinya amasezerano to sign an agreement ◊ gutegura amasezerano to draft a contract/agreement ◊ kongera amasezerano to renew a contract ◊ kuba umwe mu bashyize umukono ku masezerano to be signatory to a contract ◊ kuburizamo amasezerano to wreck/ruin/break down an agreement/convention ◊ kugirana amasezerano to enter into a contract/to agree on ◊ kugwa munsi y'amasezerano to be/fail under agreement/contract ◊ kurenga ku masezerano to infringe terms of a convention/contract ◊ kuvugurura amasezerano to amend a contract ◊ kwagura amasezerano to expand a contract ◊ kwamagana amasezerano to protest against an agreement ◊ kwica amasezerano to break an agreement ◊ kwisunga amasezerano to rely on the agreement ◊ kwivana mu masezerano to unilaterally break an agreement

## 2. vow Icyô uwîihaye Imana yiymemeza gukôra.

**ISEZERANO RYA KERA**/iseezaranoryaakéera/Isb zn

Ibitabo bya Bibiliya byanditswe mbere y'ivuka rya Yezu *the Old Testament.*

**ISEZERANO RISHYA**/iseezaranorishyá/Isb zn

Ibitabo bya Bibiliya byanditswe nyuma y'ivuka rya Yezu *the New Testament.*

*Article ISEZERANO after reworking.*

The above-outlined proposal is designed to be implemented in dictionaries with text production and text translation as main functions. And although it is exclusively exemplified for a hybrid dictionary, it can be applied with small adaptations to both monolingual and bilingual dictionaries. The KED is an ongoing project in which collocations must occupy a salient position due to the needs it is expected to cater for. In keeping with this requirement, this study has shown how to improve the access to collocations in this dictionary so that it satisfies its default functions. The pedagogic orientation of KED also justifies a partial onomasiological treatment of “difficult” collocations in which the lexical co-occurrence is also indicated in the articles of collocates with cross-reference addresses to the articles of the bases where collocations are described in detail. Therefore, the KED is a multifunctional

dictionary with text production and text translation as main functions, but with the possibility to accommodate secondary functions, e.g. text reception and cognitive functions.

#### **5.4 Kinyarwanda collocations in a dictionary designed for text reception**

As argued by Fuertes-Olivera, Bergenholz, Nielsen and Amo (2012), collocations are relevant for reuse in text production and in text translation, hence the emphasis put on text production in this chapter. However, to confine the lexicographic concept of collocation to the two lexicographic functions might be an oversimplification of the concept as this concept is itself entailed by the simplistic definition of collocations as typical word combinations that is far from being this simple and as illustrated in Section 5.2. Therefore, to qualify the argument of Fuertes-Olivera, Bergenholz, Nielsen and Amo (2012), it can be said that the prevalence of collocations in text production and translation does not rule out a detailed treatment of collocations in dictionaries in which text reception prevails as main or secondary function. Rather, depending on the language in question, the types of collocations attested in that language, and the level of proficiency of the intended audience, some (not all) types of collocations should be taken into account in text reception and be described in detail. Suffice to say, when such collocations are recorded, a particular type of logic must be followed.

Naturally, text reception dictionaries envisage the lexicon exclusively from a semasiological perspective while text production dictionaries encompass both semasiological and onomasiological angles. In text production, it is important that the user has a conceptual or a semantic representation that he/she wants to express hence the reason why the dictionary is consulted, i.e. to find the appropriate word or word combination. In a text production perspective, the semantic representation that the user has in mind leads him/her to the key word – here the base of the collocation – that directs him/her to the exact address where the collocation is. For example, if at the end of the search the user's assumption is infirmed, he/she will abandon the word and embark on a new hypothesis, but the initial semantic representation will remain the same. The base of the collocation prevails over the collocate even when the user knows the collocation and wants only to ascertain its meaning. In a situation of text production, it is preferable that the user must have seen the lexical item in order to find the corresponding lemma sign representing it in a dictionary. Thus, the inclusion policy implemented for collocations should reflect this semasiological nature of decoding dictionaries. In other words, collocations should also be treated in the microstructure of the collocators.

When selecting collocations to be represented in a decoding dictionary the types of collocations likely

to pose challenges in text reception, should be considered. The lexicographers should also take into account the level of proficiency of the users in the language in question. When considering collocations and text decoding, two different trends are reported: opaque collocations and transparent collocations. Generally, transparent collocations pose no problems in text reception and should not be taken into account. Similarly, for adult users with a high command of the language, most opaque collocations do not constitute an impediment in text understanding. However, I am of the opinion that collocations should compulsorily be recorded in a decoding dictionary with the following conditions:

- A support or achievement verb in which the specification of the base of the collocation is compulsory in the definition of the verb: In this case, the base appears in the definition of the verb. Because the verb cannot be defined without reference to the element representing the base of the collocation, the definition can be deemed as a comment on the sense of the collocation containing that support verb. If there are illustrative sentences, they are not addressed at the verb as such, but at the whole collocation of which the verb is a constituent. Thus, the collocation enjoys a non-lemmatic addressing in the article because it is not mentioned explicitly as a treatment unit but is only contextualized in the illustrating item. When the verb has many senses from a reception perspective, which is often the case, the lexicographer should indicate the respective use of the verb by signaling in square brackets the object or subject it takes with a given sense. In some dictionaries Kinyarwanda collocations are given both as items illustrating the use of the verbs and as treatment units at the same time. Note for example the five collocations highlighted in italics in the article of the lemma *gutéera* VIII in *Dictionnaire Rwandaïs-Français de l'INRS*, published by Jacob.

**gutéera**, -ye VIII (cfr *gutá*), **fixer, coller une chose sur une autre** [to fix/apply/stick sth on sth]. *Gutéera ibifuungo* : coudre des boutons (sur un vêtement) [to fix buttons on a cloth]. *Gutéera ikiremo* : repiécer [to mend a cloth]. *Gutéear ibára* : orner d'un dessin [to decorate]. *Gutéeara ikiraká* : réparer une chambre à air à l'aide d'une rustine [to stick a patch on the rubber, to repair the rubber with a patch]. *Gutéera iraángi* : peindre [to paint]. (...).

The French definition (in bold) that precedes the collocations in the above-mentioned article does not give a paraphrase of meaning of the verb; rather, it is a minimal paraphrase of all collocations in which the treated lemma stands as a support verb. These collocations roughly express the idea of “fixing or applying something on something else”. This idea is not inherent to the verb; it is the property of lemmata representing the bases of these collocations. These lemmata equally express the sense of something meant to be fixed or applied to something else. Therefore, the contextual definition is addressed at the collocations, i.e. the collocational

context of the verb and not the verb as an isolated lexical item. As microstructural co-textual items, the collocations are addressed at the main lemma but they are also elevated to the level of treatment units further to article-internal procedures of non-lemmatic addressing in that they have a comment on semantics (paraphrases and equivalents).

- With predictable collocations where the collocate expresses an intrinsic property of the basis, so much so that the collocator cannot be defined regardless of the basis even though it has its own independent existence as a lexical item: These collocations should be treated implicitly in the comment on semantics of the collocator, either in the definition or in the example, if the latter is provided. For example, in the Kinyarwanda collocation *umusatsi w'irende* (sleek hair, lit. hair sleek), the noun *irende* is a modifier collocator. The article of the lemma in *Inkoranya Iciriritse y'Ikinyarwanda* (Kinyarwanda Elementary Dictionary) is the following:

**IRENDE** /ireénde/ zn 5,6

**Umusatsi worohereye akenshi ugakunda kuba muremure no kwhihina.** Umusatsi w'irende urushya iyogosha. [Soft, long and curly hair. Soft hair is difficult to shave off. ].

The keyword of the collocation, *umusatsi* has a generic meaning (hair) and is a hyperonym of *irende* (soft when speaking of hair) just because the latter refers to one of the properties of the former, and the illustrative sentence contains the collocation “*umusatsi w'irende*” as co-textual support of the lemma *irende*. This collocation would have been sufficient to illustrate the lemma because the remainder of the sentence - *Umusatsi w'irende urushya iyogosha-* has nothing to do with the lemma *irende*; rather, it illustrates the whole collocation *umusatsi w'irende* as the syntactic agreement of the two main parts of the sentence suggests: *Umusatsi w'irende urushya iyogosha*.

- Another type of collocation that deserves a place in a text decoding dictionary is opaque collocations containing single distribution words or idioms as collocators. The singularity and unconventionality of these collocators make the collocations in which they appear more or less opaque. When the user encounters such an opaque collocator in a text and finds that he/she does not understand it, he/she may search for the meaning in the dictionary. But the collocation as a whole will not be studied only the collocator will. Yet if such a collocator is not in the collocation, it does not have its own independent existence and its meaning is abstract and indescribable in isolation. The meaning is a lexical function and as such, it only occurs in combination with the keyword of the collocation. These collocators are support adjectives and support adverbs, following the

example of support verbs. They are indefinable irrespective of the basis and the whole collocation. Be that as it may, the collocator must be lemmatised. Its comment on semantics should contain the collocation in which it appears and the explanation of that collocation. Consider the following collocations:

*Umujinyá w'umurándezanzúzi* (terrible/wild rage, literally: a rage making somebody wild to the extent of uprooting squashed plants): The collocator *umurándezanzuzi* should be lemmatised and receive the whole collocation within its microstructure, as follows:

**UMURANDURANZUZI/umurándezanzúzi/** zn 3,  
**Inm ◊ Umujinya w'umurándezanzuzi:** Umujinya mwinshi cyane umeze nk'ibisazi utera umuntu gukora ibantu mu buhubutsi akaba yakwica n'umuntu [terrible rage, wild rage].

*Kugeenda biguru ntenge* (to walk slowly, lit. to walk in such a way that one's heavy legs bother one's steps, walk as if one is unable to lift one's heavy legs): The collocator *biguru ntenge* is an exocentric nominal idiom with adverbial function (see sub-section 4.3.3.3) exclusively found in the above-mentioned collocation. It can be lemmatised and hosts the collocation in the following way:

**BIGURU NTEGE /biguruntegé/** Isb/zn ≡ Umgr,  
**Inm ◊ Kugeenda biguru ntenge:** Kugenda buhoro ku bushake [to walk slowly].

- Finally, when treating a lemma sign representing a cultural-bound lexical item (Gouws 2006), the lexicographer should in the co-text entries provide the typical cultural-bound collocations in which that lemma stands as a basis. In Kinyarwanda, lexical items referring to *king*, *churn*, *cow*, *milk*, *drum*, *crushing mortar*, *millstone*, *bow*, *wood-made milk jar*, to mention a few, fall within the scope of cultural-bound lexical items that form part of many collocations. In a dictionary combining both text reception and text production functions, such cultural data provided in the inner texts positioned in an integrated microstructure (Gouws & Prinsloo 2010) or in a semi-integrated microstructure (Gouws 2006), will primarily be directed at text production (Gouws 2006: 33).

## 5.5 Kinyarwanda collocations and knowledge-related needs

Usually a cognitive-oriented dictionary assists users in expanding their knowledge of grammar and

vocabulary in a foreign language. Tarp (2009d: 161) indicates that the ultimate goal of learning a foreign language is to acquire language skills that enable the learner to communicate in the language in question. As to language learning, he states that “dictionaries can only assist foreign-language skills *indirectly* if these skills can be developed by mediation of foreign-language knowledge and communication” (Tarp 2009d: 161). This happens when a learner consults a dictionary to increase his knowledge of vocabulary and grammar. The knowledge thus acquired is not yet foreign-language skills. Describing the relationship between foreign-language skills and foreign-language knowledge resulting from a conscious study of grammar and vocabulary, Tarp (2009d: 162) formulates that:

Foreign-language knowledge only turns into foreign-language skills if it is internalised, something which also happens with communication as the mediating element. Based on existing foreign-language skills and conscious thought arising from the foreign-language knowledge acquired, learners in a communication situation can decode or formulate a text and thereby activate new words and grammatical phenomena, which in turn affect and increase their foreign-language skills with a view to improving future communication.

Therefore, with respect to its cognitive function, the dictionary can assist foreign-language learners in an indirect way. If communication problems arise when communicating in a foreign language, dictionaries with communicative functions can directly provide the needed assistance (direct purpose of dictionary consultation) and contribute also to the improvement of foreign-language skills (indirect purpose of dictionary consultation), (Tarp 2009d: 163).

As referred to above, the concept of language learning refers to a foreign language because native languages are in most cases learnt unconsciously. Therefore, if Tarp’s perception of the language learner were applied to Kinyarwanda, it would be envisaged in the context of a foreign-language speaker learning Kinyarwanda as a foreign language. However, without wishing to alter and to dilute the words of Tarp (2009d) which are unquestionable, it is not excluded in our case to expand the concept of language learning to Kinyarwanda speakers learning some skills of their mother tongue which are provided as part of schooling. Yet the methods and ways in which conscious native-language learning interacts with the unconscious acquired language skills have not received the same interest as seen in foreign-language learning.

Although collocations are not lexical items, they are part of phraseological knowledge and when learning (consciously or unconsciously) a given language, a collocation has to be learned as a whole. The fact that most collocations are not predictable provides a compelling argument for their inclusion not only in text production dictionaries, but also in cognitive dictionaries because what is not

predictable must be studied consciously and explicitly. As has been shown, if collocations are studied from the point of view of predictability, they are split into two classes namely: regular and irregular collocations. Regular collocations are predictable. Assumingly, most collocations are in the second class of irregular collocations. This implies that only irregular collocations should be explicitly recorded in a general language dictionary meant for cognitive purposes. However, while predictable collocations need no explicit treatment in a general dictionary with a cognitive function because they are treated in definitions or as illustrative sentences, they should be explicitly treated in a specialised dictionary of collocations.

Regarding the selection of the headwords, there should be no difference in principle between a general dictionary and a special dictionary of collocations. In either case, the model designed in Section 5.3 for text encoding dictionaries can be implemented in a cognitive dictionary that excludes predictable or regular collocations. From a cognitive point of view, the onomasiological treatment of collocations in the articles of their keywords allows a neat representation of lexical knowledge in a dictionary because close collocations are grouped together around a single node or concept. Thus, it is undeniable that this approach has many didactic benefits for learners.

In addition, when determining the scope of the dictionary with respect to the treatment of collocations, due consideration must be given to the characteristics and needs of the intended audience. The requirements (in terms of collocations) of a dictionary for lexicographically inexperienced or less skilled users will differ from those of a dictionary prepared for more skilled and experienced users and those of a dictionary compiled for non-native speakers. Also depending on the degree of opacity of collocations, the spectrum of information provided for opaque and cultural-bound collocations should be much wider than in transparent collocations. Finally, another way of dealing with collocations in a cognitive function is by providing encyclopaedic knowledge about collocations as part of the middle matter or back matter texts. This can be useful in a text production oriented dictionary as well as in a cognitive oriented dictionary. By containing collocations, the outer text can familiarize users with the peculiarities and functioning of collocations by means of sample exercises, and thus contributing to increase the users' knowledge of this phraseology phenomenon.

## 5.6 Conclusion

The elaboration of a function-driven model for the inclusion and treatment of collocations in

Kinyarwanda lexicography was the object of discussion in this chapter. It has been shown that collocations are of more interest to text production than to any other lexicographic function. In fact, collocations pose more challenges for text production than text reception, provided they are transparent. Most of the time, the user knows the keyword; what he/she needs is the most appropriate collocator to put the keyword in syntactic configuration with respect to the conceptual and semantic representations he/she wants to express. In addition, in Kinyarwanda dictionaries, collocational meanings are often confused with intrinsic polysemes of verbs while collocations ought to be treated in the articles of keywords. This is the case for most support verbs. The consequence of this practice is the congestion of verbal articles and the risk for the user not being able to access collocational entries if he/she does not know the full form of the collocation. This seriously undermines the user-friendliness of the dictionary. The model proposed in this chapter can help mitigate this problem.

Although the text encoding dictionaries have received much attention in the first section of this chapter, it is recognised that collocations also play a certain role in text reception and knowledge-oriented dictionaries; and suggestions have been made as to how selection and recording of collocations in dictionaries fulfilling the respective functions should be done. It is expected that this chapter has provided Kinyarwanda lexicographers with decisive information as to how to appropriately record collocations in dictionaries with respect to various lexicographic functions and the needs of users.

## CHAPTER 6: THE LEXICOGRAPHIC REPRESENTATION OF KINYARWANDA PRAGMATEMES

### 6.1 Introduction

In this chapter the focus is on the lexicographic inclusion and treatment of pragmatic phrases referred to as pragmatemes according to the Meaning-Text terminology. Due to the lack of academic recognition of the concept of pragmateme in Kinyarwanda linguistics and lexicography, there is a need to clearly define it as a phraseological unit before exploring its lexicographic status along with the challenges it poses. After this introductory remark, the second section (6.2) characterises the concept of pragmatemes focusing on the definition of the notion of pragmateme in the context of the MTT (6.2.1), the role of the context of communication (6.2.2) and the conceptual representation that triggers the use of a pragmateme (6.2.3) and distinguishes pragmatemes from other phraseological expressions namely collocations and idioms, already discussed in this thesis (6.2.4). In the third section (6.3) the importance of pragmatemes as a dictionary data category (6.3.1) is discussed and the need for a systematic treatment of pragmatemes in Kinyarwanda dictionaries (6.3.2) is emphasised. Then two different models for the inclusion of pragmatemes in Kinyarwanda lexicographic works, i.e. the inclusion in the central list (6.3.3) and in the outer text (6.3.4) are developed and discussed. Some concluding remarks are drawn from the whole chapter in the last section (6.4).

### 6.2 Characterisation of pragmatemes and the importance of pragmatemes in communication

#### 6.2.1 Definition of pragmateme

Mel'čuk (In press) considers a pragmateme as “a compositional phraseme if it is restricted in its signified and its signifier by the extralinguistic situation in which this phraseme is used.” However, although pragmatemes are typically multi-lexemic in essence, their analysis cannot exclude those lexemic pragmatemes made of a single word (Papadopoulou 2010: 210; Vorobey 2011: 288). Gross modo, pragmatemes can appear in the form of:

- a single word: *Muraaho!* (Hello!) [someone greeting somebody else after a long time without meeting]
- a simple phrase: *Ntaa nzira* (no entry!, no thoroughfare!) [on a sign on a path or a fence]
- a complex sentence: *Ayo magaambo arí muu nyugúti ziberámye ní twe twaáyaanditse dútýo* (Italics ours) [as author's remark in a textual quotation]

An overview of Kinyarwanda pragmatic phrasemes allows distinguishing lexemic constructions from multi-lexemic constructions. The lexemic constructions in question have the same properties as prototypical pragmatic phrasemes, with the only difference that they consist in one full lexical unit that, out of the pragmatic context, is used by speakers in other free sentences. A good example is the pragmateme “*Twéese!*” (The same to you!), that is used in response to health wishes addressed to somebody who sneezes (*Urakire*: God bless you) or the greeting formulae “*Muraaho!*” (Hello!). Kinyarwanda is a holophrastic language, thus it is not surprising to get many morpho-semantic units in a single form. This conception is justified since, to refer to the same conceptual representation in the same situation, a given language uses a lexemic pragmateme while another indifferently uses a multi-word pragmateme (Kin: *Urakoze* versus Eng: *Thank you*) as Vorobey has a fortiori indicated (Vorobey 2011: 288).

Moreover, the communicational situation can take on the role of the missing elements (*Imiínsi myíinshi!* (Nice to see you) [a person greeting another after a long period without meeting] where the pragmateme is a N+Adj structure that in a full form could be something like <sup>#</sup>*Imiínsi myíinshi yari iíshize tútabonána nóone ndííshiimye kukó duhuuyé!* (It is nice to see you after a long period of separation). Long before pragmatics became a formalised linguistic discipline, Blinkenberg (1928: 5-6) suggests that the enunciation context plays a pivotal role in supplementing the missing information in single term phrases. He claims that in the very beginning of the development of language, single-term expressions were the singular form of expressions. It is only at a later stage of development that the language got free from the support of the context and became self-reliant. At this level, the language has a more developed syntax, but that does not mean that the single-term sentence sank into oblivion. Rather, it is still ubiquitous in everyday linguistic interactions, and surprising as this may seem the language remains essentially linked to the situation of communication. Therefore, the definition of the concept of pragmateme can be extended to lexemic pragmatemes as well.

Longer pragmatically restricted sequences exists, but because of their strong unity, these are still considered as being single pragmatemes. A representative example of the complex pragmateme is *Itégeko rihana umuuntu wéese uzíigaana iyi noóti* (Counterfeiting is punishable by law). However, there are complex pragmatemes that can be confused with the co-occurrence of two pragmatemes. For example, the greeting expressions *Murannya muranyáara!* (How is the baby?), and *Muraaho murakómá!* (Are you still alive?), are both made of two juxtaposed verbs conjugated in the same time and mode and they do not constitute pragmatemes. These pragmatemes have lost coordinating

conjunctions due to ellipsis.

However, in other cases, we can get a co-occurrence of two different pragmatemes. For example, *Nohéli Nziizá n'Umwáaka Mushyá Muhiíre!* (Merry Christmas and Happy New Year!), may appear together on a festive greeting card. Therefore, pragmatemes must be processed separately.

From this discussion it can be derived that only two concepts play a decisive role when determining what a pragmateme is. These concepts are the conceptual message one wants to verbalize and the specific situation of enunciation, respectively postulated as conceptual representation (ConceptR) and extra-linguistic or communication situation (SIT) in the MTT. More particularly, the SIT constitutes a crucial point in determining the distinction between pragmatemes from other types of phrasemes and an important data in lexicographic treatment of pragmatic phrasemes as well.

### **6.2.2 Pragmateme and situation of communication (SIT)**

The situation of communication shortened as SIT is the key concept in the study of pragmatic phrasemes. With the SIT, we refer to the extralinguistic situation in which a person is and from where and when he/she wants to express something. Therefore, the key defining condition of a pragmateme is the presence of that extralinguistic situation that determines the utilisation and interpretation of pragmatemes. The definition provided above implies that for a language user wishing to express a given ConceptR in a given communicational extralinguistic situation, only one pre-conceived signifier and one signified are possible; the said ConceptR cannot be expressed otherwise. The ConceptR picks up, so to speak, all the available information in a given SIT that can help verbalise what this information means to the SIT. In other words, the SIT determines the ConceptR, which in turn, determines the representation of the utterance.

In the production of an utterance, the SemR is consecutive to the ConceptR. The transition between ConceptR and SemR is the point of departure of phraseologisation of a phrase into a pragmateme. At this level, the ConceptR gets a specific signified ("AB") in relation to the SIT AB. On the other hand, the utterance which results from the PhonR is also formed according to the SIT AB that requires a particular signifier / AB /. In other words, the SIT AB determines what will be the meaning "AB" and the signifier / AB / of the pragmateme AB, leaving out various alternatives that the rules of the language L allows # "AB1", # "AB2", etc.

Of course, this restriction does not exclude exceptions. There are a number of ConceptR, which can be expressed by more than one pragmateme (two or three), formed in accordance with regular rules of L,

but the synonymic and variation possibilities are still limited, compared to paraphrasing possibilities and communicational flexibility characterising natural languages in general. The possible extralinguistic situation may be related to social interaction such as greetings, conversational formulae, slogans, correspondence, journalism, scientific writing, road signs and markings, public places indications and notices, label notices, civilities, requests, etc. The communicational situation has to be indicated in the dictionary, because it indicates to the user, especially the second language learner, when and how to utilise the expression.

### **6.2.3 Pragmateme and conceptual representation (ConceptR)**

The ConceptR is the informational content the user has in mind and wants to express using a pre-fabricated way. For example, for the pragmateme *Ni twe dushaatsé gutsiindagiriza* (Emphasis ours/his/mine, etc.), the conceptual representation is *aka gacé k'inyandukuuro ní kó dushaaká gutsiindagiriza* (this part of quotation is emphasised by us). Therefore, the conceptual representation along with the speech act and stylistic items are part of the comment on pragmatics of a pragmateme in a dictionary. This aspect will be addressed in sub-section 6.3.2.3.

However, it is important to bear in mind that conceptual information is not a semantic description of the pragmateme. Indeed, since the latter is semantically compositional, there is no need to define it; the meaning is what the phraseme literally means, that is the result of lexical constituents (if many) and grammatical rules governing the acceptable semantic utterances in the language in question. It is interesting to note that for many pragmatemes, the informational message has no relation with the meaning of pragmatemes in question. For example, the pragmateme *Ubiikwa kure y'írumuri* (Keep in dark place, lit.: *It is kept far from light*) expresses the conceptual representation {*You must keep this drug safe from light*}.

Obviously, the notion of conceptual representation as was well developed by the Meaning-Text linguistics, reminds us of conceptual information in cognitive linguistics. Some linguists in cognitive linguistics trends have been using it to support the hypotheses of the mind working independently from language and pre-linguistic conceptualisation of the world reality.

### **6.2.4 Pragmatemes vs idioms and collocations**

The phraseologisation of pragmatemes is made in the transition between ConceptR and SemR. In contrast, the phraseologisation of idioms takes place in the transition between deep Semantics Representation (DSemR) and deep syntactic representation (DSyntR). In other words, the

phraseological nature of pragmatemes is due to pragmatic factors, while that of idioms is due to semantic factors. However, this distinction between pragmatemes and idioms does not imply that a pragmateme cannot be an idiom at the same time; that is, in a given SIT, one can express a given ConceptR with an idiomatic pragmateme. Such is, for example, the formulae *To whom it may concern* is a idiomatic pragmateme. As argued in Chapter 3, the same is possible with collocations (e.g. *wet paint*), though the latter appear in Surface Syntactic Representation in general. Normally, the meaning "AB" of a pragmateme AB is the set of meanings "A" and "B" such as ("AB" = "A  $\oplus$  B"). Similarly, the signifier of pragmateme AB is the set of signifiers / A / and / B / such as (/ AB / = / A  $\oplus$  B /). The phraseologisation of a pragmateme, again, concerns the restrictions imposed by SIT AB to the signified "AB", which cannot be replaced by another, although the rules of the language L may provide a synonymous expression for "AB". The situation of communication prescribes the use of a specific signified and a specific signifier to express a given conceptual representation and precludes the use of any other expression which is semantically equivalent. However, the phraseological fixation of some pragmatemes may occur at the same semantic level. In that case, the pragmateme is unrestrictedly expressible and, though you cannot use an equivalent meaning, for "A  $\oplus$  B" you can choose any one of several possible quasi-synonymous expressions that the rules of the language L allow.

### **6.2.5 Importance of pragmatemes in communication**

Sinclair was among the first to outline the importance of pre-fabricated expressions in a functional perspective when presenting his idiom principle. Sinclair (1991: 110):

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation.

According to Mel'čuk (In press), however, the creation of phrasemes or what he calls phraseologisation, is mainly motivated by the insufficient number of lexemic signifiers contrary to the big and always increasing number of signified in natural languages that lead pre-existing free combinations to lose their liberty of selection to become a bound expression with a unified signified. This process is referred to as phraseologisation of former free combinations. Another important strategy - but less interesting in this framework - exploited by natural language to offset the lack of signifier is the "ambiguization" or polysemy, which is the association of many signified to one signifier (Mel'čuk, In press).

The above views help us to understand the role of pragmatemes in communication. Since their use is restricted by extralinguistic factors that do not allow other choices, it is of paramount importance to grasp a good number of pragmatemes for effective communication. Pragmatemes have many semantic and cultural nuances. Their functioning is rather complex and depends on many extralinguistic pragmatic factors. In addition, the knowledge of stereotypical utterances such as greetings, civilities formulae, proverbial expressions, signs and posters and notices on packages, is required by the initial levels of foreign language acquisition (Vorobey 2011: 290). These formulas are very critical to creating and sustaining human relationships and interactions.

In the context of intercultural communication, pragmatemes are also indispensable when for example a traveller meets another linguistic community and needs to know the exact expression to use in a given situation. Thus, a non-native speaker risks making mistakes when formulating the Kinyarwanda daily greeting expressions: *Mwiíriwe!* and *Mwaaramutse!*. These greeting formulae are differently used according to the time of the day. For example, the formulae *Mwiíriwe!* (Good afternoon!) is used from pm hours while *Mwaaramutse!* (Good morning!) is used from 05:00 to the end of am hours. In other cases, the speaker has to grasp all nuances underlying the use of pragmatemes according to the situation, relations between speakers, degree of politeness, etc.

To sum up, pragmatemes are pragmatically restricted and even native speakers may want to know the accurate formulae used in a given situation of communication for concepts that are expressed through pragmatemes. In the context of second language use, it is also useful to have the source language pragmatemes with their target language equivalents in order to facilitate the communication function inherent to bilingual dictionaries. As Gouws pointed out (2004a: 86), this is absolutely important because “cultural differences between different language groups imply that a phrase in one language cannot merely be translated on a word by word basis to produce a valid target language phrase.” The inappropriate manipulation of pragmatemes may negatively affect the communication. Moreover, the importance of pragmatemes in both oral and written text production is critical. This explains the importance of the lexicographic treatment of pragmatemes. However, the role of pragmatic phrasemes transcends the traditional reference works (dictionaries). They play a vital role in conveying pragmatic information in instructional information tools such as notices, user manuals, search engines or human-machines communication interfaces, information materials, etc. wherein the pragmatic dimension of information is of utmost importance. These operative tools (although not discussed in this research) can also be seen as consultable lexicographical tools in the broader sense of lexicography, satisfying

spontaneous operative and interpretative users' needs according to Bergenholz and Bothma (2011: 62) and Tarp (2008a: 120, 126-127). Having defined the concept of pragmateme and stressed the importance of pragmatemes in a dictionary as communicational tool and in lexicography in general, a discussion on the lexicographic treatment of pragmatic phrasemes will follow.

## **6.3 A Model for the lexicographic representation of pragmatemes in Kinyarwanda lexicography**

### **6.3.1 Preliminary considerations**

As discussed above, there is no more room for doubt concerning the need for processing pragmatemes in the dictionary. Even so, the treatment of pragmatemes in current Kinyarwanda dictionaries is not explicit. They occasionally appear as usage examples without any specific indication. Therefore, it is difficult for a dictionary user to access a pragmatic phrase.

In addition to problems linked to their identification, the inclusion and description of pragmatemes in general dictionaries are not easy at all. Even in more resourced languages where users can easily find dictionaries and phrasal books containing pragmatemes, their treatment remains poor relation of modern lexicography in general. Indeed, these works do not satisfy the users' needs since they are mostly elaborated with regard to reception, rather than text production. In addition, Blanco (2011) laments the fact that they do not give the necessary information for adequate usage and correct comprehension of pragmatemes. Gouws (2006: 29) also mourns the underestimation of cultural phrases or pragmatic phrases in dictionaries and insists on the fact that when considering the inclusion of cultural data in a dictionary, pragmatic phrasemes should draw the lexicographers' attention not because they may contain cultural-bound lexemes to exemplify, but "on account of the fact that cultural information can be retrieved from them", as pragmatic phrasemes. Besides Gouws, the need for the explicit treatment of pragmatemes in general language dictionaries as well as in lexicographic tools with these phrases as the only subject matter, has been expressed by different scholars in linguistics (Fonagy 1982; Martins-Baltar 2000) and some lexicographers have included routine formulae in their dictionaries as part of the back or middle matter, although contemplative and transformative theoretical investigation of the lexicographic treatment of pragmatemes is recent (Blanco 2010; Gouws 2006; Papadopoulou 2012; Vorobey 2011).

Regarding the lexicographic treatment of pragmatemes, we must note that unlike the idiom, the pragmateme is not considered a lexical item deserving its own article in the central list structure of a dictionary, although it is a phraseological item having a conceptual message, a signifier and syntax. The

internal argument structure of a pragmateme does not allow elevating the pragmatemes to the level of a first level treatment unit within the macrostructure. Thus a question arises: "If pragmatemes cannot be presented as primary level treatment units in the central list, how should they be treated lexicographically?"

Two options are provided for this problem. The first approach viewing the central list as the ideal venue for pragmatic phrases has been directly discussed by Gouws (2006) and implicitly by Blanco (2010) and Papadopoulou (2012) in the MTT perspective. The second option appeals to the trans-textual functional approach, namely to present pragmatemes within the integrated back matter or middle matter, (Gouws 2000, 2004a, 2006, 2007a; Gouws & Prinsloo 2005; Steyn 2004). It is developed in the framework of the textual approach to lexicography (Gouws 2000; Hausmann & Wiegand 1989; Kammerer & Wiegand 1998) and the functional theory of lexicography (Bergenholtz & Bothma 2011; Bergenholtz & Tarp 2003; Tarp 2000).

### **6.3.2 Pragmatemes as part of the central list**

The article-internal treatment of pragmatemes raises three main questions with respect to the choice of the guiding element, the type of the general microstructure and the scope of the description of pragmatemes. These problems relate to the selection of the guiding element, the appropriate type of microstructure, and the scope of the content of the microstructure. The following paragraphs address these issues through theoretical insights from lexicography and linguistics.

#### **6.3.2.1 How should the guiding element under which a pragmateme is listed be selected?**

The first issue relates to the choice of the lemma under which a given pragmateme should be treated. A prevailing idea is that in a general dictionary, a pragmateme should be stocked under the article of the lemma representing its headword (lexical anchor). From cognitive and text reception perspectives, this lemma should be the singular pragmateme itself when it is a single word or one of the lexical constituents of the pragmatemes that is culture-marked. In other words, the singular pragmateme is included in the article of a headword taken from it. e.g.:

***kuRAMUKa 1 Kumara ijoro ugafatiira umuûnsi udâpfuuyê*** (Iri jaambo barikoreesha bâramutsa mu gitôndo cyâangwâ bâseezera nîmugôrooba cyâangwâ nîjoro). Merewe nâabâ sîinzî kô ndamûka. Waaramutse? Uraramuke.

In the above article of the lemma *kuramuka*, the cultural value of the lemma is indicated at the first subcomment on semantics because it does not cover all the senses of the lemma. Besides the paraphrase of meaning, an additional comment in brackets indicates that this lexical item is used when

greeting people in the morning or present farewell wishes in the evening. Then the pragmatic remark is followed by three illustrative items of which the last two are pragmatemes, i.e. *Waaramutse?*: Good morning and *Uramuke*: Goodbye (evening). It is important to note that these pragmatemes do not receive treatment data providing their pragmatic value because the latter is given in the preceding pragmatic comment.

However, some pragmatemes may contain no cultural-bound or pragmatic words. In this case, the choice of the keyword is not culturally or pragmatically oriented; the pragmatemes are treated as mere illustrative entries in a dual addressing procedure, i.e. functioning as lemmatic addressing non-lemmatic addresses. For example, the lemma *kubá* has different subcomments on semantics in DIMO. The first subcomment on semantics contains numerous phrasemes, mainly idioms and collocations with one pragmateme following default illustrative free phrases. These expressions function at once as lemmatic addressing entries and non-lemmatic addresses in that they are co-text entries vis-à-vis the lemma sign and the subject of treatment on the right side. Thus, the pragmateme *Urabehó* and its plural polite form *Murabehó* (Goodbye, if not going to see for a while) are provided with pragmatic value as conventional farewell wishes used at leave-taking. In a similar fashion, the first subcomment on semantics of the lemma *umunsi* host the pragmateme *Umuúnsi mwiza* (have a nice day). Unfortunately, a comprehensive coverage of pragmatic phrases cannot be achieved using this casual approach to the presentation of pragmatemes in the dictionary (cf. Gouws 2006: 29). Thus, the comment on semantics of the lemma *Nohéli* (Christmas day) does not include any season's greeting expression containing this culture-bound lexical item, e.g. *Nohéli nziizá* (Merry Christmas!). This lemma is marked with regards to culture and therefore the lexicographers should endeavour to indicate its culture-bound nature using the appropriate label and providing relevant cultural phrases.

However, although pragmatemes represent a valuable asset in cognitive function and to a certain extent in text reception activities, the main interest for pragmatemes lies more in text production. From a text production orientation, the headword does not necessarily appear in the pragmateme in question, but is a lexical item that expresses the speech act intended by the user and often the situation of communication of the pragmateme. For example, the keyword of the pragmateme *Uhéegereye waapfá* (Danger of death) in a situation of communication [warning on an electric cabin or pole] would be the lexical items *umubúuro* (warning) expressing the speech act of warning. Similarly, the pragmateme *Ni twe dushaatsé gutsíindagiriza* (Italics ours) with ConceptR *{this part of quotation in question is emphasised by ourselves}* and the situation of communication [emphasis in a quotation] should be

listed under the keyword *gutsíindagiriza* (to emphasise).

Sometimes, as is the case in *Ni twe dushaatsé gutsíindagiriza* (Italics ours) the lexical item expressing the speech act expressed by the pragmateme, may be found in the pragmateme in question. In the pragmateme *Umugórooba mwitzá* (Good evening), the key words should be *guseezera* (saying good bye), *kwíifuuriza* (wishing) or merely the word *umugórooba* (evening) that expresses the situation of communication in which the pragmateme is usually used. It is important for the convenience of the users that this pragmateme be recorded in the articles of the lemmata representing these three words. If such procedure is well explained in the user guide, from the point of view of text production, it is easier to access a pragmateme via the lemma expressing its speech act or its situation of communication than it is when the pragmateme is recorded in the article of a headword taken from it irrespective of its communicative value.

The pragmatic value of pragmatemes is temporally and spatially fixed; it has nothing to do with its compositional meaning and its lexical constituents. Rather, it is related to the situation of communication and speech act of the pragmateme. Therefore, a linguistic knowledge of any lexical item that appears in a given pragmateme provides the user with a cue to produce that pragmateme even if he/she has never learnt it and memorised it as such. Since the situation of use of pragmatemes is extralinguistic, their representation and access in a dictionary should involve extralinguistic dimensions to guarantee a successful retrieval of information. Therefore, an onomasiological presentation of pragmatemes in the articles of lemmata expressing their speech acts and/or situations of communication is the most satisfactory approach to article-internal presentation of pragmatemes with regards to text production needs.

### **6.3.3.2 Which type of microstructure is appropriate for article-internal inclusion of pragmatemes?**

Once the issue regarding the selection of the guiding element is solved, at least theoretically, the following problem concerns the type of microstructure and where precisely in the microstructure pragmatemes should be accommodated. To the best of my knowledge, this issue has found a satisfying answer in the suggestion by Gouws to provide cultural phrases in a special text slot attached to the end of the article displaying a semi-integrated microstructure (Gouws 2006: 33-34). Such a semi-integrated microstructure has two different microstructural components. The first default component contains the relevant (sub-) comments on semantics and the respective co-textual entries. This is a normal integrated microstructure. Then the foregoing component is supplemented by a non-integrated component

attached to the end of the last sub-comment on semantics and its co-textual entry. In this way, pragmatemes should be regarded as secondary level treatment units attached to the article of the conceptual main lemma through a nesting procedure; nesting because the alphabetical order will certainly be violated given the formal discrepancies between the guiding element and the pragmateme. Elevating pragmatemes to such a level is in line with the following recommendation formulated by Gouws (2006: 29, 33):

Although not lexicalised these phrases are more than mere illustrative examples presented as co-text entries performing a lemmatic addressing procedure. (...) Cultural phrases need to be elevated to treatment units which are the targets of procedures of non-lemmatic addressing, confronting the user with more than just their status as cultural entries.

In comparison with the co-textual entries of the default microstructural component, the pragmateme accommodated in the second textual block of the semi-integrated microstructure could qualify as a non-integrated co-textual entry elevated to the status of a second level treatment unit had the main lemma used as guiding element been taken from the pragmateme in question. In the same condition and depending on whether text reception is taken into account, an article-internal cross-reference system could be developed to explicate the semantic relations between the pragmatemes in the non-integrated component and the relevant sub-comments on semantics in the preceding component displaying the integrated microstructure (Gouws 2006: 33).

What is said above presupposes similarities between the pragmateme and the lemma representing the guiding element of the article. However, presenting pragmatemes under the article of the lemma taken from the pragmateme is subject to debate because of the following reason. The lexicographic sublemmatisation of pragmatemes in a monolingual dictionary makes sense only when pragmatemes are envisaged from a text production perspective. In this perspective, the lemma used as guiding element does not necessarily occur in the pragmateme, it is only a conceptual guiding element. The user is likely to know the meaning of the conceptual guiding element, because in the pre-consultation phase, the situation of communication and the conceptual representation which have prompted the dictionary consultation are formulated in consideration of the lemma taken as the conceptual guiding element of the pragmateme. In other words, the selection of the guiding element is determined by the conceptual message and the situation of communication, not by the formal aspect of the pragmateme. For example, the Kinyarwanda pragmateme *Murakóze* (thank you) is formally a conjugated form of the verb *gukora* (to do). However, it is conceptually related to the verb *gushiiimira* (to thank) which should be the guiding element of the text slot accommodating this thanksgiving formula in the Kinyarwanda

text production dictionary, as far as the article-internal treatment is concerned.

Thus, the recommendation by Gouws (2006: 33) regarding the provision of article-internal cross-references between the default integrated microstructure and the phrases in the attached text block is not applicable in the model developed in this chapter, due to its text production orientation. Pragmatically, the pragmatemes are non-compositional albeit that they are semantically compositional phrasemes. Therefore, a pragmateme could be used as an integrated illustrative entry only when it is meant to illustrate a pragmatic sense of the lemma, or when its pragmatic nature is marked via appropriate labels as shown above in the lexicographic articles of the lemmata *kuramuka* and *umuúnsi*.

Regarding the cluster-internal ordering procedure, instead of applying an alphabetical arrangement within the nested text block, it is suggested that the pragmatemes are arranged according to their frequency of use, starting with the most frequent one when the additional text slot contains more than one pragmateme. In addition, a special structural indicator (Cult, Prag, etc.) should precede the sub article stretch containing nested pragmatemes. In a Kinyarwanda dictionary, the structural indicator **UKO BAVUGA** (How to say) or merely **BAVUGA** (It's said...) should be used while waiting for an appropriate term for the concept of pragmatemes. A similar suggestion has been made in the previous chapter in relation to sublemmatisation of idioms.

### **6.3.2.3 What should the comment on pragmatics contain?**

The sub-article headed by a pragmateme should contain two compulsory data categories: a comment on the form which is the pragmateme itself and a comment on pragmatics (instead of a comment on semantics). Attention is given to the comment on pragmatics in the following section.

However, a preliminary terminological remark is necessary with regard to the terms: comment on semantics and comment on pragmatics, the latter being introduced in the previous paragraph. According to Nkomo (2008: 126), the term 'comment on semantics' refers to "data categories which seek to offer users an understanding of the *meaning of the lemma* and consequently its *usage*" (Italics mine). In the framework of the general theory of lexicography, the comment on semantics includes both semantic and pragmatic data since pragmatic explanations are a kind of lexical paraphrase. Therefore, in the article-internal presentation of pragmatemes, the latter remain within the scope of the comment on semantics of the lemma sign functioning as guiding element. In this respect data provided as treatment of a specific pragmateme as comment on semantics could entail a confusion with the overall comment on semantics within which the pragmateme and the data provided about it are

positioned. In this context the term comment on pragmatics is more precise and specific.

Whereas the term comment on semantics is a superordinate term encompassing both semantic and pragmatic data, I reserve the use of the term ‘comment on pragmatics’ for those pragmatic data provided to explain the usage of a pragmateme presented as a non-lemmatic address and treatment unit in the article structure of the lemma functioning as either a formal or conceptual guiding element. A comment on pragmatics is thus a subset of a comment on semantics, i.e. a sub-section of the comment on semantics dedicated to a pragmatic description of a pragmateme within the article structure of the lemma sign used as guiding element. In this study, the term comment on pragmatics is not used – and it should not be understood – as a concurrent of the term ‘comment on semantics’ which has a broader spectrum of data, but rather in a complementary way, as a way of designating a particular category of data dealt with in the comment on semantics.

In a monolingual dictionary, the comment on pragmatics of a pragmateme does not contain a lexical paraphrase of the expression. Because a pragmateme is semantically compositional, its semantic meaning is obvious and it does not need semantic treatment. What is not compositional is the pragmatic modalities of the pragmateme (where, when and how it is used). That is what the comment explicates. To exemplify this contrast, the proposed comment provided on the phrase “*Pleased to meet you*” is: “an expression used when you have just met someone” (Gouws 2006: 34). There is nothing about the semantics of this expression in this comment except the pragmatic context indicating when it is used.

Thus, the term ‘comment on pragmatics,’ is appropriate in this case because it precisely and specifically characterises the type of data which are given as way of treatment of the pragmateme article-internally (Geeraerts 2003: 86-87). In a bilingual dictionary, the comment on pragmatics is confined to the target language equivalent of the pragmateme of the source language and eventual comment on the cultural gap between the source language pragmateme and the target language equivalent where deemed necessary. In a monolingual dictionary, typical data categories in the comment on pragmatemes include the situation of communication, the speech act, the language register or style level and the conceptual representation. The article slot containing these data can be complemented by a co-text entry exemplifying the pragmatemes in use, but as a pragmateme is in most cases a complete sentence, this is not always required.

In the comment on pragmatics, the indication of the situation of communication (SIT) is indispensable since it determines, together with the speech act, whether a phrase is a pragmateme or a free phrase.

The situation of communication restricts the pragmateme in question to the conceptual message and the intended speech act. It provides two types of information: where or when a pragmateme is restrictedly used and the language style (written/oral). The following data that is very critical in the comment of pragmatics, concerns the speech act: to congratulate, to give advice, to warn, to greet, to thank, to ban or to prohibit, etc. For example, the act of speech of the note *Ubiikwa ahó abáana batagerá* (Keep out of reach of children) on the package, is advising; the pragmateme *Birabújijwe kunywéera itaábi hánó* (No smoking) is meant to prohibit while *Hánó hari imbwá iryáana* (Beware of the dog) expresses a warning, and so on. In most cases, if not all, the speech act corresponds with the word chosen as the guiding element of the main article.

In addition, the indication of the language register (formal, informal) that provides precision about the context of pragmateme use is part of the microstructure. Difference of language registers must be taken into account if one wants to explain or understand the raisons d'être behind variants in pragmatemes. Other good examples of register-based variations are *Bite byaawe?* and *Amakuru yanyu?* (How are you?), where the first is labelled as vulgar or informal while the second sounds formal. However, the same conceptual representation can be expressed through different pragmatemes or variants due to different situations of communication. For instance, *Umwáaka mushyá!* and *Muzáagiré Umwáaka Mushyá Muhiíre!* (Happy New Year!), are variants expressing the same conceptual message but in different situations, since the former occurs in both oral and written communications, while the latter usually occurs only in writing, more specifically on greeting cards.

The last but still important type of information contained in the microstructure of a pragmateme is the conceptual representation (ConceptR) of the pragmateme (a reflection of the reality that the speaker wants to express). As stressed earlier, this conceptual representation or message is not the paraphrase of meaning of the pragmateme in question, but a pre-linguistic conceptualisation of the reality in the mind of the speaker. What matters in a dictionary of pragmatemes is not its semantic definition, but the mental message a pragmateme is meant to express.

This description will not be complete without providing an example of how the proposed process functions in a lexicographic modelling of pragmatemes. I present the development of an article for the pragmateme *Tegereza gató, nyaábuneka* (Please hold the line) by which I highlight how the above-mentioned aspects are taken into account. As mentioned above, we begin with an indication of the communication situation and the keyword. This pragmateme is used in a telephone conversation when a receptionist or a switchboard operator requests the caller to hold on to the call until he/she has

succeeded in establishing a connection between the caller and the person he/she wants to talk to, reassuring him/her that he/she is working to make that connection. Therefore, the keyword in Kinyarwanda is the word *terefoóne* (telephone) and the SIT can be summarized as follows: [*mu kigaaniiro kurí terefoóne* (in a telephone conversation)]. With respect to the speech act, the overall pragmatic intention of the use of this pragmateme is to request the caller to be patient until the required connection has been established. Therefore, the speech act in this case is *gusaba* (to request).

With regard to this pragmateme, special attention should be given to the issue of language register and variants. It is thus necessary to indicate the register and introduce the variants (if any) of *Tegereza gató*, *nyaábuneka*, in this instance. To be specific, the pragmateme in question is used in formal language but it is also informally used without the ending word *nyaábuneka* (please), i.e. the form *Tegereza gató*. When used in formal language, the speaker expresses an attitude of respect towards the interlocutor (*kíubaha*: formal) while in colloquial setting the expression of respect is not required. Both the speech act and the language express the user's attitude towards the interlocutor. For example, had the user chosen the former phrase expressing an informal tone, the speech act could have been an order instead of a request. Lexicographers should be aware of this interrelation. Finally, the conceptual representation concludes this analysis. In this pragmateme, the communicational intention of the user is *{ntuúkupe ndimó ndagerageza kugúhuza n'úuwó ushaaká kuvúgiisha}* (do not hang up until I finish to establish the connection or to put you through to the person you want to talk to). Thus, I have a complete sub-article of the pragmateme that appears under the semantic comment of the lemma *terefone* under the structural marker **BAVUGA**, as below. Apart from the pragmateme described above and its variant, I have extended the pragmateme text slot to seven items. It is possible that some of them could be presented as pragmatic items in the non-integrated part of the microstructure in the articles of lemmata functioning as their headwords, as has been discussed above. Such a dual distribution of pragmatemes will ensure an optimal access to pragmatic data and will make the dictionary polyfunctional so that it satisfies not only dictionary consultation needs arising during text production tasks, but also those occurring in text reception contexts.

### **TEREFONE /terefoóne/ <Far téléphone**

**Icyuma bavugiramo amajwi akagenda mu mikwege cyangwa akanyura mu kirere ikindi cyuma kikayakirira kure.** Iyo ngira terefone mba mvuganye n'incuti yanje ikora i Rubavu.  
**BAVUGA**

**1 Tegereza gató nyaámuneka:** *Invugo yiyubashye ukoresha kuri terefone usaba uhamaagaye kudakupa mu gihe urimo ugerageza kumuhuza n'uwo ashaka.* [Hold on please.]

**2 Mugume ku murongo nyamuneka:** *Invugo yiyubashye ukoresha kuri terefone usaba uhamagaye kudakupa mu gihe urimo ugerageza kumuahuza n'uwo ashaka.* [Hold the line please.]

**3 Tegereza gato,** Imt 1: *Invugo itiyubashye cg y'abamenyeranye ukoresha kuri terefone usaba uhamagaye kudakupa mu gihe urimo ugerageza kumuahuza n'uwo ashaka.* [Hold on please.]

**4 Mwiiháangaane turihó turagerageza kubáhuuza n'úuwó muhamágaye:** *Invugo ikoreshwa n'imashini isubiza terefone isaba uhamagaye kwihangana mu gihe bagishakisha uwo ahamagaye.* [Sorry. We are trying to connect you/to put you through.]

**5 Ndavugana nande?:** *Invugo yiyubashye ukoresha kuri terefone ushaka kubaza uhamagaye izina rye iyo utamumenye* [Who is this?]

**6 Nyabuneka wavuga witoonze?:** *Invugo yiyubashye ukoresha usaba uwo muvugana kuri terefone kugabanya umuvuduko w'amagambo akavuga yitonze.* [Speak more slowly please.]

**7 Hari uri kubahamagara:** *Invugo uwakira terefoni rusange akoresha amenyesha uwo bashaka ko hari ushaka kumuvugisha hifashishijwe uburyo bw'ihuzamiyoboro cyangwa kuri terefone imwe.* [There is a call for you.]

One of the conclusions that can be drawn from the foregoing discussion is that the article-internal lexicographic treatment of pragmatemes in a text production dictionary is the best alternative compared to the current unsystematic treatment of pragmatemes. I have shown how such an inclusion policy could work in the lexicographic treatment of pragmatic phrasemes if it were to be implemented in the existing Kinyarwanda dictionaries. The proposed model is also as general as possible in order to be generalised to other languages and to various types of dictionaries. In spite of this, however, it is necessary to point out that the inclusion of pragmatemes in the central list does not solve all problems, as far as the access to pragmatemes is concerned.

Concerning the treatment of pragmatemes within the microstructure, some doubts remain as to whether the users will be able to access them swiftly. Firstly, this is due to the fact that it may still require higher reference skills and linguistic knowledge to guess the main lemma under which the pragmateme is presented, since the main lemma is not always a word taken from the lexical constituents of the pragmateme as was argued in the production-oriented model. Secondly, due to the potential dissimilarities between the guiding element and the pragmateme, it will not be possible to secure the continuity of the vertically alphabetical ordering of the main lemmata and sublemmata (including pragmatemes) included in the central list. This failure to apply the alphabetical arrangement will diminish the speed of access to pragmatemes via the outer access structure of a dictionary. Another criticism of this model is that the creation of a zone for pragmatemes within the microstructure could result in some articles being too long and dense, thus complicating the dictionary consultation process. Because all these problems stress the shortcomings of the listing of pragmatemes in the microstructure, it would be advisable to explore the potentials of the outer texts in the treatment of pragmatemes. Notwithstanding the foregoing, the above-sketched micro-architecture of the articles of pragmatemes

remains valid.

### **6.3.3 A thematic treatment of pragmatemes in outer texts**

#### **6.3.3.1 Rationale for the inclusion of pragmatemes in outer texts**

The theory of a dictionary as carrier of lexicographic texts referred to in this research posits that when designing the data distribution programme of a given dictionary the lexicographer should make use not only of the central list but also of the outer text sections of the dictionary for data presentation and treatment to ensure that the function(s) and genuine purpose of the dictionary are achieved. In this line of thought, Gouws (2010: 60) argues that

The planning of the data distribution should not only be concerned with the data to be presented in the central list but also with data, regarded by the lexicographer as important to the intended target users, which cannot be accommodated in the default articles of the dictionary. The front and back matter sections offer the lexicographer the opportunity to include this data and present it in one or more outer texts of the dictionary. In a general learners' dictionary, the outer texts can play a vital role in the transfer of data, especially to support the user in getting familiar with systematic aspects of the specific language, e.g. its grammar and its pronunciation, but also *cultural values, etc.* (*Italics mine*)

As far as integration into the genuine purpose of the dictionary is concerned, Gouws and Prinsloo (2005: 59) note the distinction between two types of outer texts: integrated outer texts and non-integrated outer texts. Integrated outer texts are the outer texts that contribute in one way or another to the accomplishment of the genuine purpose of dictionary. On the one hand, they help the user retrieving the needed information from the dictionary (e.g. abbreviations, grammatical information, guide to the dictionary) or by accommodating some lexicographic data which would be regarded as macrostructural data but that lexicographers find appropriate to place in outer texts depending on the function and the type of dictionary or on the nature of data being presented (e.g. some multiword items, sub-lexical items such as prefixes and suffixes, proper names, etc.). Pragmatemes fall within the scope of integrated function-adhering outer texts, as exemplified by an increasing number of monolingual and bilingual communication-oriented dictionaries using the back matter or middle matter as a venue for pragmatic phrases (Gouws 2000, 2004a, 2006, 2009a).

Several reasons make the outer texts the best venue for the inclusion and treatment of pragmatemes. The main one is that the addressing structure is simpler and clearer than in the treatment in the central list. The user does not have to wander through the dictionary when looking for a needed pragmateme. He/she is not obliged to use his/her intuition to guess where the pragmateme is placed. Rather, taking a quick glance at the table of contents suffices to direct the user to where the desired data can be found.

However, in order to achieve this goal the user needs to know that the table of contents has to be consulted and this demands some kind of dictionary culture that is normally acquired with time and through use. The second advantage of having the pragmatemes treated in the same area of the dictionary is that this zone becomes a language acquisition tool, especially in the second language-learning context. It adds a cognitive dimension to the prevailing function of the dictionary, elevating the latter to a poly-functional status. For a foreign speaker, the acquisition of pragmatemes is a prerequisite in building his/her communicational skills in the L2 because most of them are routine phrases used in everyday conversations (Vorobey 2011: 290). Furthermore, unlike the treatment in the central list, an outer text treatment of pragmatemes helps mitigate dense articles and remote addressing related problems. The inclusion of pragmatemes in the back matter of the dictionary constitutes in my opinion the best way to follow. However, in accordance with the transtextual functional approach (Gouws 2004a, 2007a), pragmatemes could also be given as co-text and/or sublemmata entries in the central list but with a pragmatic label/marker and a cross-reference marker guiding the user to the relevant back matter text in order to make them polyaccessible (Gouws: personal communication, 9/07/2013). This dual access makes it possible for the user to reach the pragmatemes through two search roads. Although the back matter text containing pragmatemes is the default venue of the dictionary consultation procedure, the user may also reach a given pragmateme by looking up the headword under which that pragmateme is given in the central list through the procedure of article-internal presentation. From that entry, a cross-reference will also guide the user to the relevant back matter section where the pragmateme is treated; cf. Gouws (2006: 27).

### **6.3.3.2 Why the onomasiological presentation?**

In this research, I consider the pragmatemes from a text production perspective. In that perception and with respect to pragmatemes, the user consults the dictionary in order to find a pragmateme expressing the conceptual message he/she has in mind, in the situation in which he/she wants to communicate. In this context, the alphabetical arrangement is not appropriate for the listing of pragmatemes. Even in a dictionary meant for providing assistance in translation (bilingual dictionaries), the pre-consultation process follows the same pattern: the user locates the source language pragmateme in the dictionary using the same search criteria as in a monolingual dictionary, then the user looks up its equivalent.

The onomasiological or conceptual listing of pragmatic phrases facilitates the consultation procedure in both monolingual and bilingual dictionaries. In this connection, the lexicographers have to apply a thematic macrostructure. This approach is followed in many dictionaries with communication

functions, e.g. the *Language in use* section in Robert and Collins (2002) and the *Useful phrases according to function* in *Oxford-Duden German Dictionary* (Scholze-Stubenrecht & Sykes 1999) referred to by Gouws (2004a: 86) and the routine formulae back matter text in the *Deutsch-Madagassisches Wörterbuch* (cf. Gouws 2009a: 81-82), just to mention a few examples. It is now worth analysing how the lexicographers should proceed when encoding pragmatemes in the back matter following a thematic arrangement principle.

The number of themes may differ from one language to another and from one culture to another. Similarly, some themes may be underdeveloped in one language while they are well furnished in the other, but most of them do not change across the languages. Below is the sketch of the proposed thematic arrangement of Kinyarwanda pragmatemes, i.e. main themes, groups and sub-groups without pragmatemes. This is followed by a detailed presentation of the theme *kuburira* (warning), as illustration of the proposed presentation.

### **Kwifuriza (wishing)**

- Urugendo rwiza (Safe journey)
- Kugaruka amahoro (Safe return)
- Iminsi mikuru y'umwaka (Season's greetings)
- Ibihe rusange (General periods' wishes)
  - Ijoro ryiza (Good night)
  - Umunsi mwiza (Nice day)
  - Umugoroba mwiza (Good evening)
  - Igitondo cyiza (General) (umunsi, ijoro, umugoroba, akazi)
  - Wikendi nziza (Good week-end wishes)
- Kuryoherwa (Wishing a good meal)
- Isabukuru (Birthday wishes)
- Koroherwa (Get well wishes/quick health recovery)
- Amahirwe mu kazi (Good luck in career wishes)
- Gutsinda (Good luck in exam wishes)
- Kubaho (Wishing long life)
- Ubuzima bwiza (Wishing good health)
- Amahoro (Wishing peace)
- Ubukwe bwiza (Wedding wishes)
- Kwibaruka (Congratulations for new parents)
- Gutunga no gutunganirwa (Wishing happiness and prosperity)

### **Gushimira (thanks)**

- Ibyiza (For good)
- Ibibi (ironically for a misdeed)
- Mu musozo w'ibaruwa (concluding a letter)
- Gutega amatwi disikuru (concluding a speech)

### **Kurahira (swearing)**

- Ku mana (by god)

Ku masano (by the name of a relative/parent)  
Ku wakugabiye (by the name of the chief))  
Ku mwami (by the king)  
Kuri Yesu (by Jesus)

### **Kwihanganisha (sympathising/ condolences)**

Kwihanganisha (sympathising)  
Guhumuriza (comforting)  
Mu gusoza ibaruwa/ubutumwa (ending a letter)

### **Kuramutsa/Gusuhuza (greeting)**

Abo mudaherukanye (if have not seen each other in a while)  
Abo muherukanye (if have seen in a while)  
Mu gitondo (in the morning)  
Ku manywa (at day)  
Abamenyeranye (very familiar, with close friends)

### **Gusezera (Saying goodbye, farewell expressions)**

Abo mutazongera kubonana vuba (if not going to see each other again for a while)  
Abo mukunda kubonana (hoping to meet again someday/one day)  
Ku manywa (at day/during the day)  
Nimugoroba (in the evening)  
Abamenyeranye (with acquaintances)

### **Kuburira (warning)**

Ku ndangabintu (on goods labels)  
Imitin'ibijyanye na yo (on drugs)  
Ibimeneka (breakable goods)  
Ibihumanya (poisonous/harmful products)  
Ibiribwa (food stuff and beverage)  
Ahantu (on places)  
Ahari amashanyarazi (electricity poles and cabins)  
Imirase (source of radiation)  
Ibyapa byo ku mhanda (road signs)  
Ahantu harinzwe (protected property, access-restricted place)  
Ahasize irangi ribisi (freshly painted surface)  
Ku nyandiko z'ubutegetsi (on official documents)  
Ku noti (on bank notes)  
Ku nyandiko ndangamuntu (on identity documents)  
Ku mabwiriza y'ubutegetsi (on official/administrative instructions)

### **Gutumira (inviting)**

Ku meza (to the table)  
Ubukwe (to a wedding)  
Inama (to a meeting)

### **Gusaba (requesting)**

Imvugo zo gusaba mu kinyabupfura (general polite formulae)  
Gusaba imbabazi (apologising)  
Kwsegura (apologising for a misdeed)  
Kwsegura (apologising anticipatively)  
Gutegereza (request for waiting)  
Kuri terefone (on telephone)  
Kuri mudasobwa (in computer dialog box)

Ku biro (on the office door)  
Uruhushya/uburenganzira (permission)  
Serivisi (request for a service)  
Amakuru (request for information)

**Gutegeka (commanding, order)**

Gucecekesha (to silence, to hush, to shut up, to prohibit from speaking)  
Ku nyandiko z'imiti (on drug packaging leaflet/package)

**Kubuza (prohibiting from)**

Kwinjira ahantu harinzwe (entering an access-restricted area)  
Guhagarara ahatemewe (bad parking)  
Kunyura mu kanyatsi (to walk on the lawn)  
Gusakuza mu isomero (talking in the library)  
Kwinjirana igikapu mu iduka/isomero (bringing a bag into a shop/library)  
Kwanduza ubwiherero (letting a toilet unclean)  
Gusakuza kwa muganga (making a noise in the hospital)  
Kunywera itabi ahatemewe (smoking in public or on dangerous sites)

**Gutsiindagiriza (emphasising)**

Mu mvugo (in oral communication)  
Mu nyandiko (in writing)

**Kumenyesha (announcements)**

Ivuka (announcing the birth of a baby)  
Kubika (obituary)  
Kuranga ibigurishwa/ibikodeshwa (For sale/rent, on advertisement notice on a car, land, house, etc.)

**Guha ikaze (welcoming)**

Guha ikaze abashyitsi, abatumirwa, abakumva (welcoming the visitors, the invitees, the audience)  
Utanga ikiganairo (welcoming a speaker)

**Kwitaba (replying to a call)**

Terefone (answering the telephone)  
Interlocutor (answering Present!)

**Gusezeranya (promising)**

Mu kwamamaza (in advert)  
Igihembo (of generous reward)  
Ukomeje (more assertively)

Sixteen themes have been identified but this number is neither exhaustive nor static; it is likely to evolve along with an empirical data collection. Each of them is a speech act relevant to the specific context of communication and conceptual message. Each main theme is divided into different groups which may also be divided into smaller sub-groups.

### 6.3.3.3 Trial of thematic arrangement of selected pragmatemes with “warning” as speech act

I use the theme *KUBURIRA* (warning) to illustrate what a conceptual treatment of pragmatemes could look like in both monolingual and bilingual dictionaries of Kinyarwanda.

- Treatment of “warning” pragmatemes in a Kinyarwanda monolingual dictionary

## **KUBURIRA (WARNING)**

### **Ku ndangabintu (on goods' labels)**

IMITI (medicines/drugs)

- ◊ *Ubikwa aho abana batagera* (Keep out of the reach/sight of children)
- ◊ *Ubikwa kure y'ubutote n'urumuri* (Keep away from humidity and light)
- ◊ *Ubikwa muri firigo* (Keep refrigerated)
- ◊ *Usigwa ku mubiri* (For external use only)

IBIRIBWA N'IBINYOBWA (food stuff and drinks)

- ◊ *Bikoresha bitarengeneje* (Best before)
- ◊ *Byakorewe* (Product of/Made in)

IBINTU BYANGIZA UBUZIMA (harmful goods)

- ◊ *Itabi ririca* (Caution, smoking is harmful to your life)

IBINTU BIMENEKA UBUSA (breakable goods)

- ◊ *Terura neza* (Fragile, handle with care)

### **Ahantu (On places)**

KU BYAPA BYO KU NZIRA (on road signs)

- ◊ *Akayira kamatungo* (Cattle crossing)
- ◊ *Nta nzira* (No entry/no thoroughfare)
- ◊ *Inzira irafunze* (Diversion)
- ◊ *Imirimo!* (Caution, work in progress!)
- ◊ *Hanyurwa n'abahatuye* (Only inhabitants are allowed)

KU BYAPA AHANTU HARINZWE (on protected/access-restricted area)

- ◊ *Hano hari imbwa iryana* (Beware of the dog)
- ◊ *Hararinzwe* (Protected)
- ◊ *Birabujije kwinjira hano utahakora* (Unauthorized entry forbidden/ No entry without authorization)

KU BYAPA AHANTU HARI AMASHANYARA / IMIRASE  
(electricity/radiation)

- ◊ *Uhegereye wapfa* (Danger of death)

KU BYAPA AHO AHO BASIZE IRANGI (freshly painted area)

- ◊ *Ntiwegame!* (Wet paint)

### **Ku nyandiko z'ubutegetsi (on official documents)**

KU NOTI (on bank note)

◊ *Itegeko rihana umuntu wese wigana iyi noti* (Counterfeiting is punishable by the law)

KU NDANGAMUNTU (on identity card/passport)

◊ *Uzayikoresha binyuranyije n'amategeko azahanwa* (Whoever uses it contrary to the law will be punished)

KU MABWIRIZA Y'UBUTEGETSI (on administrative instruction)

◊ *Uzabirengaho azahanwa n'amategeko* (The contravention is punishable by the law)

The English equivalents and translations given in the brackets in this example are for the sake of comprehension for non-Kinyarwanda speakers; it should not appear in a mono-functional monolingual dictionary. The two following examples show how the same theme should be represented in different sections of a bilingual Kinyarwanda – English dictionary directed at text production in either language. The first section is directed at text production in English by Kinyarwanda speakers while the second is directed at text production in Kinyarwanda by English speakers.

#### a) Kinyarwanda – English section for texts production in English

In order to help the user to produce the English form when being familiar with the Kinyarwanda form, this section starts with the Kinyarwanda pragmateme and then the English equivalent.

#### **KUBURIRA (WARNING)**

##### **Ku ndangabintu (On notices)**

##### **IMITI (DRUGS)**

◊ Ubikwa aho abana batagera

= ***Keep out of the reach/sight of children***

◊ Ubikwa kure y'ubutote n'urumuri

= ***Keep away from humidity and light***

◊ Ubikwa muri firigo

= ***Keep refrigerated***

◊ Usigwa ku mubiri

= ***For external use only***

##### **IBIRIBWA N'IBINYOBWA (FOOD STUFF AND BEVERAGES)**

◊ Bikoreshe bitarengeje

= ***Best before***

◊ Byakorewe

= ***Product of/Made in***

##### **IBINTU BYANGIZA UBUZIMA (HARMFUL PRODUCTS)**

◊ Itabi ririca  
= ***Caution, smoking is harmful to your life***

#### IBINTU BIMENEKA UBUSA (FRAGILE OBJECTS)

◊ Terura neza  
= ***Fragile, handle with care***

#### Ahantu (Places)

#### KU BYAPA BYO KU NZIRA (ON ROAD SIGNS)

◊ Akayira k'amatungo  
= ***Cattle crossing***  
◊ Nta nzira  
= ***No entry/no thoroughfare***  
◊ Inzira irafunze  
= ***Diversion***  
◊ Imirimbo!  
= ***Caution, work in progress!***  
◊ Hanyurwa n'abahatuye  
= ***Only inhabitants are allowed***

#### KU BYAPA AHANTU HARINZWE (ON PROTECTED PROPERTIES)

◊ Hano hari imbwa iryana  
= ***Beware of the dog***  
◊ Hararinzwe  
= ***Protected***  
◊ Birabujijwe kwinjira hano utahakora  
= ***Unauthorized entry forbidden/ No entry without authorization***

#### KU BYAPA AHANTU HARI AMASHANYARAZI / IMIRASE (ON ELECTRICITY OR RADIATION-EMITTING SITES)

◊ Uhegereye wapfa  
= ***Danger of death***

#### KU BYAPA AHO AHO BASIZE IRANGI (ON PAINTED AREA)

◊ Ntiwegame!  
= ***Wet paint***  
**Ku nyandiko za Leta (Official documents)**

#### KU NOTI (BANK NOTES)

◊ Itegeko rihana umuntu wese wigana iyi noti  
= ***Counterfeiting is punishable by the law***

#### KU NDANGAMUNTU (IDENTITY CARD)

◊ Uzayikoresha binyuranyije n'amategeko azahanwa  
= ***Whoever uses it contrary to the law will be punished***

## KU MABWIRIZA Y'UBUTEGETSI (OFFICIAL INSTRUCTIONS)

◊ Uzabirengaho azahanwa n'amategeko  
= *The contravention is punishable by the law*

### b) English – Kinyarwanda section for texts production in Kinyarwanda

In order to help the user to produce the Kinyarwanda form when being familiar with the English form, this section starts with the English pragmateme and then the Kinyarwanda equivalent.

## 14. WARNING (KUBURIRA)

### 14.1 On labels (Ku ndangabintu)

#### MEDICINES(IMITI)

◊ Keep out of the reach/sight of children  
= *Ubikwa aho abana batagera*  
◊ Keep away from humidity and light  
= *Ubikwa kure y'ubutote n'urumuri*  
◊ Keep refrigerated  
= *Ubikwa muri firigo*  
◊ For external use only  
= *Usigwa ku mubiri*

#### FOOD STUFF AND DRINKS (IBIRIBWA N'IBINYOBWA)

◊ Best before  
= *Bikoresha bitarengengeje*  
◊ Product of/Made in  
= *Byakorewe*

#### HARMFUL GOODS (IBINTU BYICA)

◊ Caution, smoking is harmful to your life  
= *Itabi ririca*

#### BREAKABLE GOODS (IBINTU BIMENEKA)

◊ Fragile, handle with care  
= *Terura neza*

### On places (Ahantu)

#### ON ROAD SIGNS (IBYAPA BYO KU MIHANDA)

◊ Cattle crossing  
= *Akayira k'amatungo*  
◊ No entry/no thoroughfare  
= *Nta nzira*  
◊ Diversion  
= *Inzira irafunze*  
◊ Caution, work in progress!

= ***Imirimo!***

◊ Only inhabitants are allowed

= ***Hanyurwa n'abahatuye***

#### ON PROTECTED/ACCESS-RESTRICTED AREA (AHANTU HARINZWE)

◊ Beware of the dog

= ***Hano hari imbwia iryana***

◊ Protected

= ***Hararinzwe***

◊ Unauthorised entry forbidden/ No entry without authorisation

= ***Birabujijwe kwinjira hano utahakora***

#### ELECTRICITY/RADIATION (AMSHANYARAZI/IMIRASE)

◊ Danger of death

= ***Uhegereye wapfa***

#### FRESHLY PAINTED AREA (AHASIZE IRANGI RIBISI)

◊ Wet paint

= ***Ntiwegame!***

#### **On official documents (Ku nyandiko z'ubutegets)**

#### ON BANK NOTE (INOTI)

◊ Counterfeiting is punishable by the law

= ***Itegeko rihana umuntu wese wigana iyi noti***

#### ON IDENTITY CARD/PASSPORT (INYANDIKO NDANGAMUNTU/Z'INZIRA)

◊ Whoever uses it contrary to the law will be punished

= ***Uzayikoresha binyuranyije n'amategeko azahanwa***

#### ON ADMINISTRATIVE INSTRUCTION (AMABWIRIZA Y'UBUTEGETSI)

◊ The contravention is punishable by the law

= ***Uzabirengaho azahanwa n'amategeko***

Pragmatemes are sometimes used in situations of communication which do not necessarily involve translation, but immediate text production. This inclusive model can assist translators; however, the primary function is to assist users in text production rather than in translation. Concerning text translation, for instance the translation of a notice text or a legal text containing pragmatemes, the treatments presented in (c) and (d) are suitable for translation of pragmatemes from English into Kinyarwanda and vice versa.

#### c) English – Kinyarwanda section for texts translation from English into Kinyarwanda

#### **14. WARNING**

## On labels

### MEDICINES

- ◊ **Keep out of the reach/sight of children**  
= Ubikwa aho abana batagera
- ◊ **Keep away from humidity and light**  
= Ubikwa kure y'ubutote n'urumuri
- ◊ **Keep refrigerated**  
= Ubikwa muri firigo
- ◊ **For external use only**  
= Usigwa ku mubiri

### FOOD STUFF AND DRINKS

- ◊ **Best before**  
= Bikoreshe bitareengeje
- ◊ **Product of/Made in**  
= Byakorewe

### HARMFUL GOODS

- ◊ **Caution, smoking is harmful to your life**  
= Itabi ririca

### BREAKABLE GOODS

- ◊ **Fragile, handle with care**  
= Terura neza

## On places

### ON ROAD SIGNS

- ◊ **Cattle crossing**  
= Akayira kamatungo
- ◊ **No entry/no thoroughfare**  
= Nta nzira
- ◊ **Diversion**  
= Inzira irafunze
- ◊ **Caution, work in progress!**  
= Imirimo!
- ◊ **Only inhabitants are allowed**  
= Hanyurwa n'abahatuye

### ON PROTECTED/ACCESS-RESTRICTED AREA

- ◊ **Beware of the dog**  
= Hano hari imbwa iryana
- ◊ **Protected property**  
= Hararinzwe
- ◊ **Unauthorized entry forbidden/ No entry without authorization**

= Birabujijwe kwinjira hano utahakora

## ELECTRICITY/RADIATION

◊ **Danger of death**

= Uhegereye wapfa

## FRESHLY PAINTED AREA

◊ **Wet paint**

= Ntiwegame!

## on official documents

## ON BANK NOTE

◊ **Counterfeiting is punishable by the law**

= Itegeko rihana umuntu wese wigana iyi noti

## ON IDENTITY CARD/PASSPORT

◊ **Whoever uses it contrary to the law will be punished**

= Uzayikoresha binyuranyije n'amategeko azahanwa

## ON ADMINISTRATIVE INSTRUCTION

◊ **The contravention is punishable by the law**

= Uzabirengaho azahanwa n'amategeko

## d) Kinyarwanda – English section for texts translation from Kinyarwanda into English

### 14. KUBURIRA (WARNING)

#### Ku ndangabintu (on labels)

##### IMITI (medicines)

◊ **Ubikwa aho abana batagera**

= Keep out of the reach/sight of children

◊ **Ubikwa kure y'ubutote n'urumuri**

= Keep away from humidity and light

◊ **Ubikwa muri firigo**

= Keep refrigerated

◊ **Usigwa ku mubiri**

= For external use only

##### IBIRIBWA N'IBINYOBWA (food stuff and drinks)

◊ **Bikoreshwe bitarengeje**

= Best before

◊ **Byakorewe**

= Product of/Made in

##### IBINTU BYANGIZA UBUZIMA (harmful goods)

◊ ***Itabi ririca***

= Caution, smoking is harmful to your life

IBINTU BIMENEKA UBUSA (breakable goods)

◊ ***Terura neza***

= Fragile, handle with care

**Ahantu (On places)**

KU BYAPA BYO KU NZIRA (on road signs)

◊ ***Akayira k'amatungo***

= Cattle crossing

◊ ***Nta nzira***

= No entry/no thoroughfare

◊ ***Inzira irafunze***

= Diversion

◊ ***Imirimo!***

= Caution, work in progress!

◊ ***Hanyurwa n'abahatuye***

= Only inhabitants are allowed

KU BYAPA AHANTU HARINZWE (on protected/access-restricted area)

◊ ***Hano hari imbwa iryana***

= Beware of the dog

◊ ***Hararinzwe***

= Protected

◊ ***Birabujijwe kwinjira hano utahakora***

= Unauthorized entry forbidden/ No entry without authorization

KU BYAPA AHANTU HARI AMASHANYARAZI / IMIRASE  
(electricity/radiation)

◊ ***Uhegereye wapfa***

= Danger of death

KU BYAPA AHO AHO BASIZE IRANGI (freshly painted area)

◊ ***Ntiwegame!***

= Wet paint

**Ku nyandiko za Leta (on official documents)**

KU NOTI (on bank note)

◊ ***Itegeko rihana umuntu wese wigana iyi noti***

= Counterfeiting is punishable by the law

KU NDANGAMUNTU (on identity card/passport)

◊ *Uzayikoresha binyuranyije n'amategeko azahanwa*  
 = Whoever uses it contrary to the law will be punished  
 KU MABWIRIZA Y'UBUTEGETSI (on administrative instruction)

◊ *Uzabirengaho azahanwa n'amategeko*  
 = The contravention is punishable by the law

The presentations displayed in (c) and (d) can also be regarded as hybrid treatments because beside the translation of pragmatemes from English into Kinyarwanda and from Kinyarwanda into English, such sections can support text production in English and Kinyarwanda respectively.

#### **6.3.3.4 Observations on the inclusion of pragmatemes in outer texts and access structure related issues**

It can be argued that the treatment proposed in (c) and (d) displays a remote access structure for the translator. The translator has before his/her eyes the source language pragmateme, whereas he/she only wants its equivalent in the target language. To facilitate the access, the lexicographer should complement the treatment in example 3 by an alphabetical list of pragmatemes whereby each pragmateme is provided with its equivalent in either language or a number indicating the theme and group under which it is treated in detail. The same type of list should be used when text production in L2 by users with low level of language proficiency is at the forefront. In fact, users without a high command of the second language tend to think in their own language, which means that they endeavour to access foreign language pragmatemes via their counterparts in their native language, thereby doing a kind of translation. Such a list can help the user rapidly accessing the L2 equivalent of a pragmateme he/she knows in L1 if that equivalent exists in the dictionary. Notwithstanding the foregoing, the importance of the complementary list should not be overestimated because it may lead the user to restricting him/herself to a variant which may have escaped the attention of the compilers while the main text provides many alternatives to the problem to be solved. We must conclude that there is no single recipe for easy retrieval of information. The compilers could and should make use of various structures in order to ensure unproblematic access to pragmatic phrases and easy retrieval of the needed information.

To this end, besides the main text, the back text of pragmatemes must contain a table of contents and a concise introductory text. If the outer text giving pragmatemes occurs in a bilingual dictionary, it should be placed either as a back matter or a middle matter directed at the dictionary as a whole or as a back matter text in the secondary frame structure of each word list (cf. Gouws 2004a: 71-72). In addition, each back matter text containing pragmatemes must contain its own introduction and table of

contents. Whether the option of middle matter text is taken, it will be more user-friendly to present the respective tables of contents and introductions in parallel on the same pages at the beginning of the middle matter because the content of the two sections is parallel in general.

The table of contents will guarantee a rapid outer access structure to the themes as containers of pragmatic phrases. The themes or speech acts which are frequently used should precede those that have a restricted use in Kinyarwanda (e.g. greeting versus commanding, wishing versus swearing, etc.). The introduction will help users familiarise themselves with the inner structure and the purpose and content of the outer text. Moreover, the use of running heads and a particular colour for the borders or edges of pages should not be underestimated.

The inner access structure of the text can be implemented by means of an elegant and user-friendly presentation of data. The lexicographer should consider the appropriate use of font styles and structural indicators. For example, the main theme should be written between the lines in bolded capitals, preceded by the number of order. The groups within one theme should be highlighted using bold typeface, a particular font colour and a two-level number of order (e.g. 14.1, 14.2., 14.3) while sub-groups should be framed. The pragmateme should be introduced by a special non-typographical structural marker, a vertical or horizontal lozenge for instance, and written in bold. Where equivalents or translations are provided, it should be underwritten with(out) an equivalence symbol.

In order to ensure the participation of this outer text to the achievement of the overall lexicographic function of the dictionary and the poly-accessibility to pragmatemes, pragmatic-bound entries of the central list should be linked to the relevant pragmatemes through cross-references. To this end, the cross-reference item should indicate the number of the theme and group under which the relevant pragmateme is listed. In this way, the lexicographer can ensure that the pragmatemes' section is poly-accessible because the user can access them either via the central list address to the back/middle matter or via a direct consultation of the outer text.

The recording of pragmatemes in the back matter section has many advantages provided the lexicographer indicates clearly in the front matter how to access them and presents them as user-friendly as possible. For the lexicographer to help the user become familiar with the pragmatemes in the back matter, the latter should have an introduction and a table of contents as secondary outer-texts. In order to help the user getting a quick grasp of the text layout, the introduction should be as clear and concise as possible. The function of the table of contents should be to outline the structure of the text

and help the user reach as swiftly as possible the section of interest to him/her. It is up to the lexicographer to determine the themes of the content of a back matter text for pragmatemes, taking into account the user profile and the function of the back matter. However, to facilitate the user in accessing the data, the titles of sections and sub-sections (if applicable) should as much as possible reflect different situations of communication or conceptual representations that pragmatemes are meant to express, e.g. thanks, greetings, wishes, congratulations, ending of letters, requests, etc. Within a section/sub-section, the ordering should be based on the frequency of usage, the dialogue structure and/or the language register, from the most formal pragmatemes to the most informal ones. The lexicographer should also supply stylistic labels to the pragmatemes where relevant (e.g. informal, rare). In so doing, the lexicographer will provide the dictionary users with a valuable text production auxiliary that will contribute to the proper use of Kinyarwanda.

## 6.4 Conclusion

To sum up this chapter, both the phraseological status and the lexicographic recording of Kinyarwanda pragmatic phrasemes were explored. The concept of pragmateme has been defined using linguistic principles stemming from the MTT before assessing it from a lexicographic perspective. The objective was to know whether the pragmatemes deserve a place in a dictionary alongside other phrasemes, namely idioms, collocations and proverbs. In this connection, a lack of scholarly interest in Kinyarwanda lexicography was revealed. Because cultural phrases, including pragmatemes, play a role in achieving the genuine purpose and satisfying the lexicographic functions of the dictionary, the need to have pragmatemes recorded in the dictionaries to meet various users' needs was highlighted. It was noted that the lexicographic formalizing of pragmatemes requires guiding principles stemming mainly from the trans-textual functional approach to lexicography. Therefore, with focus on text production and to a lesser extent on translation and cognitive related functions, two models for the lexicographic representation of pragmatemes was proposed, implemented and evaluated. The first approach favours a presentation of pragmatemes in the central list, in a non-integrated component of an article displaying a semi-integrated microstructure, using a procedure of nesting. In this model, pragmatemes are targets of procedures of non-lemmatic addressing. The second approach consists in presenting pragmatemes in the back matter section in monolingual dictionaries and in the middle matter section in bilingual dictionaries. The ultimate objective was to propose a theoretically motivated and user-driven inclusion and treatment policy of Kinyarwanda pragmatemes.

Since the two approaches have both weaknesses and strengths, one has to weigh up the pros and cons.

On balance, treating pragmatemes as part of an integrated function-adhering outer text is more effective than doing it within the central list, as far as easy retrieval of information by users with low reference skills is concerned. However, in a monolingual and multifunctional dictionary, the two models are not necessarily incompatible. They can rather be combined to make the pragmatic data poly-accessible and poly-functional. To mitigate the shortcomings of the alphabetical principle in the outer text, the thematic arrangement of pragmatemes was proposed as the best arrangement approach of pragmatemes in the outer-text as far as the focus is on communication-oriented functions. However, the two approaches are not mutually exclusive in bilingual lexicography where the role of the alphabetical list of pragmatemes complementing the onomasiological list is recognised.

## CHAPTER 7: GENERAL CONCLUSION

### 7.1 Introduction

By way of conclusion, this chapter maps out the major results of this thesis and assesses whether the thesis has answered the research questions raised in the general introduction. Besides providing a summary of the main results, a review of the research problem with respect to research findings was made. Finally, the limitations of this study was highlighted and it was indicated where further lexicographic research should be conducted in order to make a comprehensive evaluation of lexicographic aspects not covered in this study and to ensure that Kinyarwanda lexicographers put the user at the centre of dictionary compilation in view of better dictionaries in the near future.

### 7.2 Summary of research findings

The ultimate goal of this research was to formulate a typology of Kinyarwanda phrasemes and to design lexicographic models for the treatment of the respective types of phrasemes. The major findings of this research have been presented in seven chapters.

In Chapter 1 the foundation was laid by defining the research problem, the motivation, the purpose, the delimitation and the hypotheses of the research. This chapter has also been used to expose the methodological and theoretical approaches that have come to be employed in this research. The first chapter contains a background to the sociolinguistic profile of Kinyarwanda, stressing its geo-political and demographic importance as trans-border language and it is stressed that in such a context, the importance of dictionary compilation cannot be ignored. The chapter ends with a synoptic view on the content of the thesis.

Chapter 2 has been dedicated to the overall theoretical framework of this research. It recognises that a user-friendly representation of lexicographic data in a dictionary demands a strong commitment of dictionary compilers to the application of insights from theoretical lexicography. It has presented in the first place the function theory of lexicography as a plausible approach to data presentation in a dictionary. The concept of lexicographic functions has been emphasised and was then used in formulating function-driven inclusion policies for different types of phrasemes. Some concepts of the structural theory of dictionaries have been reviewed with special focus on the access structure, the frame structure, the macrostructure, the microstructure and mediostructure which are crucial aspects in metalexicography. In the discussion of the frame structure, the role of outer texts in the achievement of the genuine purpose of the dictionary and fulfilment of lexicographic function(s) was highlighted. This

led to the discussion of the concept of a functional trans-textual approach (Gouws 2007a) which has been very effective in this research. This helped our understanding of how various structural components of a dictionary should be utilised to render phraseological data more accessible than they currently are in the context of Kinyarwanda lexicography. These structures have been negotiated in formulating the proposed models in Chapters 4 to 6.

In Chapter 3 the concept of phraseme is defined and a critical review of the treatment of phrasemes in Kinyarwanda linguistics and lexicography is carried out. This chapter has revealed that Kinyarwanda phraseology is idiom-biased. This is confusing and has caused that different categories of phrasemes are seen as one. For example, it has been shown that pragmatemes are ignored; and collocations, that are mistakenly under-represented in dictionaries, are included as idioms or merely as illustrative items. A fresh linguistic perspective on the concept phraseme and a better knowledge of its wide-ranging nature would play a decisive role in the formulation of suitable inclusion policies specific to different types of phrasemes. In addition, despite the cautions against the use in lexicography of insights from linguistics, it has been shown in this chapter that, everything in its place, the role of linguistic insights in defining and categorising phrasemes should not be overlooked in the general lexicography.

Furthermore, it was made clear that, in addition to confusing various data categories, the general lexicography of Kinyarwanda lacks theoretical lexicography-driven policies for the inclusion and treatment of phrasemes. Among the problems identified is the lack of lexicographic principles for the selection of the guiding element for the sublemmatisation of phrasemes. The overall observation that stands out from the investigation into representative general dictionaries of Kinyarwanda is the use of the second constituent in recording phrasemes (generally not the verb). Listing the phrasemes in the microstructure of the main lemma is the most generalised method, the distribution of phrasemes to various senses of the main lemma and mixing the idioms with other data categories have also been observed. The lack of an inclusion policy for phrasemes and inconsistency in the application of the set principles in some dictionaries, particularly concerning the distribution of idioms in various sub-comments on semantics was noted. The gap of knowledge in the lexicographic representation of Kinyarwanda phrasemes has been defined and the overall motivation for undertaking the research provided.

Chapters 4 to 6 provided an instance of Gouws's argument according to which the structures devised for a dictionary "should not be imposed on a dictionary but they should result from the functions and the genuine purpose of that dictionary" (Gouws 2010: 60). In an effort to design a lexicographic

treatment model of phrasemes, the concepts of lexicographic functions and dictionary structures were brought to the fore.

In Chapter 4 the focus was on the lexicographic treatment of Kinyarwanda idioms and proverbs. In this chapter, I have developed a model according to which idioms can be treated as either primary level treatment units or secondary level treatment units. The former option consisting in straight lemmatisation of idioms has been strongly recommended for text production dictionaries because it makes access to idioms easier in this situation. In addition, Kinyarwanda idioms always have a lexical word in first position, which facilitates the lemmatisation. The second approach, the sublemmatisation of idioms using a niching procedure and the idiom's opening word as guiding element, has been discussed in the context of the text reception function. The challenges lexicographers could face if they envisage either of the approaches and how to overcome them were discussed. Furthermore issues related to the description of idioms according to various lexicographic functions were examined.

With proverbs, neither the straight lemmatisation of proverbs in the central list nor the microstructural treatment of proverbs as illustrative sentences as viable approaches for the treatment of proverbs in Kinyarwanda was considered, due to the shortcomings and user-unfriendliness of these approaches. However, it was recommended that the frame structure of the dictionary should preferably be employed to present proverbs in general dictionaries. Furthermore the moving and treating of proverbs in the outer texts, precisely in the back matter section or in a section created for this purpose as second section of the central list was suggested. The thematic ordering for text production purposes was proposed while the alphabetical listing should be used for text reception and to a lesser extent for text production purposes (for polyfunctional dictionaries). In addition, the microstructural listing of proverbs as part of a semi-integrated microstructure was explored. However it was recommended that to be fully profitable to users, i.e. to make proverbs poly-accessible, this approach should be complemented by an additional alphabetical list of proverbs positioned in the back matter and directing the user through a cross-reference system, to the articles in which proverbs are described in the central list. Nevertheless, the decision to implement this model should be taken with due regard to the consequence on the structure of articles in order to avoid dense article and data congestion.

In Chapter 5 a function-driven model for the lexicographic treatment of Kinyarwanda collocations was formulated using relevant insights from metalexicography. Early in the chapter theoretical clarifications on the concept of collocation in phraseology were provided. Then the treatment of collocations in dictionaries was addressed and suggestions were made as to the inclusion and eventual description of

collocations in different situations including text production and translation, text reception and cognitive situations. The macrostructure, the microstructure, the mediostructure along with dictionary functions emerged as lexicographic structures and tools which should play a pivotal role in the lexicographic representation of Kinyarwanda collocations. The treatment of collocations in the microstructure of the constituent taken as the base of the collocation (onomasiological approach) is firmly asserted while recognising at the same time the need for double recording in order to make collocations poly-accessible. Therefore, an appropriate cross-referencing system should be used in the article of the collocator to direct the dictionary user to the article of the base where the collocation containing that collocator is treated in detail. In order to offset the lack of visibility and the underrepresentation of collocations in Kinyarwanda dictionaries, it was shown how to distinctly include collocations as separate data categories in the microstructure instead of mixing them up with free phrases used as illustrative items.

Drawing on the typology of collocations exposed in Section 5.2.3, special recommendations regarding the inclusion of collocations in dictionaries meant for text reception or those responding to cognitive needs were made. Unlike dictionaries with text production as the sole function, collocations that are deemed necessary should be included in a semasiological way, in the articles of the collocators in dictionaries designed for text reception. In Chapter 5 it is recommended that collocations should be included in this kind of dictionary when treating the senses of support and achievement verbs that are not definable irrespective of the bases they stand with in collocations, predictable collocations in which the collocator expresses an intrinsic property of the basis, opaque collocations containing single distribution words or idioms as collocators, and when treating cultural-bound lexical items. In a cognitive-oriented dictionary, it is suggested to include only unpredictable collocations using an onomasiological inclusion approach as applied in a text reception dictionary. The focus on collocations in Kinyarwanda general language dictionaries should not eschew the importance of this type of phrasemes in LSP lexicography of Kinyarwanda and anywhere else, where the concept of collocation plays a crucial role (Jousse, L'Homme, Leroyer & Robichaud 2011). The model developed in this chapter for the treatment of collocations can be adapted to Kinyarwanda LSP dictionaries and improved if necessary.

In Chapter 6 the lexicographic treatment of pragmatemes were studied. At first, the concept of pragmateme was defined and its importance in communication underscored, especially in text

production, hence the need for an explicit treatment of pragmatemes in Kinyarwanda dictionaries. Drawing on theoretical lexicography and to a certain extent on the Meaning-Text Theory, the second section was employed to develop a two-fold model for the treatment of pragmatemes in Kinyarwanda dictionaries. The first model suggests recording pragmatemes in the central list, within the microstructure of the word taken as the guiding element of the pragmateme. This model is referred to as article-internal treatment of pragmatemes. Within the context of this approach, three fundamental questions were considered: (1) How should the guiding element under which a pragmateme is recorded be chosen? (2) Which type of microstructure is appropriate for article-internal inclusion of pragmatemes? (3) What should the comment on pragmatics contain? Solutions were provided and the article-internal treatment of pragmatemes was evaluated in a trial article.

The second model consists of a thematic recording of pragmatemes in the outer text, generally in the back matter section of a monolingual dictionary or in the middle matter in a bilingual dictionary. A sketch of the thematic structure of Kinyarwanda pragmatemes was proposed based on a corpus of pragmatemes compiled by the researcher in the framework of this thesis. The main themes correspond to relevant speech acts that are performed through pragmatemes. Using the theme *kubúurira* (to warn), I have assessed the potentials of this model for text production in monolingual, and for text production and text translation in bilingual dictionaries. This study concluded that a model favouring the pragmatically based thematic presentation of pragmatemes as a back matter (or middle matter) text is by far the most compelling lexicographic inclusion policy compared to the article-internal treatment model. Notwithstanding the abovementioned findings, these two approaches are not necessarily mutually exclusive since combining them in a monolingual multifunctional dictionary makes it poly-functional and poly-accessible.

Finally, in Chapter 7 the major findings of this thesis were summarised and the research questions reviewed before concluding that further research is recommended.

### **7.3 Research hypotheses, questions and contribution of the research results**

Regarding the hypotheses and questions that underpinned this study, the above-highlighted findings have satisfactorily corroborated all the initial assumptions and affirmatively responded to the questions formulated in the research proposal.

The first hypothesis according to which theoretical lexicography offers innovative tools and

suggestions the application of which could enable a user-friendly presentation of phrasemes in general dictionaries of Kinyarwanda, has been verified and confirmed. In this regard, this research demonstrates that, depending on the underpinning lexicographic functions, the lemmatisation of idioms as primary level treatment units or niched secondary level treatment units is feasible and more user-friendly than including them as nested secondary level treatment units or mere co-textual entries.

Regarding collocations and pragmatemes, they are phrasemes in their own right and due to their importance in text production and in language acquisition related needs; they qualify for a systematic inclusion in the dictionary in accordance with the relevant lexicographic functions and users' needs. In Chapter 5 and 6 the importance of collocations and pragmatemes in responding to both communicational and cognitive needs and the formulation of a function-driven model for the lexicographic representation of Kinyarwanda collocations and pragmatemes respectively were addressed. The concept of frame structure and the transtextual functional approach (Gouws 2007) led us to envisage the presentation of some types of phrasemes beyond the boundaries of the central list. In this framework, the outer text section of the dictionary proved to be the best venue for the presentation and treatment of Kinyarwanda proverbs and pragmatemes. Such a use of back matter texts will be an innovation in the general lexicography of Kinyarwanda because so far a back matter section has never been used in any general dictionary of Kinyarwanda. The use of outer texts was respectively discussed in Chapters 4 and 6.

Therefore, using theoretical insights from both functional and structural theories of lexicography outlined in Chapter 2 and considering the weaknesses identified in Chapter 3, Chapters 4 to 6 consists of proposals of different models for the presentation and treatment of Kinyarwanda phrasemes to be used in order to improve existing dictionaries or to create better ones. In this way, the following two research questions have been answered: *1) What adequate lexicographic principles can lexicographers rely on when formulating informed policies for the inclusion and description of various types of Kinyarwanda phrasemes, in order to ease their accessibility in Kinyarwanda dictionaries and ensure an unproblematic retrieval of information? 2) How could those insights from theoretical lexicography and particularly various structural parts of a dictionary be used in order to accommodate phrasemes in a manner which is compatible with the requirements of each type of phraseme and ensuring easy access to phraseological data in accordance with the specific situation of use or dictionary function?*

Furthermore, while recognising the importance of metalexicography in providing the guiding principles for a user-driven representation and treatment of phrasemes in Kinyarwanda dictionaries, this study

claims that the linguistic typology of phrasemes plays a certain role in identifying and defining the corresponding lexicographic data categories or data types and that the lexicographic representation models to be designed will be helpful as long as they take cognisance of the wide-ranging nature of phrasemes. In so doing, the third research question – *Could insights in linguistic approaches to phraseology contribute to lexicographic representation of Kinyarwanda phrasemes?* – was addressed and affirmatively answered.

Although researchers in theoretical lexicography must first subscribe to a consistent and pure theory of lexicography, they can also use insights from linguistics to solve problems to which theoretical lexicography is not cognisant, when formulating principles and guidelines for practical lexicography (Andersen & Nielsen 2009). For instance, in Chapter 3 it was underscored that the non-recognition of the multifaceted nature of Kinyarwanda phrasemes leads to abusive generalisations such as those related to the selection of the guiding element on the basis of compositional principle while in fact not all phrasemes are compositional. In this study, a linguistic-based classification of phrasemes into four major types (idioms, proverbs, collocations and pragmatemes) has been adapted for lexicographic purposes and used as a point of departure for model designing. This MTT-inspired classification has enabled the formulation of a precise and type-specific lexicographic representation model, as different types of phrasemes fulfil different lexicographic functions and lend themselves to different representation procedures in general dictionaries. It is the hope of the researcher that the proposed lexicographic approach that separately treats different types of phrasemes is better than the current one that makes no distinction between these differences.

However, although this study has made recourse to a linguistically inspired classification of Kinyarwanda phrasemes, it has been done only for the sake of the appropriate treatment of each category. Yet, this research remains in essence metalexicographic. Without relying solely on theoretical linguistics, this thesis has made it clear that using linguistic-driven classifications is not only feasible but also rewarding in dictionary compilation in the context of Kinyarwanda. In this way, it assumes a relative complementarity between lexicography and theoretical linguistics (cf. Rundell 2012). Therefore, while emphasising how lexicography can benefit from theoretical linguistics, it provides a theoretical contribution to the practical lexicography of Kinyarwanda and African languages in general. Above all, it contributes to existing discussions aimed at opening up and increasing the adaptability of lexicographic theories to new contexts (Andersen & Nielsen 2009; Bergenholz 1995; Bowker 2010; Chabata 2009; Hurskainen 2003; Mochiwa 2007). Notwithstanding this, linguistics is not the only

discipline likely to contribute to the development of reference works as any contribution depends on the subject matter of the dictionary and its situation of use. Theoretically, insights from linguistics are more likely to play a vital role in communication and cognitive situations (general dictionaries and specialised dictionaries) than in interpretative and operative situations (Bergnholtz & Bothma 2011; Tarp 2008a) where insights from the information science may be critical.

## 7.4 Concluding remarks and way forward

So far, the lack of a clear-cut classification of Kinyarwanda phrasemes and non-application of principles and tools of metalexicography has impaired the user-friendly lexicographic representation of this subset of language in existing Kinyarwanda dictionaries. Effort has been made in this thesis to formulate lexicographic treatment models for the specific types of Kinyarwanda phrasemes, referring to relevant lexicographic theories. In one way or the other, this dissertation has demonstrated that Kinyarwanda lexicography needs to base its decisions on sound theories of lexicography in order to present dictionary data in the perspective of potential users, while welcoming insights stemming from sound linguistic theories in order to solve linguistic problems that impede the appropriate treatment of the language.

With respect to the issues at stake, the results of the study summarised in this chapter are important. First and most important, the implementation of the formulated representation models will make it easier for users to access different types of phrasemes, according to the function prevailing in a dictionary. Secondly, the recognition of different types of phrasemes will assure a systematic coverage of phraseological units during the phase of data acquisition and processing and a type-driven treatment of phrasemes in the lexicographic work. I take this opportunity to call on the compilers of Kinyarwanda dictionaries to apply and adapt the suggestions contemplated in this research when planning new dictionaries or revising existing dictionaries.

However, lexicographers of Kinyarwanda should bear in mind that the significance of the results is confined to the lexicographical representation of Kinyarwanda phrasemes, the subject matter of this research. In fact, this investigation has not addressed in depth many lexicographic issues that are not relevant to phraseology. For example, when reviewing current dictionaries of Kinyarwanda, it was found that the user-perspective is underestimated in most Kinyarwanda dictionaries. Thus, it is difficult for users to access data presented in Kinyarwanda dictionaries and retrieve lexicographic information, even for the most educated people. The principles underpinning the compilation of Kinyarwanda

dictionaries should be revisited in the light of relevant lexicographic theories, such as those referred to in this research. Therefore, studies aimed at devising theoretical function-oriented models for specific dictionaries of Kinyarwanda are desperately needed not only because the present study has addressed only the phraseological subset of the language, but also because it was not targeting a tangible group of users in a specific situation, but different hypothetical situations of use and dictionary functions. In addition, whereas it is acknowledged that corpus-based lexicography improves significantly the description of lexicographic data, one of the problems facing any dictionary compiler in Kinyarwanda is the lack of lexicographic corpora to be used as primary source of lexicographic data. Future research should also focus on building computerised corpora for Kinyarwanda that would constitute the basis of material acquisition for future dictionaries.

Ultimately, despite its ambitious undertaking, this work may have raised more questions than provided answers, especially because problems were approached in a new way and some theories used are still at the stage of initiation in the lexicography of Kinyarwanda. However, it has opened a line of thought for further development and research in Kinyarwanda lexicography.

## BIBLIOGRAPHY

### Dictionaries

- Chimhundu, H. (Ed.). 2001. *Duramazwi Guru ReChiShona*. Harare: College Press.
- DIEL : Bizimana, S. & Kayumba, C. et al. 2010. *Inkoranya iciriritse y'ikinyarwaanda*. Butare: IRST.
- DIFO : Coupez, A., Kamanzi, Th., Bizimana, S., Sematama, G., Rwabukumba G. & Ntazinda C. et al. 2005. *Dictionnaire Rwanda-Rwanda et Rwanda-Français*. 3 volumes. Butare/Tervuren : IRST/MRAC.
- DIMO : Bizimana, S. & Kayumba, C. et al. 2011. *Inkoranya y'ikinyarwaanda mu kinyarwaanda*. Butare : IRST.
- DIMO : IRST (2002) *Inkoranya y'ikinyarwaanda mu kinyarwaanda*. 4 volumes. Kigali : Régie de l'Imprimerie Scolaire.
- Hurel, E. 1926. *Dictionnaire Français-Runyarwanda / Runyarwanda-Français*. Kabgayi.
- Jacob, I. 1983, 1985, 1987. *Dictionnaire rwandais-français - extrait du dictionnaire de l'INRS*. 3 volumes. Kigali : Régie de l'Imprimerie Scolaire.
- Kabagema, E. 2011. *Amuga y'ubucamanza Ikinyarwanda-Igifaransa-Icyongereza. Inkoranya y'amuga iciriritse*. Butare : IRST.
- KED : Nikuze, E. et al. (In progress) *Inkoranya y'Ikinyarwanda n'Icyongereza/Kinyarwanda-English Dictionary* (draft).
- Scholze-Stubenrecht, W. & Sykes, J.B. (Eds.). 1992. *Oxford-Duden German Dictionary*. Oxford: Oxford University Press.
- Schumacher, P. 1956. *Dictionnaire Phonétique Runyarwanda-Français/Français-Runyarwanda*. Kabgayi.

### Other literature

- Alberts, M. 2007. *Dictionary structures and dictionary using skills*. Afrilex, TUT, Soshanguve.
- Andersen, B. & Nielsen, S. 2009. "Ten Key Issues in Lexicography for the Future". In Bergenholz, H., Nielsen, S. & Tarp, S. (Eds). *Lexicography at crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. 355-363. Bern: Peter Lang.
- Balteiro, I. 2007. *A Contribution to the Study of Conversion in English*. Münster: Waxmann.
- Bastin, Y. et al. 1999. *Continuity and Divergence in the Bantu Languages: perspectives from a lexicostatistic study*. Tervuren: Musée Royal de l'Afrique Centrale.
- Benson M., 1990. Collocations and General-purpose Dictionaries. *International Journal of Lexicography* 3 (1): 23-34.

- Bergenholtz, H. 1995. "Linguistic information". In Bergenholtz, H. & Tarp, S. (Eds). *Manual of Specialised Lexicography. The preparation of specialised dictionaries.* 111-142. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65–80.
- Bergenholtz, H. 2006. How to do language policy with dictionaries. *Lexikos* 16: 13–45.
- Bergenholtz, H. 2009. Klassifikationen in der Linguistik und in der Lexikographie: Wortarten und Wortverbindungen. *Zeitschrift für Anglistik und Amerikanistik*. Special Issue *Classification in Linguistics and Lexicography*. (57) 3: 275-288.
- Bergenholtz, H. 2011. "Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries." In Fuentes-Olivera, P. A. & Bergenholtz, H. (Eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. 30-53. London/New York: Continuum.
- Bergenholtz, H. 2012. Concepts for monofunctional accounting dictionaries. *Terminology* 18(2): 243-263.
- Bergenholtz, H. & Bergenholtz, I. 2011. "A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool." In Fuentes-Olivera, P. A. & Bergenholtz, H. (Eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. 187-207. London/New York: Continuum.
- Bergenholtz, H. & Bothma, Th. J.D. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- Bergenholtz, H., Götz-Votteler, K. & Herbst, Th. 2009. Preface. *Zeitschrift für Anglistik und Amerikanistik*. Special Issue *Classification in Linguistics and Lexicography*. (57) 3: 215-218.
- Bergenholtz, H. & Gouws, R.H. 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica* 23 : 237-260.
- Bergenholtz, H. & Gouws, R.H. 2010. A Functional Approach for the Choice between Descriptive, Prescriptive and Proscriptive Lexicography. *Lexikos* 20: 26-51.
- Bergenholtz, H. & Gouws, R.H. 2010a. A new perspective on the access process. *Hermes- Journal of Language and Communication Studies* 44: 103-127.
- Bergenholtz, H .& Johnsen, M. 2007. Log Files Can and Should Be Prepared for a Functionalistic Approach. *Lexikos* 17: 1-20.
- Bergenholtz, H., Nielsen, S. & Tarp, S. 2009. "Introduction". In Bergenholtz, H., Nielsen, S. & Tarp, S. (Eds). *Lexicography at crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. 7-16. Bern: Peter Lang.
- Bergenholtz, H. & Tarp, S. 2003. Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes -Journal of Language and Communication Studies* 31: 171-196.

- Bergenholtz, H. & Tarp, S. 2004. The Concept of “Dictionary Usage”. *Nordic Journal of English Studies* 3(1): 23-36.
- Bhattacherjee, A. 2012. *Social Science Research: Principles, Methods, and Practices. USF Open Access Textbooks Collection*. Online on [http://scholarcommons.usf.edu/oa\\_textbooks/3](http://scholarcommons.usf.edu/oa_textbooks/3) (consulted on 22/06/2012).
- Bigirumwami, A. 1967. *Imigani migufi (Proverbes), Inshamarenga (Dictons), Ibisakuzo (Devinettes)*. Nyundo.
- Bizimana, S. 1998. *Imiteérere y'ikinyarwanda Igitabo I*. Butare: Editions de l’IRST.
- Bizimana, S. 2002. *Imiteérere y'ikinyarwanda Igitabo II*. Butare: Editions de l’IRST.
- Blanco X. 2010. « Los frasemas composicionales pragmáticos ». In Mejri, S. & Mogorrón, P. (dirs.). *Opacité, Idiomaticité, Traduction*. Alcant: Universitat d’Alacant.
- Blanco, X. 2011. Traduction des pragmatèmes dans les guides de conversation en russe. Contenus conceptuels et enjeux culturels. *Synergies Tunisie* 2: 75-84.
- Blinkenberg, A. 1928. *L’ordre des mots en français moderne. Première partie*. København: Bianco Lunos Bogtrykkeri.
- Bowker, L. 2010. “The Contribution of Corpus Linguistics to the Development of the Specialised Dictionaries for Learners”. In Fuertes-Olivera, P.A. (Ed.). *Specialised Dictionaries for Learners*. Berlin/New York: De Gruyter.
- Burkhanov, I. 2003. “Pragmatic specifications: Usage indications, labels, examples; dictionaries of style, dictionaries of collocations”. In van Sterkenburg, P. (Ed.) *A Practical Guide to Lexicography*. 102-113. Amsterdam/Philadelphia: John Benjamins.
- Cadiou, Y. (Ed). 1985. *Le Kinyarwanda, études de morphosyntaxe*. Paris : Bibliothèque pour l’Information Grammaticale.
- Chabata, E. 2002. An Advanced Dictionary? Similarities and Differences between *Duramazwi ReChiShona* and *Duramazwi Guru ReChiShona*. *Lexikos* 12: 105-116.
- Chabata, E. 2009. Linguistic Theory in the Practical Lexicography of the African Languages. *Lexikos* 19 (Supplement): 1-15.
- Chabata, E. & Nkomo, D. 2010. The Utilisation of Outer Texts in the Practical Lexicography of African Languages. *Lexikos* 20:73-91.
- Coupez, A. 1980. *Abrogé de Grammaire Rwanda, T. 1, 2*. Butare : Edition Provisoire INRS.
- Coupez, A. 1985. « Problèmes et méthodes en lexicographie bantoue ». In Clas, A. & Caprile, J.P. (Eds). *Guide de recherche en lexicographie et terminologie*. 99-112. Paris : ACCT.
- Coupez, A. & Kamanzi, Th. 1970. *Littérature de Cour au Rwanda*. Oxford : OUP.

- Crépeau, P. & Bizimana, S. 1979. *Proverbes du Rwanda*. Tervuren : MRAC.
- Crépeau, P. & Dion, G.M. 1968. Proverbes, sentences et dictions inédits du Rwanda. *Cahier Ifumba* 1.
- d'Hertefelt, M. & de LAME, D. 1987. *Société, Culture et Histoire du Rwanda. Encyclopédie bibliographique 1863-1980/87 (2 vols)*. Tervuren : MRAC.
- De Schryver, G.-M. 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- De Schryver, G.-M. & Prinsloo, D.J. 2001. «Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — with Special Reference to Cilubà, Sepedi and Kiswahili». In Mdee, J.S. and Mwansoko, H.J.M. (Eds.). 2001. *Makala yakongamano la kimataifa Kiswahili 2000. Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.
- Dobrovolskij, D. & Filipenko, T. 2007. "Russian phraseology". In Burger, H., Dobrovolskij, D., Kühn, P. & Norrick, N. (Eds). 2007. *Phraseologie/ Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/ An International Handbook of Contemporary Research*. 714-727. Berlin: Walter de Gruyter.
- Dufays, F. 1912. *Wörterbuch Deutsch-Kinyarwanda*. Trier: Mosella-Verlag.
- Ferraro, G., Nazar, R. & Wanner, L. 2011. "Collocations: A Challenge in Computer Assisted Language Learning". In Boguslavsky, I. & Wanner, L. (eds.). *Proceedings of the 5th International Conference on Meaning- Text Theory*. 69-79. Barcelona, September 8 – 9, 2011.
- Firth, J.R. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis. Special Volume*. 1-32. London: Philological Society.
- Fonagy, I. 1982. *Situation et signification*. Amsterdam: Benjamins.
- Fuertes-Olivera, P.A. & Tarp, S. 2011. Lexicography for the third millennium: Cognitive-oriented specialised dictionaries for learners. *Ibérica* 21: 141-162.
- Fuertes-Olivera, P., Bergenholz, H., Nielsen, S. & Amo, M.N. 2012. Classification in Lexicography: The Concept of Collocation in the Accounting Dictionaries. *Lexicographica* 28: 293-308.
- Geeraerts, D. 2003. "Meaning and definition". In: van Sterkenburg, P. (Ed.) *A Practical Guide to Lexicography*. 83-93. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R. H. 1991. Toward a Lexicon-based Lexicography. *Dictionaries: Journal of Dictionary Society of North America* 13. 75-90.
- Gouws, R. H. 1996. Idioms and Collocations in Bilingual Dictionaries and their Afrikaans Translation Equivalents. *Lexicographica* 12: 54–88.
- Gouws, R.H. 1993. Afrikaans Learner's Dictionaries for a Multilingual South Africa. *Lexikos* 3: 29-47.
- Gouws, R.H. 1999. Mediostructural representation, textual condensation and user-orientation in the

WAT X. *Lexicographica* 15: 4-37.

- Gouws, R.H. 2000. "Strategies in Equivalent Discrimination". In Mogensen, J.E. *et al.* (Eds). *Symposium on Lexicography IX*. 99-112. Tübingen : Niemer (Lexicographica : Serie Maior).
- Gouws, R.H. 2001. The Use of an Improved Access Structure in Dictionaries. *Lexikos* 11: 101-111.
- Gouws, R.H. 2002. Niching as a Macrostructural Procedure. *Lexikos* 12: 133-158.
- Gouws, R.H. 2003. "Types of articles, their structure and different types of lemmata". In Van Sterkenburg, P. (Ed.) *A Practical Guide to Lexicography*. 34-43. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H. 2003a. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110.
- Gouws, R.H. 2004. Monolingual and Bilingual Learners' Dictionaries. *Lexikos* 14: 264-274.
- Gouws, R.H. 2004a. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R. H. 2005. "Die zweisprachige Lexikographie Afrikaans-Deutsch – eine metalexikographische Herausforderung". In Perkov *et al.* (Eds). *Proceedings: Colloquium on bilingual lexicography with German*, Reihe Germanistische Linguistik: Hildesheim, Georg Olms.
- Gouws, R.H. 2006. The Selection, Presentation and Treatment of Cultural Phrases in a Multicultural Dictionary. *Lexicographica* 22: 24-36.
- Gouws, R.H. 2007. "Sublemmata or main lemmata : A critical look at the presentation of some macrostructural elements". In Gottlieb, H. & Mogensen, J.E. (Eds). *Dictionary Visions, Research and Practice*. 55-70. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H. 2007a. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.
- Gouws, R.H. 2009. "Sinuous lemma files in printed dictionaries: Access and lexicographic functions". In Nielsen, S. & Tarp, S. (Eds). *Lexicography in the 21st Century. In Honour of Henning Bergenholz*. 3-21. Amsterdam/Philadelphia: JohnBenjamins.
- Gouws, R.H. 2009a. The integrated outer texts in recent English and German learners' dictionaries: A critical comparison. *Lexicographica* 25: 67-90.
- Gouws, R.H. 2010. "The Monolingual Specialised Dictionary for Learners". In Fuertes-Olivera, P.A. (Ed.). *Specialised Dictionaries for Learners*. 55-68. Berlin/New York: De Gruyter.
- Gouws, R.H. 2010 a. Fixed word combinations as second level treatment units in dictionaries. In Ďurčo, P. (Ed.). *Feste Wortverbindungen und Lexikographie – Kolloquium zur Lexikographie und Wörterbuchforschung*. 51-63. Berlin: De Gruyter.
- Gouws, R.H.&Prinsloo, D.J. 2005. *Principles and practice of South African lexicography*. Stellenbosch: SUN Press.

Gouws, R.H. & Prinsloo, D.J. 2010. "Thinking out of the box – perspectives on the use of lexicographic text boxes". In *Proceedings of Euralex 2010*. Published online: <http://www.euralex.org/members-only-proceedings/euralex-2010-proceedings> (consulted on 27/09/2012).

Grossmann, F. & Tutin, A. (Eds). 2003. *Les collocations: analyse et traitement*. Amsterdan : de Werelt.

Guiraud, P. 1967. *Structures étymologiques du lexique français*. Paris : Larousse.

Guiraud, P. 1979. *L'étymologie*. Paris: PUF.

Guthrie, M. 1948. *The classification of the Bantu languages*. London: Oxford University Press.

Guthrie, M. 1971. *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages*. vol. 4. Farnborough: Gregg Press.

Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Hausmann, F.J. 1989. «Le dictionnaire des collocations». In Hausmann, F.J., Reichmann, O., Wiegand H.E. & Zgusta, L. et al. (Eds), *Wörterbcher, Dictionaries, Dictionnaires*, vol. 1. 1010-1019. Berlin / New York: Walter de Gruyter.

Hausmann, F.J. 1990. "De quoi se compose l'article du dictionnaire de langue? L'importance du sous-adressage". In Magay, T. & Zigány, J. (Eds) *BudaLEX'88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 September 1988*. 59-66. Budapest: Akadémiai Kiado.

Hausmann, F.J. 1991. «Collocations in Monolingual and Bilingual English dictionaries». In Ivir, V. & Kalogjera, D. (Eds). *Languages in Contact and Contrast*. 225-236. Berlin / New York: Mouton de Gruyter.

Hausmann, F.J. 1997. «Tout est idiomatique dans les langues». In Martins-Baltar, M. (Ed.), *La locution entre langue et usages*. Fontenay/St. Cloud : ENS Éditions.

Hausmann, F.J. 2004. "Was sind eigentlich Kollokationen?" In Steyer, K. (Hr.) *Wortverbindungen – mehr oder weniger fest*. Institut für deutsche Sprache. Jahrbuch2003. 309–334. Berlin/New York: de Gruyter.

Hausmann, F.J. & Blumenthal, P. 2006. Présentation : collocations, corpus, dictionnaires. *Langue française*.150 : 3-13.

Hausmann, F.J. & Wiegand, H.E. 1989. "Component Parts and Structures of General Monolingual Dictionaries : A Survey". In Hausmann, F.J. et al. (Eds). *Wörterbücher/Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*. Volume I. 328–360. Berlin/New York: De Gruyter.

Heid, U. 2004. "On the presentation of collocations in monolingual dictionaries". In *Proceedings of the Eleventh EURALEX International Congress* 2: 729-738. Lorient: UBS.

Heid, U. & Weller, M. 2009. Extraction de collocations à partir de corpus : un système pour l'allemand. *Revue de la lexicologie* 25 : 25-42.

Hill, F. 2004. *Idioms*. Munich: GRIN Publishing GmbH Published online on: <http://www.grin.com/en/e-book/71028/idioms> (23/07/2012)

Hitayezu, Th. 1999. *Matumizi ya Kiswahili katika uwanja wa jeshi: Mtazamo wa kiisimujamii*. (Unpublished Master Dissertation). Butare: National University of Rwanda.

Hurel, E. 1920. *Grammaire Kinyarwanda*. Alger.

Hurskainen, A. 2003. New Advances in Corpus-based Lexicography. *Lexikos* 13: 111-132.

Jouannet, F. (éd) 1983. *Le Kinyarwanda, Langue bantu du Rwanda. Etudes linguistiques*. Paris : SELAF.

Jousse, A.-L. 2003. *Normalisation des fonctions lexicales non standard dans les collocations*. (MPhil Thesis).Paris : Université Paris 7- Denis Diderot.

Jousse, A-L., L'Homme, M.C., Leroyer, P. & Robichaud, B. 2011. "Presenting collocates in a dictionary of computing and the Internet according to user needs". In Boguslavsky, I. & Wanner, L. (Eds.) *Proceedings of the 5th International Conference on Meaning-Text Theory*. 134-144. Barcelona, 8-9 September 2011.

Kagame, A. 1953. *Imigani y'imigenûrano*. Kabgayi : Editions royales.

Kagame, A. 1977. Quelques problèmes de la sémantique du Kinyarwanda. *Etudes Rwandaises*, Numéro spécial. 112-125.

Kammerer, M. & Wiegand, H. E. 1998. Über die textuelle Rahmenstruktur von Wörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224–238.

Kayigana, C. 1971[2007] *Imigani n'inshoberamahanga bisobanuye*. Kigali.

Kayigema, L.J. 2010. *Loanword allocation in Kinyarwanda*. Master thesis. Pretoria: UNISA.

Kilgarriff, A. 2012. "Pedro A. Fuertes-Olivera and Henning Bergenholz (eds.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*". *Kernerman Dictionary News* (July 2012): 26-29.

Kimenyi, A. 1979. *Studies in Kinyarwanda and Bantu Phonology*. Edmonto: Arbertha Linguistic Research.

Kimenyi, A. 1980. *A relational grammar of Kinyarwanda*. Berkeley/Los Angeles: University of Carfornia Press.

Kimenyi, A. 1999. Cow metaphors. *Wihogora Rwanda* 5: 78-85.

Kimenyi, A. 2000. "Journey metaphors." In *31st Annual Conference on African Linguistics (ACAL2000)*, 2-5 March 2000. Boston: Boston University.

- Kimenyi, A. 2002. *A Tonal Grammar of Kinyarwanda: An Autosegmental and Metrical Analysis*. Lewiston: The Edwin Mellen Press.
- Kimenyi, A. 2004. "Kinyarwanda (Bantu)". In Booij, G. E., Lehmann, Ch. & Mugdan, J. (Eds) *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word formation*. Vol. 2, 1527-1545. Berlin: Walter de Gruyter.
- Kromann, H-P., Riiber, T. & Rosbach. 1984. "Active and passive bilingual dictionaries: the Ščerba concept reconsidered". In *Proceedings of Lexeter '83*. 207-215. Tübingen : Ed. II.
- Larivière, L., 1998 «La constitution d'un répertoire de collocations verbales : Une hiérarchisation à trois niveaux ». In Guevel, Z. & Valentine, E. (dirs). *Traduction et langage de spécialité. Approche théoriques et considérations pédagogiques*. 179-208. Québec : CIRAL (Université Laval).
- Louw, Ph. 2000. An Integrated Semasiological and Onomasiological Presentation of Semantic Information in General Monolingual Dictionaries as Proposed in H.E. Wiegand's Semantics and Lexicography. *Lexikos* 10: 119-137.
- Louw, Ph. 2006. Inclusion Strategies for Multi-word Units in Monolingual Dictionaries. *Lexikos* 16: 95-103.
- Mabika Mbokou, L. 2002. Les proverbes dans *Kongo Proverbs and the Origins of Bantu Wisdom* par Mukumbuta Lisimba. *Lexikos* 12: 275-289.
- Mabika Mbokou, L. 2006. *A Model for the Macro- and Microstructure of a Yipunu French School Dictionary* (Unpublished doctoral dissertation). Stellenbosch: Stellenbosch University.
- Martins-Baltar, M. 2000. « Construire un dictionnaire d'énoncés (unilingue comme plurilingue): à propos de Dicomotus ». In Szende, Th. (ed.). *Approches contrastives en lexicographie bilingue*. 301-317. Paris: Champion.
- Mel'čuk, I. (In press). Tout ce que nous voulions savoir sur les phrasèmes, mais ... *Cahiers de lexicologie, revue internationale de lexicologie et de lexicographie*. Paris : Classique Garnier. Online : <http://olst.ling.umontreal.ca/pdf/Mel'čukPhrasemes2011.pdf>. (Consulted on 14/09/2011).
- Mel'čuk, I. 1988. *Dependency Syntax: Theory and Practice*. Albany: State University of New York Press.
- Mel'čuk, I. 1997 *Vers une linguistique Sens-Texte. Leçon inaugurale*. Paris: Collège de France.
- Mel'čuk, I. 1998. "Collocations and Lexical Functions". In Cowie, A. (Ed). *Phraseology. Theory, Analysis and Applications*. 23-53. Oxford: Clarendon Press.
- Mel'čuk, I. 2004. Verbes supports sans peine. *Lingvisticae Investigationes* 27 (2) : 203-217.
- Mel'čuk, I. 2010. « La phraséologie en langue, en dictionnaire et en TALN ». In *Proceedings of the 17th Conference on Natural Languages Processing (TALN'10)*, July 2010, Montréal, Canada.

- Mel'čuk, I. 2012. *Semantics : from meaning to text*. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, I., Clas, A. & Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain la-Neuve : Duculot.
- Merriam, A.P. & Armstrong, B.W. 1954. Banyarwanda Proverbs. *JL American Folklore* 67: 272-284.
- Mihindou, G.-R. 2006. *A Theoretical Model for a Yipunu-English-French Explanatory Dictionary of Medical Terms*. (Unpublished doctoral dissertation). Stellenbosch: University of Stellenbosch.
- MNR 2005. *Ibirali by'insigamigani. Igitabo cya mbere. (Icapwa rya 2)*. Butare: Régie du Musée Nationale du Rwanda.
- Mochiwa, Z.s.m. 2007. Kiswahili Verbs: A Lexicographical Challenge. *Lexikos* 17: 114-133.
- Mojela, V. M. 2004. Etymological Aspects of Idiomatic and Proverbial Expressions in the Lexicographic Development of Sesotho sa Leboa - A Semantic Analysis. *Lexikos* 14: 331-339.
- Mphahlele, M.C. 2003. The Lexicographic Treatment of Sublexical and Multilexical Items in a Northern Sotho Monolingual Dictionary: A Challenge for Lexicographers. *Lexikos* 13: 154-167.
- Mpofu, N. 2002. Lexicographical Developments in the Shona Language as Reflected in the Making of the *Duramazwi GuruReChiShona (DGS)*. *Zambezin* XXIX (ii): 149-155.
- Mpofu, N. 2007. The User Perspective in Lexicography: The Lemmatisation of Fixed Expressions in *Duramazwi Guru reChiShona*. *Lexikos* 17: 360-371.
- Mulhall, C. 2010. "A Semantic and Lexical-Based Approach to the Lemmatisation of Idioms in Bilingual Italian-English Dictionaries". In Dykstra, A. & Schoonheim, T. (eds.) *Proceedings of the XIV Euralex International Congress*. 1355-1371.
- Mulihano, B. & Mineduc. 1980. *Ibirali by'insigamigani : Icyiciro I*. Kigali: MINESUPRES.
- Mulihano, B. & Minesupres. 1986. *Ibirari by'insigamigani : Icyiciro II*. Kigali: MINESUPRES.
- Munyakazi, L. 1992. Amabaango yîkinyarwaanda atânaze. *Etudes Rwandaises* 2 (1) : 233-246.
- N'Diaye Corréard, G. 2008. Thème, rhème et syntaxe. *Sudlangues* 10: 1-21. Published online on <http://www.sudlangues.sn/> (28/10/2012).
- Nielsen, S. 1994. *The Bilingual LSP Dictionary. Principles and Practice for Legal Language*. Tübingen: G. Narr.
- Nielsen, S. 1995. "Access Structure". In Bergenholz, H. & Tarp, S. (Eds): *Manual of Specialised Lexicography: The preparation of specialised dictionaries*. Amsterdam/Philadelphia: John Benjamins: 219-223.
- Nielsen, S. 1995a. "Alphabetic macrostructure". In Bergenholz, H. & Tarp, S. (Eds). *Manual of Specialised Lexicography. The preparation of specialised dictionaries*. 190-195.

Amsterdam/Philadelphia: John Benjamins.

Nielsen, S. 1995b. "Microstructure". In Bergenholz, H. & Tarp, S. (Eds). *Manual of Specialised Lexicography: The preparation of specialised dictionaries*. 200-210. Amsterdam/Philadelphia: John Benjamins.

Nielsen, S. 1995c. "Structural Components". In Bergenholz, H. & Tarp, S. (Eds.): *Manual of Specialised Lexicography: The preparation of specialised dictionaries*. 167-187. Amsterdam/Philadelphia: Benjamins.

Nielsen, S. 1995d. "Systematic macrostructure". In Bergenholz, H. & Tarp, S. (Eds). *Manual of Specialised Lexicography. The preparation of specialised dictionaries*. 195-200. Amsterdam/Philadelphia: John Benjamins.

Nielsen, S. 1999. Mediostructures in Bilingual LSP Dictionaries. *Lexicographica* 15: 90-113.

Nielsen, S. 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.

Nikuze E. 2006. *Structures lexico-sémantiques des expressions idiomatiques du kinyarwanda*. Unpublished thesis. Butare: UNR.

Nikuze, E. 2008. Locutions et collocations : esquisse de typologie des expressions idiomatiques du kinyarwanda à la lumière de la Théorie Sens-Texte (TST). *Bulletin de l'IRST. Numéro spécial* : 149-196.

Niyibizi, S. 1980. *Le Swahili au Rwanda. Aperçu historique, linguistique et sociolinguistique*. Unpublished thesis. Butare: UNR.

Nkomo, D. 2008. *Towards a Theoretical Model for LSP Lexicography in Ndebele with Special Reference to a Ndebele Dictionary of Linguistic and Literary Terms*. Unpublished M.Phil. thesis. Stellenbosch:Stellenbosch University.

Nkongoli, L. & Kamanzi, Th. 1957. *Proverbes du Rwanda*. Tervuren: MRAC.

Nkusi, L. 1986 a. « Etymologie et ethymologia dans les traditions orales du Rwanda ». In Nkusi, L. (Ed). *Etudes d'Ethnolinguistique et de Littérature Orale Rwandaise*. Ruhengeri : GERLA-UNR.

Nkusi, L. 1986. « Quelques éléments programmatifs pour une étude de l'ethnographie de la communication au Rwanda ». In Nkusi, L. (Ed). *Etudes d'Ethnolinguistique et de Littérature Orale Rwandaise*. Ruhengeri : GERLA-UNR.

Nkusi, L. 1995. *Analyse syntaxique du Kinyarwanda y compris ses dialectes et avec référence spéciale à la syntaxe des formes simples de la littérature orale rwandaise*. Paris : Université Paris V. (Thèse de doctorat d'Etat).

Ntakirutimana, E. 1996. *Historia ya Kiswahili*. Butare : NUR.

Ntakirutimana, E. 2002. *La langue Swahili comme base d'unification dans la région des Grands lacs*

- africains. (PhD Thesis). Québec : Université Laval.
- Ntakirutimana, E. 2007. Kiswahili Nchini Rwanda : Kupanda na Kushuka. *Etudes Rwandaises* 15 : 12-21.
- Ntampaka, A. 1992. *Représentation et expression de l'action en kinyarwanda : le cas de la diathèse verbale*. Unpublished thesis. Ruhengeri: UNR.
- Ntawigira, P. 1997. *Kiswahili Nchini Rwanda: Mchango katika mradi wa utengenezaji wa lugha*. Unpublished thesis. Butare: UNR.
- Ntazinda, Ch. 1980. *Le complément d'objet en Kinyarwanda*. Unpublished thesis. Butare: UNR.
- Overschelde, V.G. 1972. *Amarenga 1150/ 1150 Dictons*. Kigali.
- Papadopoulou, E. 2010. *Diccionario monolingüe coordinado para enseñanza/aprendizaje del griego modern por parte de hispanohablantes y para traducció automática griego-español*. Doctoral Dissertation. Barcelona : Autonomous University of Barcelona.
- Papadopoulou, L. 2012. “Pragmatemes (Πραγματηματα): Theory and their “Première” In Greek”. In Gavriilidou, Z., Efthymiou, A., Thomadaki, E. & Kambakis-Vougiouklis, P. (eds). *Selected papers of the 10th International Conference of Greek Linguistics*. 480-491. Komotini/Greece: Democritus University of Thrace.
- Patsala, P. 2006. “On the Distinction and Presentation of Idioms and Collocations in Bilingual English-Greek Dictionaries”. In *Proceedings of the Euralex International Congress*. 1029-1035.
- Pauwels, M. 1953. Le Munyarwanda et ses proverbes I, II, III, IV. *Grands Lacs* 9 (163) : 47-54 ; 10 (164) : 49-55; 2 (166) : 48-54 ; 3 (167) : 9-16.
- Polguère, A. 2000. Towards a Theoretically Motivated General Public Dictionary of Semantic Derivations and Collocations for French. In *Proceedings of EURALEX’2000*. 517-527. Stuttgart, Germany.
- Polguère, A. 2003. *Lexicologie et sémantique lexicale: notions fondamentales*. Montreal : Presses de l’Université de Montréal.
- Prinsloo, D.J. 1994. Lemmatisation of verbs in Northern Sotho. *South African Journal of African Languages*, 14(2):93–102.
- Prinsloo, D.J. 2011. A Critical Analysis of the Lemmatisation of Nouns and Verbs in isiZulu. *Lexikos* 21: 169-193.
- Prinsloo, D.J. & Gouws, R.H. 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10: 138-156.
- Ptaszynski, M.O. 2009. Function-related Secondary User Needs and Secondary Data? A Critical Examination of Some Central Concepts in the Modern Theory of Lexicographical Functions. *Lexikos* 19: 242-261.

- Rundell, M. 2006. "More than One Way to Skin a Cat: Why Full- Sentence Definitions have not been Universally Adopted." In Corino, E., Marello, C. & Onesti, C. (Eds). *Atti del XII Congresso Internazionale di Lessicografia : Torino, 6-9 settembre 2006.* 323–337. Alessandria: Edizioni dell'Orso.
- Rundell, M. 2012. "'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical". In Fjeld, R. V. & Torjusen, J. M. (Eds.). *Proceedings of the 15<sup>th</sup> EURALEX International Congress.* 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Rwabukumba, G. 1980. *Les compléments du verbe introduits par les suffixes enrwanda.* Unpublished thesis. Butare : UNR.
- Rwabukwandi, E. 1981. *Etude des circonstants en Kinyarwanda.* Unpublished thesis. Butare :UNR.
- Rwanda Focus. 2009. Rwanda's unofficial second language: Luganda. *Rwanda Focus*, 1<sup>st</sup> April 2009. Online: <http://focus.rw/wp/2009/04/rwandas-unofficial-second-language-luganda/>
- Sierra, G. & Hernández, L. 2011. A Proposal for Building the Knowledge Base of Onomasiological Dictionaries. *Journal of Cognitive Science* 12: 215-232.
- Sinclair, J. 1991. *Corpus, Concordance, Collocations.* Oxford: Oxford University Press.
- Smit, M. 2002. The Systematic Development of Wiegand's Metalexicography as Demonstrated in *Kleine Schriften. Lexikos* 12: 290-310.
- Soami, L.S. 2010. *Towards the Development and Application of Representative Lexicographic Corpora for the Gabonese Languages.* Unpublished Doctoral Dissertation. Stellenbosch: Stellenbosch University.
- Spohr, D. 2012. *Towards a Multifunctional Lexical Resource: Design and Implementation of a Graph-Based Lexicon Model.* Berlin/Boston: Walter de Gruyter.
- Steyn, M. 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.
- Steyn, M. 2006. Aspects of the Micro-Architecture in Learner's Dictionaries. *Lexicographica* 22: 138-151.
- Sugira, D. 2004. *Ufundishaji wa Kiswahili Katika Shule za Sekondari Nchini Rwanda: : Mtazamo wa Kiisimujamii.* Unpublished thesis. Butare : UNR.
- Tarp, S. 1999. Theoretical Foundations of the So-called Cross-reference Structures. *Lexicographica* 15: 114-137.
- Tarp, S. 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S. 2001. Lexicography and the linguistic concepts of homonymy and polysemy. *Lexicographica* 17: 22-39.

- Tarp, S. 2003. The Usefulness of Different Types of Articles in Learner's Dictionaries. Review of "Ursula Wingate: The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German. Tübingen: Niemeyer 2002". *Hermes* 30: 215-234.
- Tarp, S. 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.
- Tarp, S. 2004a. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.
- Tarp, S. 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.
- Tarp, S. 2008a. The Third Leg of Two-legged Lexicography. *Hermes – Journal of Language and Communication Studies* 40: 117-131.
- Tarp, S. 2009. "Beyond Lexicography: New Visions and Challenges in the Information Age". In: Bergenholz, H., Nielsen, S. & Tarp, S. (Eds). *Lexicography at crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. 17-32. Bern: Peter Lang.
- Tarp, S. 2009a. "Reflections on data access in lexicographic works". In: Nielsen, S. & Tarp, S. (Eds.). *Lexicography in the 21st Century. In Honour of Henning Bergenholz*. 43-62. Amsterdam/Philadelphia: John Benjamins.
- Tarp, S. 2009b. Homonymy and polysemy in a lexicographical perspective. *Zeitschrift für Anglistik und Amerikanistik. Special Issue:Classification in Linguistics and Lexicography*. (57) 3: 289-305.
- Tarp, S. 2009c. Reflections on lexicographic user research. *Lexikos* 19: 275-296.
- Tarp, S. 2009d. The foundation of a theory of learners' dictionaries. *Lexicographica* 25: 155-168.
- Tarp, S. 2010. Reflections on the Academic Status of Lexicography. *Lexikos* 20: 450-465.
- Tarp, S. 2010a: "Functions of Specialised Learners' Dictionaries". In Fuertes-Olivera, P.A. (Ed.): *Specialised Dictionaries for Learners*. 39-53. Berlin/New York: De Gruyter.
- Tarp, S. 2011. Pedagogical Lexicography: Towards a new and strict typology corresponding to the present state-of-the art. *Lexikos* 21: 217-232.
- Tarp, S. & Gouws, R.H. 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Tono, Y. 2010. A critical review of the Theory of Lexicographical Functions. *Lexicon (Journal of the Iwasaki Linguistic Circle)* 40: 1-26.
- Tutin, A. 2005. Le dictionnaire de collocations est-il indispensable ? *Revue française de linguistique appliquée* X (2) : 31-48.

- Tutin, A. 2010. *Sens et combinatoire lexicale : de la langue au discours I: Synthèse*. Grenoble : Université Stendhal.
- Tutin, A. & Grossmann, F. 2002. Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*. VII (2002/1) : 7-25.
- Uwilingiyimana, J. C. 2000. *Structures étymologiques du lexique du kinyarwanda*. Unpublished thesis. Butare: UNR.
- Vorobey, M. 2011. “French-Spanish-Russian Pragmatemes Dictionary”. In *Proceedings of the International conference on the Meaning-Text Theory, Barcelona 8-9 September 2011*. 287-294.
- Vrbinc, A. & Vrbinc, M. 2011. Treatment of Multi-word Lexical Items in the Dictionary: the Current Situation and the Potential Problems Facing Dictionary Users. *Eesti Rakenduslingvistika Ühingu Aastaraamat* 7: 249–263.
- Welker, H.A. 2009. An overview of Wiegand’s metalexicographic works. <http://www.let.unb.br/hawelker/Wiegand.pdf> (Last date of consultation: 24/08/2012).
- Wiegand, H.E. 1984. “On the Structure and Contents of a General Theory of Lexicography”. In Hartmann, R. R. K. (Ed.). *LEX'eter '83 Proceedings*. 13–30. Tübingen: Max Niemeyer.
- Wiegand, H.E. 1996. A theory of lexicographic texts: an overview. *South African Journal of Linguistics* 14 (4) : 134-149.
- Wiegand, H.E. (Ed). 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin/New York: De Gruyter.
- Zgusta, L. 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.
- Žolkovskij, A. & Mel’čuk, I. 1965. O vozmožnom metode i instrumentax semantičeskogo sinteza [On a Possible Methodology and Tools for Semantic Synthesis (of texts)]. *Naučno-texničeskaja informacija* 5: 23-28.
- Žolkovskij, A. & Mel’čuk, I. 1966. O sisteme semantičeskogo sinteza. I: Stroenie slovarja [On a System of Semantic Synthesis (of texts). I: Building the Dictionary]. *Naučno-texničeskaja informacija* 11: 48-55.
- Žolkovskij, A. & Mel’čuk, I. 1967. O sisteme semantičeskogo sinteza. II: Pravila preobrazovaniija [On a System of Semantic Synthesis (of texts). II: Paraphrasing Rules]. *Naučno-texničeskaja informacija* 2: 17-27.