This book provides a broad overview of current work on South African languages, language resources and language technologies. While it provides a fairly comprehensive overview, it also ties together the most recent knowledge state here, and is therefore truly innovative... The book is therefore informed by current international trends in the respective fields of science, and feeds back into them... There is absolutely no doubt that the book has an academic peer audience and is directed at specialists in the field.

Prof. Axel Fleisch, University of Helsinki, Finland

This Festschrift, containing 20 stimulating articles by colleagues and students, is a fitting tribute to his 65th birthday. The book covers a wide range of subjects dealing with phonetics and phonology, language description and resources, lexicography and terminology, and language technology research and development. Each article presents the results of experiments completed or the exposition of viewpoints held in connection with these subject fields. However, the presentations do not just end here. They also contain pointers to further problems for research and discussion. These articles are therefore not complete in themselves, but open-ended, leaving not only the possibility for reconsidering the research and discussions presented, but also, because the information and viewpoints given are original and innovative, for stimulating thought and evoking reaction. These articles are therefore meant for specialists in the relevant fields who can fully appreciate the experiments and arguments and respond to them in an informed manner.

Dr Johan du Plessis, Bureau of the WAT, Stellenbosch, South Africa

Dr H. Steve NDINGA-KOUMBA-BINZA is Research Fellow at the Centre for Text Technology, North-West University in Potchefstroom, South Africa, and Assistant Professor of phonetics and phonology at Omar Bongo University in Libreville, Gabon. His research interests include African vowel systems and syllable structures for use in text and speech technology development and writing and spelling systems. He has authored and co-authored numerous publications, including the books Civili, langue des Baloango: Esquisse historique et linguistique (Lincom Europa 2010; co-author: PA Mavoungou) and A phonetic and phonological account of the Civili vowel duration (Cambridge Scholars Press 2012). He received his Doctor Litteratum (PhD) degree under Justus C. Roux in 2008 at Stellenbosch University, South Africa.

Prof. Sonja E. BOSCH is Chair of the Department of African Languages at the University of South Africa in Pretoria. She teaches Zulu acquisition courses and undergraduate as well as postgraduate courses in morphology and syntax. She is project leader of a research group dealing with computational morphological analysis of African languages, as well as coordinator of the African Wordnet project. As researcher, she has organised several workshops nationally and internationally and is an NRF-rated researcher. Her publications include more than 50 peer-reviewed journal articles, conference proceedings and book chapters, as well as three books published nationally and internationally. She has been a friend, a colleague and co-project leader to Justus C. Roux for many years.

3387947819209
ISBN 978-1-920338-79-4
Language Science and Language Technology in Africa:
A Festschrift for Justus C. Roux

Edited by
Hugues Steve Ndinga-Koumba-Binza
&
Sonja E. Bosch

SUN PRESS
ACKNOWLEDGEMENT TO REVIEWERS

Every contribution in this volume was subject to a double blind peer-review process. We wish to acknowledge the participation of the following colleagues in the production of this book. Some of them had to review the same chapter two or three times, and others reviewed more than one chapter. Thank you!

Mariëtta Alberts (NWU)          Shamila Naidoo (UKZN)
Etienne Barnard (NWU)           Steve Ndinga-Koumba-Binza (NWU)
Ian Bekker (NWU)                Thomas Niesler (SU)
Ansu Berg (NWU)                 Mildred Nkolola-Wakumelo (Zambia)
Sonja E. Bosch (UNISA)          Dion Nkomo (RU)
Willem Botha (WAT)              Blanche Nyangone Assam (UWC)
Emmanuel Chabata (Zimbabwe)     Sulene Pilon (NWU)
Febe de Wet (CSIR)              Rigardt Pretorius (NWU)
Johan du Plessis (WAT)          Danie Prinsloo (UP)
Rufus Gouws (SU)                Martin Puttkammer (NWU)
Marissa Griesel (NWU)           Aditi Sharma Grover (CSIR)
Hendrik Groenewald (NWU)        Frenette Southwood (SU)
Maxwell Kadenge (WITS)          Elsabé Taljard (UP)
Inge Kosch (UNISA)              Charl van Heerden (CSIR)
Albert E. Kotzé (UNISA)         Gerhard van Huyssteen (NWU)
John Lubinda (Botswana)         Linda van Huyssteen (TUT)
Ludwine Mabika Mbokou (Gabon)   Bertus van Rooy (NWU)
Paul Achille Mavoungou (Gabon)  Daan P. Wissing (NWU)
Cindy A. McKellar (NWU)         Sabine Zerbian (Germany)

The editors are also grateful to Professor Axel Fleisch (Department of World Cultures, University of Helsinki, Finland) and to Dr Johan C.M.D. du Plessis (former editor of *Lexikos*, Bureau of the Woordeboek van die Afrikaanse Taal, South Africa) for their thorough and insightful review of the entire volume.
CONTENTS

JUSTUS C. ROUX’S BRIEF CURRICULUM VITAE v

PREFACE xi

PART 1: PHONETICS AND PHONOLOGY

STRESS ASSIGNMENT IN BLACK SOUTH AFRICAN ENGLISH
Sabine Zerbian 1

A COMPARATIVE STUDY OF EXPLOSIVE [b] AND IMPLOSIVE [ɓ] IN NGUNI
Shamila Naidoo 21

ON THE PHONETIC AND THE PHONOLOGICAL STUDIES OF SOUTHERN AFRICAN LANGUAGES
John Lubinda 33

ON THE MANY FACES OF TONE IN SOUTHERN SOTHO: A CASE STUDY
Daan Wissing 47

SYLLABIFICATION OF CONSONANTS IN SESOTHO AND SETSWANA
Mildred Nkolola-Wakumelo, Liketso Rantso, Keneilwe Matlhaku 61

PHONETIC DATA AND PHONOLOGICAL THEORY: A REPORT FROM THE CIVILI VOWEL DURATION ISSUE
Hugues Steve Ndinga-Koumba-Binza 83

PART 2: LANGUAGE DESCRIPTION AND RESOURCES

THE ORATURE-GRAMMAR INTERFACE: ON “RHYMES” IN AFRICAN ORAL VERBAL ART
H. Ekkehard Wolff 99

BOOTSTRAPPING THE DEVELOPMENT OF MORPHOLOGICAL ANALYSERS FOR "DISPERSED" NGUNI LANGUAGES - A LINGUISTIC INVESTIGATION
Sonja E. Bosch 123

FROM 'BELEAGUERED' TO 'EMANCIPATED' ORTHOGRAPHY: THE CASE OF NORTHERN SOTHO
Inge Kosch 145

A SURVEY OF BILINGUALISM IN MULTILINGUAL GABON
Ludwine Mabika Mbokou 163
PART 3: LEXICOGRAPHY AND TERMINOLOGY

TOWARDS A YILUMBU DICTIONARY OF IDIOMATIC PHRASES
Paul Achille Mavoungou 177

LANGUAGE STANDARDISATION AND OTHER LANGUAGE AND CONTENT RESOURCES: SABS
TC 37 - A MIRROR COMMITTEE OF ISO/TC 37
Mariëtta Alberts 199

ENRICHING A DICTIONARY DATABASE WITH MULTI WORD EXPRESSIONS
Thapelo Otlogetswe 215

DICTIONARY BASIS AND LEMMATISATION FOR AN ENCYCLOPEDIC DICTIONARY OF YILUMBU
Hugues Steve Ndinga-Koumba-Binza and Gilles Saphou-Bivigat 237

INNOVATIVE STRATEGIES IN MACROSTRUCTURAL CHOICES
Rufus H. Gouws 251

PART 4: HLT RESEARCH AND DEVELOPMENT

FREQUENCY-BASED DATA SELECTION FOR STATISTICAL MACHINE TRANSLATION WITH SCARCE RESOURCES
Cindy A. McKellar and Hendrik J. Groenewald 271

VOICE USER INTERFACE DESIGN FOR EMERGING MULTILINGUAL MARKETS
Gerhard van Huyssteen, Aditi Sharma Grover, Karen Calteaux 291

THE RELATIONSHIP BETWEEN THE AUTOMATIC ASSESSMENT OF ORAL PROFICIENCY AND OTHER INDICATORS OF FIRST YEAR STUDENTS’ LINGUISTIC ABILITIES
Febe de Wet, Thomas Niesler, Christa van der Walt 309

CONCEPTS FOR DIFFERENT TYPES OF INFORMATION TOOLS
Henning J. Bergenholtz 327

‘MARKET SPEAK’ IN IGBO: A SPEECH SYNTHESIS TRAINING PROJECT
Dafydd Gibbon, Ugonna Duruibe, Jolanta Bachan 339

INDEX 360
JUSTUS C. ROUX’S BRIEF CURRICULUM VITAE

1947  Born 15 January 1947 in Brakpan, South Africa.

1967  Received BA degree majoring in Afrikaans and Dutch, Northern Sotho and Zulu from the then Potchefstroom University for Christian Higher Education (PUCHE).

1968  Received an Honours degree in African languages from PUCHE.

1969  Appointed Junior Lecturer in African languages at PUCHE.

1971  Received a Master of Arts in African languages from PUCHE.

1971  Promoted to Lecturer in African languages at PUCHE.

1972  Appointed Lecturer in African languages at Stellenbosch University. This same year, he married Cornelia and they have three grown-up children – Ruzanne, Christelle and Justus – and two grandchildren.

1979  Received his Doctor Litteratum (DLitt) degree in African Languages and General Linguistics from Stellenbosch University.
1980 Promoted to Senior Lecturer in African languages at Stellenbosch University. At the same period he became member of the Suid-Afrikaanse Akademie vir Wetenskap en Kuns; a membership that he still holds to date. He also took on a membership of the African Language Association of Southern Africa (ALASA), an organisation of which he is still a member and has held various executive positions ranging from Chairperson of the Western Cape Chapter (1980, 1984-1988), Adjudicator for ALASA prizes for various publishers (1986), Member of the Board and Executive of ALASA (1986-1989, 1995-1997), and Secretary to the National Executive Board of ALASA (1995-1997).

1981-2004 He was member of the Linguistics Society of Southern Africa (LSSA).

1982 Appointed Head of the Department of African Languages at Stellenbosch. He was later re-appointed to this position five times until 2000.

1983 Appointed Member of the Faculty Executive Committee, Faculty of Arts at Stellenbosch University. He had this position renewed three times until 1999.

1983-1985 Member of Faculty Committee on Student Feedback.

1984 Attended two specialised international training courses. The first course, on “Theory and practice of Suggestopedic teaching”, was held at the Institute for Language Teaching of the University of Stellenbosch and presented by Prof Charles Schmidt of Lind Institute, San Francisco, USA. The second course, on “Methods and Techniques of Accelerative Learning”, was held by the Department of Continuing Education at the University of Houston, Houston, USA.

1985 Promoted to Full Professor (Ad Hominem) in African Languages at Stellenbosch University. During the same year, he also attended an international training course held by the Institute for Language Teaching of the University of Stellenbosch on “Educational Cognitive Styles - the development of creative thinking skills”, presented by Prof J. Hand (USA) and Ms Beatrice Capdeville (Venezuela).

1986-1990 Appointed member of the Management Committee of the Institute for Language Teaching at Stellenbosch University. He chaired the Committee from 1989 to 1990.
1988-1990 Appointed member of the Senate Sub-Committee for Language Laboratories. It is also in this period that he was appointed member of the Selection Committee of the Human Sciences Research Council on Modern Languages, Arts and Information Science, responsible for adjudication of national research and travel grants from 1989 to 1990. He was simultaneously (until 1991) member of the Management Committee of the National Working Group on Computational Linguistics, functioning under the auspices of the Human Sciences Research Council.

1990 Elected Chairperson of the Committee for Language Laboratories. He was re-elected in 1991 and in 1992.

1993-2004 Founded an NRF unit and became its Director, i.e. the Research Unit for Experimental Phonology at the University of Stellenbosch (RUEPUS). It is during this period that he became Chairperson of the Special Interest Group on Laboratory Phonology functioning under the auspices of the African Language Association of Southern Africa. He was simultaneously founder member of the South African Foundation for Language and Speech Technology Development in 1999 until the organisation discontinued its activities in 2004.

1994 Became member of the Sub-Committee A (Arts & Social Sciences) of the Research Committee of Stellenbosch University until 1999.

1994-1995 Was member of the Management Committee of the Bureau for Continuing Education at Stellenbosch University.


1997-1998 Elected Chairperson of the Management Committee for Computer Based Education in the Humanities. The following year he became Chairperson of the Management Committee of the Computer Use in the Humanities (HUMARGA) until 2003.

1998-1999 Appointed Vice-Dean of the Faculty of Arts at Stellenbosch University.

1999 Founding member of the Special Interest Group for Language and Speech Technology of the African Language Association of Southern Africa (ALASA-SIG).

2000 Appointed member of the Steering Committee for Language and Information Technology Development by the Department of Arts,
Culture, Science and Technology (DACST) and the Pan South African Language Board (PanSALB).

2001
Was once again Vice-Dean of the Faculty of Arts at Stellenbosch University from September to December 2001. This same year he acquired life membership of the Academy of Science of South Africa. It is also in this year that he was appointed Co-ordinator of the Advisory Panel on Human Language Technology Development in South Africa by the Minister of Arts, Culture, Science and Technology; he served in this position until 2002.

2002
Became member of the Academic Research Rating Panel for Language and Linguistics of the National Research Foundation.

2002 – 2007
Appointed Professor of African Languages at Stellenbosch University on a part-time contract.

2003 – 2006
Appointed Chairperson of the South African Technical Committee for the Standardisation of Terminology and other resources SABS TC/37. It is in this capacity that he was member of the South African delegation to the Annual International Meetings of ISO TC/37 in 2005 in Warsaw, Poland, in 2006 in Beijing, China, in 2007 in Provo, USA, and in 2010 in Dublin, Ireland.

2003-2009
Appointed member of the Steering Committee for Human Language Technologies Implementation by the Minister of Arts, Culture, Science and Technology.

2005-2008
Appointed Director of the newly founded Stellenbosch University Centre for Language and Speech Technology (SU-CLaST).

2005
Served as part-time Director of CatchWord Language and Speech Technologies (Pty) Ltd (a Stellenbosch University spin-off company).

2009-2010
Appointed part-time Senior Researcher at SU-CLaST. In the same period, he was nominated member of the HLT Expert Panel (HLTEP) by the Minister of the Department of Arts and Culture (DAC) for a period of five years to assist DAC in deploying HLT Research and Development projects in South Africa. In 2010, the Minister also appointed him Co-ordinator of a Task Team to develop a Blueprint for a Resource Management Agency for the National Centre for Human Language Technologies (NCHLT-RMA).
2009-2011 Appointed Senior Researcher at the Centre for Text Technology (CTeXT) on the Potchefstroom Campus of the North-West University. Since 2009 he has been member of the Committee for Advanced Degrees, Research Unit for Language and Literature within the South African Context, and member of the Executive Committee of the Research Unit for Language and Literature within the South African Context since 2011.

2010 Since 2010, he is the Vice-President of the African HLT Association.

2011 Appointed Director of the Research Unit for Languages and Literatures in the South African Context on the Potchefstroom Campus of the North-West University as from May 2011. He is also member of the Advisory Panel for NRF Researcher Rating (Humanities and Social Sciences) since the same year.
Preface

The genesis of this book dates from January 2010, when the editors met on the occasion of the first National HLT Network workshop at the Council of Scientific and Industrial Research (CSIR) in Pretoria, and agreed to compile a volume in honour of Professor Justus Roux to celebrate his 65th birthday in 2012. When suggested to a number of colleagues, the idea was warmly welcomed and many contributed, as is evident in the contents of the book.

The number of contributions (20 papers with 29 contributors), the calibre of the contributors (half a dozen are NRF-rated researchers, 10 internationally known professors and 4 world leaders in their respective fields), as well as the diversity of contributors’ origins (9 African and European countries, i.e. Botswana, Denmark, Gabon, Germany, Lesotho, Nigeria, South Africa, Poland and Zambia) are testimony to the academic stature of the man being honoured through this book. Both his personality and his work have had an impact on the lives and careers of many of his friends and colleagues.

Professor Justus Christiaan Roux has, in fact, been the study leader, the promoter, the mentor, the colleague, or simply the friend of many worldwide. Close friends of his and colleagues are cited among the world top scholars. This explains the 100 plus visits he has made to top universities in Europe, America and Asia not only for world class conferences and workshops, but also for research stays, partnership and board meetings of several organisations such as the International Phonetics Association (IPA), the International Speech Communication Association (ISCA), the International Committee of Written Language Resources (WRITE) and the International Standards Organization (ISO) to name a few.

Born on 15 January 1947 in Brakpan (now part of the Ekurhuleni Metro Municipality), South Africa, Justus Roux attended the Carletonville High School from 1960 to 1964, before enrolling for a BA degree in the (then) Potchefstroom University for Christian Higher Education (now Potchefstroom Campus of the North-West University). He graduated in 1967 majoring in Afrikaans and Dutch, Northern Sotho and Zulu. He later obtained from the same institution an Honours degree (1968) and a Master of Arts in African Languages (1971). His research and scientific education has been marked with several training courses (e.g. Suggestopedic teaching, Accelerative learning, language technologies, etc.) throughout his long career that started as an academic in the position of Junior Lecturer in African Languages at Potchefstroom University for Christian Higher Education in 1969. He later became Lecturer in African Languages from 1969 to 1971 in that same institution.

In 1972 he moved to Stellenbosch University where he was successively Lecturer, Senior Lecturer and Professor of African Languages before retiring from teaching.
in 2002. He was the Chair of the Department of African Languages at Stellenbosch six times (1982, 1986, 1988, 1992-1993, 1997-1999, and 2000). He was also the Vice Dean of the Faculty of Arts at Stellenbosch University for two different terms (1998-1999 and 2001-2002). In 1993, he founded the Research Unit for Experimental Phonology at the University of Stellenbosch (RUEPUS) which contributed in sparking laboratory phonology and language technologies research and development in South Africa. One of the major outcomes produced in RUEPUS was the successful development of a multilingual African Speech Technology system (i.e. a fully automated telephone-based multilingual query and booking system) for five official languages of South Africa, i.e. South African English, Zulu, Xhosa, Southern Sotho and Afrikaans.

Through the African Language Association of Southern Africa Special Interest Group (ALASA-SIG) on human language technologies that he co-established, he promoted human language technologies research and development not only in South African institutions (universities, private research organisations and companies) but also in other African countries such as Gabon, Ghana and Morocco together with a number of colleagues. It is in this trend that RUEPUS merged with the Digital Speech Processing Group (DSP Group) of the Stellenbosch University Department of Electrical and Electronic Engineering, with which it had already had an active collaboration over a period of twenty years, to form the Stellenbosch University Centre for Language and Speech Technology (SU-CLaST) in 2005. Justus Roux was subsequently appointed Director of SU-CLaST that same year.

His work and passion for promoting human language technologies in the African context led him to organise, co-organise or host conferences and workshops nationally and abroad in the field or related fields. One would note the following to mention the four most recent:


(ii) the International Speech Communication Association’s (ISCA) Tutorial and Research Workshop on Multilingual Speech and Language Processing (MULTILING 2006), Stellenbosch, 9-11 April 2006.

(iii) the pre-conference workshop on Networking the development of resources for African languages, at the 5th International Conference on Language Resources and Evaluation (LREC 2006), 21-28 May, 2006. Genoa, Italy.

Justus Roux’s dedication to the development of language resources and language technologies in South Africa is reflected in his activities on a high level, e.g. his role of co-ordinator of the Advisory Panel on Human Language Technology Development in South Africa, from 2001 to 2002; and subsequently his membership of the Steering Committee for Human Language Technologies Implementation from 2003 – 2009. Both appointments were made by the Minister of Arts, Culture, Science and Technology. In 2009 he was nominated as member of HLT Expert Panel (HLTEP) by the Minister of the Department of Arts and Culture (DAC) for a period of five years to assist the DAC in deploying HLT Research and Development projects in South Africa. Based on his experience in the field of resource management and his collaboration with international language resource projects, he was also appointed co-ordinator of a Task Team to develop a Blueprint for a Resource Management Agency for the National Centre for Human Language Technologies (NCHLT-RMA) in 2010.

This volume brings together work from the fields of phonetics and phonology, morphology, sociolinguistics, terminology, lexicography and language technology research as a reflection of Justus Roux’s wide range of academic interests. The common denominator in the majority of contributions is their reference to African languages (including Black South African English). The chapters in this book are organised in four topical sections. The first is concerned with phonetics and phonology; the second with various field studies grouped into the topic of language description and resources, i.e. morphology, sociolinguistics etc.; the third with terminology and lexicography; and the fourth with language technology research.

The first section contains six chapters. Chapter 1, by Sabine Zerbian, is a study of stress assignment in Black South African English (BSAE) at the word level. The study uses new empirical data to test the stress algorithm for BSAE proposed in previous studies such as van Rooy (2002) who claimed that there are indicators of a single stable system for stress assignment in Tswana English across speakers. Among other findings, the study confirms that the algorithm can account for many of the observed stress patterns and thereby refutes, in line with van Rooy (2002), the previously held impression that stress assignment in this variety is either idiosyncratic or restricted to the penultimate syllable.

Chapter 2, a contribution by Shamila Naidoo, is a comparative analysis of the explosive [b] and implosive [ɓ] consonants in three Nguni languages, i.e. Swati, Xhosa and Zulu. An experimental method and a corpus of minimal pairs are used to conduct a qualitative and quantitative examination of the parameters burst amplitude and closure duration in order to determine the degree of similarity between the studied consonantal sounds in the three languages.
Chapter 3 is a contribution by John Lubinda who protests against the weakness of phonetic and phonological descriptions of the majority of languages in Southern Africa. He pleads the case for a more scientific approach to the study of the sounds and sound systems of these languages, using available technical devices of the phonetics laboratory that guarantee a greater degree of objective precision, and applying appropriate phonological models that ensure descriptive adequacy.

Chapter 4, which is the only tone study of the volume, is a proposal by Daan Wissing. It is an acoustic description and an assessment of tonological characteristics of Southern Sotho. Findings strongly suggest that tone in Southern Sotho is no clear-cut phenomenon, as described or assumed in various writings.

Chapter 5 is a joint contribution by Mildred Nkolola-Wakumelo, Liketso Rantso and Keneilwe Matlhaku whose paper discusses the occurrence of derivative syllabic consonants in Sesotho and Setswana. The authors apply theoretical perspectives drawn from Distinctive Feature Phonology in order to show that the derivative syllabic consonant is characteristic of these two languages both belonging to the Sotho-Tswana group. This chapter is a collaboration of three colleagues based in three different Southern African countries, namely Botswana, Lesotho and Zambia.

Chapter 6, by Hugues Steve Ndinga-Koumba-Binza, concludes this first section with a report on the interaction between phonetic data and phonological theories with reference to the study conducted on the Civili vowel duration issue. The paper defines the nature of efficient data for a systematic study within the framework of the so-called Phonetics-Phonology Interface Debate (P-PID).

Section two comprises four chapters. In chapter 7, Ekkehard Wolff makes a study of the orature-grammar interface with reference to rhymes in African oral verbal art. In fact, the existence of “rhyming” as a salient aesthetic device has long been negated in traditional African verbal art, apart from assumed copies of Arabo-Islamic models. It is herein shown that, in the Chadic language Lamang in Nigeria, rhyme patterns are also salient in elevated discourse and narratives. A number of illustrative examples from two different speakers and different discourse genres are introduced and discussed.

Chapter 8, by Sonja Bosch, is a linguistic investigation of the feasibility of bootstrapping the development of morphological analysers for two ‘dispersed’ Bantu languages, by using an existing prototype of a Zulu morphological analyser. Cross-linguistic morphological similarities and distinguishing grammatical features between Zulu and two ‘dispersed’, resource-scarce Nguni languages, namely Zimbabwe Ndebele (S44) and Tanzanian Ngoni (N12) are examined with reference to their significance for bootstrapping purposes. The investigation focuses on the morphotactics and the morphophonological alternations of the languages involved.
In chapter 9, the contribution by Inge Kosch aims to illustrate how the orthography for Northern Sotho gradually freed itself from impractical and elaborate conventions to become a more user friendly writing system. The European missionaries who reduced the language to writing were informed by various factors in their choice of symbols, ranging from simple transfer from their source language on the one hand to scientifically motivated choices on the other. A high premium was placed on a practical orthography that was close to a phonetic orthography. This approach, however laudable, proved to be too technical for implementation by its target users.

In chapter 10, Ludwine Mabika Mbokou reflects on the notions of bilingualism in the multilingual context of Gabon, a French speaking state in central Africa. A commonly held view among the Gabonese people is that bilingualism only refers to a person who is able to speak two different European languages such as English and French. The author shows that bilingualism is a relative concept with specific reference to Gabonese children who, individually raised in a bilingual environment, are introduced to a second and even a third language at either a later or earlier stage of their childhood. As a result, the younger Gabonese generations are divided into two major groups where French and the native languages are wrestling for the initial language position.

Section three comprises five chapters. Chapter 11 by Paul Achille Mavoungou is an account of the planning of a Yilumbu idiomatic Dictionary. In order to discuss Yilumbu idiomaticity, the author restricts his discussion to the following points: (i) the methodology and theoretical assumptions of the work, and (ii) the type of oral traditions. A comparative terminology analysis of idiomatic expressions in English and Yilumbu is subsequently made.

In chapter 12, Mariëtta Alberts gives an overview of matters such as language development, standardisation of language, standard languages, harmonisation of languages, modernisation of languages and sociolinguistic factors regarding term creation. Emphasis is placed on terminology as a source for communication and training and the role of standardisation regarding terminological principles and practice. All the aspects related to the standardisation of language and other content resources are encapsulated in the activities of and dealt with by a technical committee of the International Organization for Standardization (ISO) and a South African mirror committee at the South African Bureau of Standards (SABS). The business of these technical committees is discussed.

Chapter 13, a contribution by Thapelo Otlogetswe, proposes three strategies of identifying multiword units from a corpus of over fifteen million words to enrich a Setswana dictionary. The study is conducted on an untagged Setswana corpus using Wordsmith Tools. The proposed strategies are: the harvesting of concordance lines, the generation of concgrams and the use of word association measures. The
three strategies have been found to be effective in the extraction of multiword units.

Chapter 14 by Hugues Steve Ndinga-Koumba-Binza and Gilles Saphou-Bivigat focuses on the dictionary basis and a few lemmatisation issues of a planned encyclopedic dictionary for Yilumbu, a developing language spoken in Gabon and in Congo. An early study on the planning of an encyclopedic dictionary for this language contains a number of discrepancies on the proposed dictionary basis and the lemma selection. In this chapter, the authors re-examine the same topics and suggest new perspectives.

In chapter 15, Rufus Gouws focuses on macrostructural and lemmatisation choices in a recently revised LSP dictionary. It is shown how different kinds of multiword and compound terms receive different types of lexicographic presentations. The main emphasis is on those sublemmata condensed to partial lemmata and presented in article niches and nests. Some of these lemma types are discussed, and innovative strategies that have been followed, e.g. the use of macrostructurally-isolated and double-layered sublemmata, are analysed. Using a contemplative approach, this paper endeavours to contribute to the expansion of a theoretical model by making the procedures that are discussed accessible to future lexicographers.

Section four also comprises five chapters. In chapter 16, Cindy McKellar and Hendrik Groenewald show that the utilisation of frequency-based data selection techniques for the generation of training data results in an increased learning rate when applied to statistical machine translation systems for three resource-scarce South African languages, namely Afrikaans, isiZulu and Sepedi. The paper first gives a general introduction to statistical machine translation. This is followed by an overview of related work and frequency-based data selection. The final part of the contribution gives information about the experimental setup and presents the results of the various machine translation systems. An interpretation of the results is provided together with some directions for future work.

Chapter 17 by Gerhard van Huyssteen, Aditi Sharma Grover and Karen Calteaux aims at getting a better understanding of business and design issues related to interactive voice responses (IVRs) in a multilingual, emerging market such as South Africa, in order to shed light on the challenges relating to voice user interface (VUI) design for such markets. It is an attempt to provide a first snapshot of the situation in South Africa, and to explicate some of the challenges. Among the 34 selected South African IVRs investigated, only 9 were found to have a multilingual offering, with only 5 having some form of speech input. Cost is the major driver for multilingual IVRs overshadowing the many positive business drivers in support of multilingual IVRs.
In chapter 18, Febe de Wet, Thomas Niesler and Christa van der Walt describe quantitative indicators obtained from undergraduate university students, with a view to automatically assess their oral proficiency. By applying automatic speech recognition methods to the automatic assessment of oral proficiency and listening comprehension, these logistical difficulties can be alleviated. Results of an automatic test are compared with human evaluations of the same data, as well as with the results of written placement tests, to determine the relationship between these approaches.

Chapter 19 is a contribution by Henning Bergenholtz who focuses on concepts for different types of information tools. The author indicates that occupation of lexicography by linguistics has serious adverse effects on the theoretical development of metalexicography and the physical products of dictionary work. In particular, the British tradition in which lexicography is classified as a field without theory and for which the linguistic theories should suffice is sharply criticised. The German tradition, while arguing at a high theoretic level, also uses the linguistic theories and forms of presenting research. The author then argues that the view that dictionaries must of necessity document extensively in the form of polyfunctional dictionaries is part and parcel of this traditional perspective. This does not apply to printed dictionaries only, but allegedly also – or even specifically – to internet dictionaries. The counterthesis postulated in this contribution is that the internet makes it even more possible than printed dictionaries to produce monofunctional dictionaries which, depending on their function, can be derived from one and the same database.

Lastly, chapter 20 is another contribution by three colleagues based in three different countries including Germany, Poland and Nigeria. In this chapter Daffyd Gibbon, Ugonna Duruibe and Jolanta Bachan present a tutorial approach to speech technology infrastructure development for a less resourced Niger-Congo tone language (Igbo), with well-known components but a new integrative strategy for creating a prototype digital signal processing front-end for a Text-to-Speech synthesiser. Text parsing problems are only dealt with in passing because the main focus is on developing a synthetic voice for a restricted practical scenario, a market information system. A generic but practical strategy for training non-specialist personnel with linguistic and/or computational skills based on a restricted domain lexicon, Finite State techniques and a traditional rule-based diphone synthesis method are described.

It is indeed a pleasure to pay tribute to the authors and co-authors who contributed to this Festschrift as well as to the referees of the papers that were submitted. Appreciation is also herein acknowledged for the firm support of Professor Wannie Carstens (Director: School of Languages, North-West University, Potchefstroom Campus), Professor Hein Viljoen (former Director: Research Unit for Language and
Literature in the South African Context, North-West University, Potchefstroom Campus), Mrs Elsa van Tonder and the entire staff of the Centre for Text Technology (CTeXT). They carried this project with us from the moment they were told about it to the final stages of the publication of this book. They also managed to prevent information about this project from reaching Justus Roux’s ears during various conversations. The project received funding from the Research Unit for Language and Literature in the South African Context, Potchefstroom Campus, North-West University.

Finally, honouring Professor Justus Christiaan Roux through this volume is an immense pleasure for both the editors and the contributors.

H Steve Ndinga-Koumba-Binza
Sonja E. Bosch
PART 1:
PHONETICS AND PHONOLOGY
CHAPTER 1

STRESS ASSIGNMENT IN BLACK SOUTH AFRICAN ENGLISH

Sabine Zerbian

University of Potsdam, Potsdam, Germany
sabine.zerbian@uni-potsdam.de

1. INTRODUCTION

West-Germanic languages like English, German or Afrikaans are stress languages in which in any given polysyllabic lexical content word, such as a verb or a noun, one syllable carries the main stress. Stress is expressed acoustically by higher pitch, longer duration and/or higher intensity. In West-Germanic languages, the location of the stressed syllable within the word is not restricted to a specific syllable as it is in languages like French (last syllable) or Polish (penultimate syllable). The placement of stress in English is mainly determined by syllable weight and morpho(-syntactic) information (Giegerich 1992). There are only very few minimal pairs which are differentiated in their meaning by stress placement alone, such as ‘differ and de’fer for many speakers or noun-verb pairs such as ‘abstract and ab’stract (where a high colon appears before the stressed syllable).

Not all languages show the word-prosodic system of stress. In South African tone languages, for example, each syllable of a word carries a specific tone, either high or low. Tone is expressed acoustically by high pitch (not by increased duration), thus manipulating one of the same acoustic parameters as stress. However, in contrast to stress languages, the tonal specification of a syllable is determined lexically and/or morphologically. As a result, a word can have several high tones. Tone frequently differentiates between the meanings of words, e.g. Sotho bóna – ‘see’ versus boná – ‘they’, where an acute accent refers to a high tone and a low tone is left unmarked.

It has been reported in the literature that the West-Germanic fixed stress system poses a challenge in language contact. This holds for the acquisition of word stress in learner varieties (Dupoux et al. 1997; Dupoux et al. 2007) as well as for the establishment of word-prosodic systems in varieties emerging from genuine language contact, such as in bilingual speakers (Dupoux et al. 2009). Similarly, speakers of a language with a different word-prosodic system, such as a tone system, as well as speakers of a stress language whose stress assignment differs from the target language face difficulties with the rules that govern stress placement in a specific language (Peperkamp et al. 2010). The authors argue for
differences in the phonological representation of the first and target languages as the cause for the difficulties.

Against this background, it is not surprising that it has been repeatedly stated in the literature that one of the striking linguistic features of the contact variety Black South African English (BSAE; for a thorough discussion of this problematic term see Da Silva 2008:98f.) is stress placement at the word level that differs from that in General South African English (Lanham 1984, Wright 1996, De Klerk & Gough 2002).

The anecdotal evidence remained common “knowledge” until it was more thoroughly investigated in work by Van der Pas et al. (2000) and van Rooy (2002). Both studies reach the conclusion that word stress in Black South African English/Tswana English is less deviant from the Standard English target patterns than commonly believed. Whereas Van der Pas et al.’s (2000) analysis argues for speaker-dependent coherent stress systems, van Rooy’s (2002) work makes the strong claim that there are indicators of a single stable system for stress assignment in Tswana English across speakers.

The current contribution tests the claim by van Rooy (2002) of a speaker-independent coherent system of stress assignment in Black South African English against new empirical data. It thus tests the predictions of one theory by applying it to data from a new corpus. It honours Prof. Justus C. Roux by relating to several of his research interests, both in content and methodology. It particularly deals with suprasegmental information at the word-level, which has been the topic of his work on the Nguni languages (e.g. Roux 1995, 1998). The chapter furthermore concentrates on the reflexes in Black South African English, another focus of Prof. Roux’s work (cf. Roux & Louw 2001; Roux et al. 2005).

The remainder of the chapter is organised as follows: Section 2 presents the background to stress in BSAE and introduces van Rooy’s (2002) algorithm for stress assignment in this variety. Then, van Rooy’s (2002) algorithm is tested by comparing its predictions against a corpus of data which is described in section 3. The results are presented and discussed in section 4. Section 5 discusses the findings with respect to their implications for the phonological representation of stress in BSAE and with respect to the limitations of control over usage frequency in corpus data. The paper concludes in section 6.

2. DO WE NEED LINGUISTIC THEORY? RANDOMNESS AND SYSTEMATICITY

Previous scholars entertained the view that word stress in BSAE is “assigned idiosyncratically, very often on the penultimate syllable, following the phonological rule in Bantu languages where this syllable is lengthened” (de Klerk & Gough 2002:361; see also Hundleby 1964:80-81). In their experience “BSAE-speaking students of linguistics
indicate a very marginal ability to assign native-speaker stress pattern to words” (de Klerk & Gough 2002:361.).

Van Rooy (2002) developed two hypotheses from this position: the penult-hypothesis and the random-hypothesis. According to the penult-hypothesis, stress in BSAE is attracted to the penultimate position of a word. The preference for the penultimate position has been claimed to be due to the fact that in the South African Bantu languages, the penultimate syllable of a word is predictably lengthened. According to the random-hypothesis, stress in BSAE is assigned randomly because there is no coherent stress system in this variety of English. In his work, van Rooy (2002) argues that neither of the hypotheses can account for his data in a satisfactory way. Instead, he suggests an analysis within the theoretical framework of Optimality Theory (OT) (Prince & Smolensky 1993) which is best suited to account for the observable variation in stress placement. His analysis will be rephrased here, without going into the technical details of OT.

Starting out from the basic requirement that all lexical content words need to have one primary stress, van Rooy (2002) isolates three factors that account for almost all of the data. The first is that stress falls on the penultimate syllable, as in (1a, b). This observation is central in the penult-hypothesis, and explanatory reference is made to the prosodic system of South African Bantu languages. However, the preference for the penultimate syllable is a property of varieties of English in general and results from the tendency of English stress to fall on the right edge of the word, with the caveat that the final syllable is generally exempt from stress assignment, especially in words with more than two syllables (Chomsky & Halle 1968).

(1) Standard English Black South African English

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>‘conduct</td>
</tr>
<tr>
<td>b.</td>
<td>‘seventies</td>
</tr>
<tr>
<td>c.</td>
<td>re’lax</td>
</tr>
<tr>
<td>d.</td>
<td>de’bate</td>
</tr>
</tbody>
</table>

However, stress in BSAE does not always fall on the penultimate syllable, as seen in (1c, d), and these exceptions are not predicted by the penult-hypothesis. Van Rooy (2002) observes that in BSAE the phonological shape of the final syllable also determines stress assignment, more specifically either a consonant cluster in the coda of the final syllable (thus syllable weight; see 1c) or the presence of an underlying diphthong in the final syllable (see 1d) leads to such syllable receiving stress. For the consonant cluster the surface form is decisive, i.e. the phonetic output that a given speaker produces, not the underlying form. Van Rooy (2007) describes how the underlying and surface form of consonant clusters might differ in BSAE. However, for the diphthong + consonant sequence it is the underlying
form that counts. This is illustrated in (1d) by the vowel [e] as a BSAE allophone of the underlying diphthong /eɪ/ (as in pay) which attracts stress to the final syllable. Van Rooy (2002:151) summarises both observations in characterising the BSAE stress system as quantity-sensitive so that superheavy final syllables attract stress, where superheavy refers to syllables which contain either a vowel and two consonants in the coda, or an allophone of a diphthong and a coda consonant.

These generalisations of stress assignment in BSAE are said to apply to both morphologically simple and morphologically complex words. However, for the latter one needs to differentiate between two kinds of suffixes (or affixes more generally): on the one hand those which are incorporated into the stem for the purposes of stress assignment, called opaque suffixes by van Rooy (2002:153), and on the other hand those whose presence is ignored in stress assignment, called transparent suffixes.

The basis for the categorisation of a given affix as either opaque or transparent is somewhat unclear. Van Rooy (2002:154) seems to suspect a division along derivational (=opaque) and inflectional (=transparent) affixes, and points out that “the ostensibly transparent third person [-s] and past tense/participle [-d] suffixes of English verbs” actually follow the analysis of opaque suffixes. Other examples of opaque suffixes are the nominal suffixes -ment, -ion, -or and -ature as well as the adjectival suffixes -ant and -able. Examples of transparent suffixes are the gerund affix -ing, plural -s and comparative -er.

Van Rooy’s algorithm allows clear predictions as to which stress pattern one would expect for a given word in BSAE, provided that the underlying form (in the case of the diphthong /eɪ/), the surface segmental structure (for the coda clusters), and the morphological structure are known.

3. Reproducibility: Testing the Predictions of the Stress Algorithm

In order to test if the predictions of van Rooy's (2002) algorithm are borne out in new data, the stress patterns of polysyllabic content words in an already existing corpus of spoken Black South African English have been analysed.

The corpus contained the speech of 14 speakers of Black South African English. All participants were students at the University of the Witwatersrand, Johannesburg, aged between 19 and 30. Based on the phonological features of their speech and their performance in an English test (Quick Placement Test of English; QPT) they can be described as speakers of the mesolect (cf. van Rooy 2002). They all reported an African Bantu language as their first language and most gave English as or among their preferred languages (8/14). They obtained an average QPT score of 62/100.