OXFORD

The Life Cycle of Language

Past, Present, and Future

EDITED BY
Darya Kavitskaya AND Alan C. L. Yu

The Life Cycle of Language

The Life Cycle of Language

Past, Present, and Future

Edited by

DARYA KAVITSKAYA
AND
ALAN C. L. YU

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2023935893

ISBN 9780192845818

DOI: 10.1093/oso/9780192845818.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

To our teacher, Andrew Garrett

Contents

Acknowledgments x
Preface xi
List of Figures and Tables xiv
List of Abbreviations xviii
The Contributors xxii

I. RECONSTRUCTING THE PAST

1. The fall and rise of vowel length in Bantu 3
Larry M. Hyman

2. The rise and fall of rounding harmony in Turkic 23
Darya Kavitskaya and Adam McCollum

3. The life cycle of the Kuuk Thaayorre desiderative 34
Alice Gaby

4. Akan morphological 'reversal' in historical context 51
Mary Paster

5. Increasing morphological mismatch via category loss: The Spanish Future Subjunctive 68
Matthew L. Juge

6. Toward a non-teleological account of demonstrative reinforcement 85
David Goldstein

7. Typology and history of unusual traits in Nivaclé 99
Lyle Campbell

8. Greek ἔγνωκα and the perfect of PIE *$ǵneh_3$- 'know' 118
Jay H. Jasanoff

9. The surface position of Hittite subordinating *kuit* 125
H. Craig Melchert

10. PIE *meh_{2-} 'grow, be fruitful' and Proto-Basque *ma, *maha 'fruit': An apple by any other name… 137
Juliette Blevins

II. PHILOLOGICAL AND DOCUMENTARY PAST AND PRESENT

11. Paradigm structure in Sanskrit reduplicants 155
Donca Steriade

12. Sound symbolic words in Séliš-Ql'ispé 173
Sarah Thomason

13. Tone and morphological structure in a documentation-based grammar of Choguita Rarámuri 198
Gabriela Caballero

14. The structure of dialect diversity in Mono: Evidence from the Sydney M. Lamb Papers 214
Hannah J. Haynie and Maziar Toosarvandani

15. Recovering prosody from Karuk texts: Deciphering J. P. Harrington's diacritics 233
Clare S. Sandy

16. Stylistic differentiation in California Dene texts 254
Justin Spence

17. Winter story themes in Meskwaki: Familiar creatures seen with new eyes 268
Lucy Thomason

18. The material and the textual in documentation of Native American languages 280
Lisa Conathan

19. Community-participatory orthography development in the Máíjùnà communities of Peruvian Amazonia 291
Christine Beier and Lev Michael

20. The value of family relations for revitalization 315
Marianne Mithun

III. LOOKING FORWARD: NEW APPROACHES

21. Sound structure and the psycholinguistics of language contact 339
Molly Babel and Melinda Fricke

22. Child-directed speech as a potential source of phonetic precursor enhancement in sound change: Evidence from Cantonese 354
Alan C. L. Yu, Carol K. S. To, and Yao Yao

23. Paradigmatic heterogeneity and homogenization: Probing Paul's principle 371
Chundra Cathcart

24. Language change in small-scale multilingual societies: Trees, waves, and magnets? 386
Jeff Good

25. Gradualness and abruptness in linguistic split: A Nyulnyulan case study 399
Claire Bowern

References 412
Language Index 457
Subject Index 461

Acknowledgments

This book would not have come to fruition without the encouragement and help from many. We would like to first thank the external reviewers for all their efforts: Kathryn Franich, David Harrison, José Hualde, Brian Joseph, Hannah McElgunn, and Armin Mester. Special thanks to Henry MacKall for his assistance in compiling the references. Many thanks also to the contributors of this volume not only for their chapters, but also for the timeliness of their efforts!

Preface

Historical linguistics is a dynamic and vibrant field with multifaceted goals. Historical linguists reconstruct a language's linguistic past and reveal the history of its speakers. They elucidate mechanisms underlying language change and unravel complex interactions between people throughout the ages and space. Yet, and perhaps somewhat surprisingly, historical linguistics is not a field that concentrates exclusively on past histories, it also addresses questions concerning the development of language, be it in the past, in the present, or in the future, as it addresses questions of language change and evolution. Thus, historical linguistics deals with not only languages from a bygone past, but also languages that are still evolving. What is learnt about the past helps elucidate what the future might hold.

This volume brings together 25 chapters by an international group of linguists with diverse research backgrounds and at different stages of academic careers. The all-encompassing nature of historical linguistic research necessarily invites diverse perspectives and methodological concerns. This volume highlights this multifaceted nature of language change research with contributions from Indo-Europeanists, philologists, fieldworkers, language documentarians, theoreticians (e.g. phonologists, syntacticians, and semanticists), and experimentalists (e.g. laboratory phonologists and psycholinguists). This volume features languages from the distant past (e.g. Hittite, Sanskrit, Proto-Indo-European, Proto-Turkic, Proto-Bantu) to the present (e.g. Cantonese, Meskwaki, Mono, Nivaclé, and Séliš-Qlʼispé). It covers languages from all continents other than Antarctica.

The book is organized into three parts. The first part, "Reconstructing the past," focuses on contributions that address the bread and butter of historical linguistics. They deal with phonological, morphological, syntactic, and semantic changes that affect language development from a diachronic perspective. The first two chapters focus on sound change. Hyman addresses the history of length contrast in Bantu, distinguishing four vowel length systems. Hyman proposes that positional restrictions resulted in the loss of the vowel length contrast, and the shift of the type of prominence contributed to the creation of the contrast. Kavitskaya and McCollum provide an analysis of the rise and fall of rounding harmony in Turkic, suggesting that the loss of harmony in Turkic involves the contraction of the prosodic domain and the proliferation of invariant suffixes. Chapter 3 from Gaby deals with semantic change, discussing the co-existence of 12 desiderative constructions in Kuuk Thaayorre, which are argued to be the result of several successive waves of conventionalization. Chapters 4 and 5 concentrate on morphological change.

Paster talks about the reversal of past and perfect morphology in the negative in Akan, and Juge addresses a morphological mismatch in Spanish future subjunctive, which is attributed to category loss. Goldstein follows with a discussion of the non-teleological nature of the definiteness cycle in Romance, which refers to the development of a definite article from a demonstrative, followed by demonstrative reinforcement. Campbell addresses several unusual traits in Nivaclé at the phonological, morphological, and morphosyntactic levels. Jasanoff focuses on the development of the perfect of Greek 'recognize, know.' This chapter shows the importance of understanding the historical development in making claims about synchronic morphological irregularities. Melchert addresses the syntactic position of Hittite subordinating conjunction 'because.' Finally, Blevins argues that the source for one of the reconstructed forms for 'apple,' *méh$_2$l-o-, is not Proto-Indo-European, but Proto-Basque.

The second part, titled "Philological and documentary past and present," includes chapters that focus on the study of linguistic features on the basis of textual materials from the past or on the creation of documentary evidence that are crucial for historical and comparative research. Steriade offers new insights into the nature of reduplication in Sanskrit through the lens of modern day linguistic theory. Sarah Thomason examines the nature of sound symbolic words in Séliš-Ql'ispé, which provides an important empirical basis for the comparative study of Salishan languages. Caballero offers a meditation on the nature of grammar writing, focusing on the description of the interaction between tone and morphological structure in Choguita Rarámuri. Particularly for under-documented languages like Choguita Rarámuri, reference grammars often serve as the primary documentary evidence for historical and comparative work and are instrumental in many language reclamation efforts. Haynie and Toosarvandani examine Mono dialectology based on Sydney Lamb's fieldnotes and offer new insights into the internal subgroupings among Mono varieties. A significant part of historical linguistic research focuses on deciphering textual materials. Several contributions in Part II of this volume offer important insights in this endeavor. Sandy's contribution examines the transcriptional practice of J. P. Harrington, the prodigious fieldworker whose voluminous work on Native American languages is largely unpublished and is in often cryptically transcribed fieldnotes. Sandy focuses on Harrington's work on Karuk in particular, illuminating the many different uses of diacritics in his Karuk fieldnotes. Another dimension of philological examination of textual records involves the classification and contextualization of the textual materials themselves. Spence examines stylistic differences across texts from Athabaskan languages in California, while Lucy Thomason focuses on a set of texts from Meskwaki that share the same theme, but exhibit diverse linguistic as well as stylistic features. Conathan's chapter examines the creation of linguistic texts, problematicizing modern practices of language documentation. The case in point comes from the Massachusetts language where she examines the interactions

among oral discourse, textual documentation, and contemporary native language reclamation.

Studies regarding language change encompass not only the past and present, but also the future. Questions regarding the future trajectory of change are most pertinent in endangered language communities where the future survival and vibrancy of a language is most precarious. Part II of this volume includes contributions from scholars who are actively involved in revitalization and reclamation efforts in such communities, discussing not only the changes that endangered languages are experiencing, but also the scarce nature of the materials communities with few or no fluent speakers often have to confront and work with, and the ideological backdrop that sustains efforts of language maintenance and revitalization. Beier and Michael look at the intricacies of orthography design in Máíjùnà, proposing a model for community-participatory orthography development that emerged from collaboration between linguists and members of the communities of Peruvian Amazonia. Many languages left behind scant documentary footprints in the annals of time. Mithun suggests that impoverished historical records can be greatly enhanced if information from related languages are brought to bear.

The third part of the book, titled "Looking forward: New approaches," features contributions focusing on the theoretical and methodological basis of language change research. Babel and Fricke argue for the value of incorporating psycholinguistic findings in language contact research. Specifically, they look at how processes of cross-language interaction within an individual speaker shape sound patterns in the context of languages in contact. Yu, To, and Yao investigate the role of child-directed speech in sound change, showing that some, but not all, sound changes in progress in Hong Kong Cantonese exhibit enhancement effects, which have been hypothesized to be a source of incrementation in language change. Cathart analyzes the temporal dynamics of the leveling of vocalic and consonantal patterns of allomorphy in Middle and Early New High German using Bayesian modeling. Good studies how language change operates in small-scale multilingual societies, arguing that linguistic convergence and divergence may look different when the community exhibits what he refers to as a magnetic sociohistorical dynamics. Bowern examines changes in Nyulnyulan languages, where both gradual diffusionistic tendencies and abrupt splits are observed. These changes reflect complex sociohistorical patternings which strictly tree or wave models of language changes have difficulties capturing.

We titled this book "The Life Cycle of Language," to reflect the all-encompassing nature of historical linguistics and language change research. The goal is not to produce a handbook of historical linguistics, but to showcase the dynamism and inherently interdisciplinary nature of the field. We also want to take this opportunity to dedicate this volume to our mentor, colleague, and friend, Andrew Garrett. There is no one who better embodied the scope and spirit reflected in the contributions in this volume than Andrew.

List of Figures and Tables

Figures

1.1. Languages which have lost the PB vowel length contrast (shaded) *Source:* Guthrie (1967: 66) 7

1.2. Languages which have introduced penultimate lengthening (Dark gray squares) 14

3.1. Putative etymology of subjunctive coding of desire 46

3.2. Putative etymology of associated motion coding of desire 46

3.3. Continuum between PCIs and GCIs 48

4.1. The Tano languages *Source:* Adapted from Williamson and Blench (2000) 53

5.1. Tense, aspect, and mood in non-imperative, non-anterior (non-perfect) forms in Spanish 71

5.2. The role of the Spanish Indicative and Subjunctive in asserting propositions 71

5.3. Tense, aspect, and anteriority in Greek verbs 73

5.4. The five major Romance languages in Comrie's typology of future-time protases 81

13.1. Location of Choguita (Guachochi municipality, Chihuahua, Mexico) 199

14.1. Mono dialects according to Lamb (1957: 14–15) 215

14.2. Mono-speaking areas (modified from Babel et al. 2013: 452) 216

14.3. Page 1 from Notebook A, recording a meeting with Annie Wenz, in the Sydney M. Lamb Papers (Kinsman and Lamb 1953) 217

14.4. Approximate geographic extents of Lamb's dialect areas and distribution of speakers in Lamb's Mono field notes 219

14.5. Distribution of selected features: A shows reflexes of **ɲj* in **pohniŋya* 'skunk' and **tihiŋya* 'deer'; B shows reflexes of **ŋw* in **paŋwi* 'fish'; C shows reflexes of **mp* in **timpi* 'rock'; D shows stem for the meaning 'horse' 222

19.1. Locations of the four Máíjùnà communities in Loreto Peru (west to east): Puerto Huamán, Nueva Vida, Sucusari, and Totoya. *Source:* Modified from the base map in Gilmore et al. (2010, 299) 294

20.1. Map by Charles Edward, Public Domain *Source:* https://commons.wikimedia.org/wiki/File:Beaver_wars_map.jpg 316

20.2. Susquehannock Numerals: Campanius 1696: 160 317

20.3. Basic Verb Template: Iroquois Proper 326

22.1. Average vowel triangles (solid lines) for ADS and CDS across all the speakers, shown against each speaker's mean [a]/[i]/[u] formant values 360

22.2. Average tone triangles (solid lines) for ADS and CDS across all the speakers, shown against each speaker's mean T1/T2/T4 F0 onset and offset 361

22.3. Model predictions of the likelihood of (a) syllable-initial /n/ being pronounced as [n] and (b) labiovelars being pronounced as labio-velars 363

22.4. Word-specific percentages of labiovelars being pronounced as labiovelars 364

22.5. Model predictions of sibilant spectral mean before rounded and unrounded vowels as modulated by the speaking-rate-residualized duration of the sibilant. Error bars indicate 95% confidence intervals 367

22.6. "Caterpillar plots" for the conditional modes of the by-subject random slopes for the interaction between CONDITION and ROUND the 28 mothers. The participants are ordered by the conditional modes of the by-subject slope 368

22.7. Each panel shows a mother's the sibilant spectral means before rounded and unrounded in ADS and CDS conditions 369

23.1. Chronology of selected analogical changes represented by generalized additive model (GAM) smooth curves. Values of 1 correspond to the sound on the right-hand side of the arrow 379

23.2. Posterior estimates of trends for levels of LEMMA×MACRO-INFLECTIONAL CLASS×CHANGE. For readability, PretSg represents 1/3sg.pret.ind. and PretPl represents 2sg/1-3pl.pret.ind. Verbs that do not survive paste MHG are in parentheses 382

23.3. Relative chronologies of change types inferred from model results. For clarity, only intra-lemma chronologies are provided. Solid lines represent precedence relationships supported by 95% or more of samples (85% for dashed lines). For readability, PretSg represents 1/3sg.pret.ind. and PretPl represents 2sg/1-3pl.pret.ind 383

24.1. Lower Fungom and the surrounding region
Source: Map created by Pierpaolo Di Carlo 388

25.1. Two types of tree topologies: rakes and consistent splits 401

25.2. Rapid multiple migrations 402

25.3. Map of the Nyulnyulan languages
Source: Bowern (2012a) 404

25.4. Cognate counts across Nyulnyulan languages 408

Tables

2.1. Counts of harmony and disharmony aggregated over the entire Crimean Tatar corpus in Radlov (1896) 28

2.2. Percent harmony by preceding vowel aggregated over the entire Crimean Tatar corpus in Radlov (1896) 29

2.3. Counts and percent harmony for each town and syllable number in Radlov (1896) 30

3.1. Summary of constructions used to express desiderative mood 36
4.1. Constraints yielding morphological reversal in Ofori (2006a, 2006b) 55
4.2. Past, perfect, and negative markers in Tano languages
Source: Classifications from Stewart (1989: 225) 57
4.3. Morphological changes from Proto-Tano to Asante Twi 65
5.1. Third person plural subjunctive forms of Spanish *tener* 'to have' 71
5.2. Third-person singular forms of the active forms of Greek *γráfo* 'I write' 73
5.3. Mood distinctions in the Portuguese Future 75
5.4. Mood distinctions in the Catalan Preterit 76
5.5. Ibero-Romance Future Subjunctive forms of *ser* 'to be'/*ir* 'to go' and their Latin antecedents 78
5.6. Conditional and aspect in Romance 80
5.7. Latin and Romance tense/mood categories in conditional sentences 82
5.8. Tense and mood in subordinate clauses in Spanish and Portuguese 82
5.9. Loss of the synthetic perfective past indicative in French and Italian 83
6.1. Romance definite articles, distal demonstratives, and their precursors 88
6.2. The reinforcement of demonstratives in Romance 92
6.3. Demonstrative inventories in archaic Indo-European 93
6.4. Demonstrative inventories in Tocharian A and B 94
6.5. Demonstrative reinforcement in archaic Indo-European 95
6.6. Reinforced demonstratives in classical Greek 96
7.1. Nivaclé phonemic inventory 100
7.2. Some Matacoan cognates illustrating reflexes of Proto-Matacoan **l* 101
7.3. Cognates in Matacoan languages of Nivaclé *-eš* 'valency increasing affix' 104
7.4. Basic demonstratives in Nivaclé 105
7.5. Potential cognates of tense or tense-like markers in Matacoan languages 107
10.1. Proto-Basque/Proto-Indo-European sound correspondences after Blevins (2018) 143
13.1. Morphologically conditioned stress patterns of inflected verbs 204
13.2. Lexical and grammatical tone interaction in CR inflected verbs 207
13.3. The structure of the Prosody chapter in the CR reference grammar 210
14.1. Mono language data in the Sydney M. Lamb Papers The 'ID' refers to an Item Number in the Survey of California and Other Indian Languages at the University of California, Berkeley 218
14.2. The 37 Mono speakers represented in the Sydney M. Lamb Papers 'Location' is a geographic location corresponding to where the speaker was raised through age 18, as described in Lamb's notebooks or determined by census or other public records. 'Data' is a count of the total data points that served as the input to our analysis; for many speakers, there is likely more data available though it is less readily analyzed (e.g. it is included in a text) 220

15.1. Comparison of segmental transcriptions 235
15.2. Karuk word prosody examples 237
15.3. Most common diacritics used by Harrington 238
15.4. Summary of correspondences 252
19.1. Máíj̱ìkì vowel phonemes and corresponding graphemes in the MP and Velie orthographies 295
19.2. Máíj̱ìkì consonant phonemes, allophones, and corresponding graphemes in the MP and Velie orthographies; where the orthographies differ, MP is in the top cell and Velie in the bottom cell 295
23.1. Leveling of stem alternation between *aim-* and *am-* in the Old French present paradigm of *amer* 'love,' adapted from Trask 1996, Garrett 2008
Source: adapted from Trask 1996; Garrett 2008 372
23.2. Selected Middle and New High German Class I, II, and III preterite paradigms, displaying patterns found before and after the operation of leveling of allomorphy in stem vowels and stem-final consonants, with vowels and consonants affected by leveling shown in bold. Note that the *t* ~ *d* variation found in MHG *vinden* is unlikely to reflect variation brought about via VL (see Jones & Jones 2019: 123ff.) 376
23.3. Reflexes of *verlieren* 'lose,' *ziehen* 'pull,' and *frieren* 'freeze' in Central Bavarian, East Franconian, and East Swabian dialects, coded according to expected (unbolded) or innovated (bolded) features that are displayed 377
24.1. Independent pronouns across the five Mungbam varieties 390
24.2. Mundabli possessive pronouns and determiners 392
24.3. Sample Fang noun class agreement patterns 393
24.4. Mundabli noun class system 395
24.5. Buu noun class system 395
24.6. Abar noun class system 396

List of Abbreviations

1	first person
1CON	first conjugation
2	second person
3	third person
A	Alavese
ABL	ablative
ACC	accusative
ACT	active
ADS	adult-directed speech
ADV	adverb
AGT	grammatical agent
AOR	aorist
ART	article
B	Basque
BEN.APPL	benefactive applicative
BGD	*Becerro Galicano Digital*
CAUS	causative
CDS	child-directed speech
CISL	cislocative
CLF	classifier
CM	combining form
COMP	complementizer
COND	conditional
CONJ	conjunction
DAT	dative
DEF	definite
DEF	definite article
DEM	demonstrative
DEM: SP.PRX	speaker-proximal demonstrative
DEP	dependent
DIR	direction toward
DN	divine name
DU	dual
DUB	dubitative
DV	duplicative
EHHE	*Euskal Hiztegi Historiko-Etimologikoa* (Lakarra et al. 2019)
ENHG	Early New High German
ERG	ergative

F	feminine
FEM	feminine
FUT	future
G	Gipuzkoan
GEN	genitive
GN	place name
GO&	associated motion
HAB	habitual
HKC	Hong Kong Cantonese
HN	High Navarese
HORT	hortative
HUM	human
IDS	infant-directed speech
IE	Indo-European
IMP	imperative
IMPER	imperative
IMPF	imperfect
IMPFV	imperfective
IMPV	imperative
INCL	inclusive
IND	indicative
INDEF	indefinite
INST	instrumental
L	Lapurdian
LAT	Latin
LN	Low Navarrese
LOC	locative
M	masculine
MASC	masculine
MB	Medieval Basque
MED	mediopassive
MHG	Middle High German
MID	middle
MP	Máíjì-kì Project
N	neuter
NEG	negative/negation
NHG	New High German
NHUM	non-human
NMZ	nominalizer
NOM	nominal suffix
NOM	nominative
NPOS	non-possessed
NPST	nonpast
NVIS	demonstrative 'not visible'
OBJ	object

OBL	oblique
OEH	*Orotariko Euskal Hiztegia* (Michelena & Sarasola 1987–2005)
P	plural
P.IPFV	past imperfective
P.PFV	past perfective
PASS	passive
PAT	grammatical patient
PB	Proto-Basque
PCPL	participle
PERI	periphrastic
PIE	Proto-Indo-European
PL	plural
PLPF	pluperfect
PN	personal name
POS	possessive
POSS	possessive
PRAG	discourse pragmatic morpheme
PRES	present
PRET	preterit
PRF	perfect
PRFV	perfective
PROHIB	prohibitative
PRT	partitive
PTCL	particle
PTCP	participle
R	Roncalese
RDP	reduplication
REFL	reflexive
REL	relative
REL.PROP	relational proprietive
REM	demonstrative 'no longer extant (removed)'
REP	demonstrative 'reported'
REP	repetitive
RS	read speech
S	Salazarese
S	singular
SBJV	subjunctive
SG	singular
SPPAS	SPeech Phonetization Alignment and Syllabification
ST	stative
SUBJ	subjunctive
TAM	tense/aspect/mood
TOP	toponym
V	Bizkaian
V^	valence increase

VAL	valency-changing
VIS	demonstrative 'visible'
VL	Verner's Law
VOT	voice onset time
Z	Zuberoan

The Contributors

Molly Babel is Associate Professor in Linguistics at the University of British Columbia, where she is the director of the Speech in Context Lab. Babel's research circles around questions related to variability in sound structure. At its core, Babel's research is phonetic in nature and is cross-pollinated by empirical and theoretical insights from a range of disciplines, notably, psycholinguistics, sociolinguistics, and the cognitive sciences. Interacting language systems within-individuals (e.g. accent, dialect, and language contact; the perception-production link) and across-individuals (e.g. adaptation and accommodation) are a constant focus of her work.

Christine Beier is an Assistant Adjunct Professor in the Department of Linguistics at the University of California, Berkeley and a co-founder of Cabeceras Aid Project. Her work focuses on the documentation, description, revitalization, and revalorization of endangered languages, primarily in Peruvian Amazonia, in tandem with humanitarian work promoting the well-being of local participants and communities.

Juliette Blevins is Presidential Professor of Linguistics at the Graduate Center, CUNY where she continues research and teaching in phonology, morphology, linguistic typology, and historical linguistics, and heads the Endangered Language Initiative. Her influential book *Evolutionary Phonology* (CUP 2004) presents an explanatory theory of sound patterns and sound change, and is the foundation of subsequent research, including *Advances in Proto-Basque Reconstruction with evidence for the Proto-Indo-European-Euskarian Hypothesis* (Routledge 2018). Blevins holds a Ph.D. in Linguistics from MIT, and has taught at the UT Austin; the University of Western Australia; Stanford; the University of California, Berkeley; and the University of Leipzig.

Claire Bowern is Professor of Linguistics at Yale University. Her research focuses on the languages and history of Indigenous Australia. She has been working with speakers of endangered, Indigenous languages since 1998, both through fieldwork with Northern Australian communities and through archival work with historical records. Her work combines traditional methods of language documentation and reconstruction with computational and statistical phylogenetic methods. She is the editor of the journal *Diachronica*.

Gabriela Caballero received her BA in Linguistics from Universidad de Sonora (Mexico) and her Ph.D. in Linguistics from UC Berkeley. Her research focuses on the description and documentation of indigenous languages of the Americas (especially Uto-Aztecan languages) and the nature of intralinguistic and crosslinguistic variation in morphology and phonology. She has a deep commitment both to the careful study of particular languages and to bring data from lesser-studied languages to bear on topics in formal phonology and morphology. She is especially interested in linguistic description and collaborative language documentation projects whose products serve both academic linguists and indigenous communities.

Lyle Campbell (Ph.D. UCLA) is professor of Linguistics University of Hawai'i Mānoa (emeritus). He has been professor or visiting professor in Australia, Brazil, Canada, Finland, Germany, Mexico, and Spain, and has held appointments in Linguistics, Anthropology, Spanish, Behavioral Research, and Latin American Studies. He has published 24 books and c.200 articles, and won the Linguistic Society of America's "Leonard Bloomfield Book Award" twice, for *American Indian Languages* (1997) and *Historical Syntax in Cross-linguistics Perspective* (1995, by Alice Harris & Lyle Campbell). His specializations include: language documentation and revitalization, historical linguistics, indigenous languages of the Americas, typology, fieldwork, and Uralic languages.

Chundra Cathcart is a senior researcher in the Department of Comparative Language Science at the University of Zurich. His research deals with issues in diachronic linguistics and evolutionary linguistics, and he employs a wide range of quantitative and computational tools to address questions regarding language change.

Lisa Conathan is Head of Special Collections at Williams College, overseeing the Chapin Library of Rare Books and the College Archives. Her research centers on the creation and reuse of endangered language documentation, focusing on vernacular literacy and the revival of textual documentation. She teaches workshops that equip participants to liberate documentation through archival methods, and is on the faculty of the Rare Book School. She holds a BA in Linguistics from Dartmouth College, a Ph.D. in Linguistics from the University of California, Berkeley, and a Master of Library Science from the University of Maryland.

Alice Gaby is Associate Professor of Linguistics at Monash University, Australia. Her research focuses on three interrelated domains: semantic and structural typology; the relationship between language, culture and cognition; grammatical description and language reclamation. She has collaborated with speakers of various Paman languages (Cape York Peninsula, Australia) since 2002, producing, e.g. *A Grammar of Kuuk Thaayorre* (2017). She is Vice-President of the Australian Linguistics Society and Deputy Chair of Living Languages (formerly, RNLD).

David Goldstein is Professor of Linguistics, Indo-European Studies, and Classics at the University of California, Los Angeles. He is the author of *Classical Greek Syntax: Wackernagel's Law in Herodotus* (Brill 2016) and numerous articles spanning syntax, semantics, lexicography, linguistic variation, and phylogenetics. His research has been supported by the Fulbright Program, the National Endowment for the Humanities, and the John Simon Memorial Guggenheim Foundation.

Jeff Good is a Professor in the Department of Linguistics at the University at Buffalo, the State University of New York. His research interests center on morphosyntactic typology, language documentation, and the comparative linguistics of the Niger-Congo language family, with an emphasis on Bantu languages and their closest relatives. He is presently engaged in documentation projects focused on a number of endangered Bantoid languages of Cameroon and the multilingual practices of their speakers.

Hannah J. Haynie is Assistant Professor of Linguistics at the University of Colorado, Boulder. Her interests include linguistic diversity and language change, and she incorporates ideas and methods from geography, ecology, and biology into her research on language and

linguistics. Her work has investigated global patterns in morphosyntactic diversity as well as spatial patterns in California language diversity and genealogical relationships between California languages.

Larry M. Hyman, Professor of Linguistics at the University of California, Berkeley and Director of the France-Berkeley Fund, has worked extensively on phonological theory, tone systems, linguistic typology, and the descriptive, comparative and historical study of Bantu and other African languages within the Niger-Congo family. His publications cover both general and African linguistics including several descriptive grammars as well as theoretical, typological, and historical articles in phonology, morphology, and syntax. A past Guggenheim Fellow, Larry Hyman chaired the Berkeley Department of Linguistics from 1991 to 2002, has directed the France-Berkeley Fund since 2010, and served as 2017 president of the Linguistic Society of America.

Jay Jasanoff earned his Ph.D. at Harvard, where he is Diebold Professor of Indo-European Linguistics and Philology He has taught at Berkeley (1969–70), Harvard (1970–78, 1998–), and Cornell (1978–98). He is a historical linguist who specializes in the early Indo-European languages and their prehistory. His book, *Hittite and the Indo-European Verb* (Oxford, 2003), proposes the first major revision of the Proto-Indo-European verb system since the decipherment of Hittite in 1915. His more recent *The Prehistory of the Balto-Slavic Accent* (2017) tries to show how the PIE accentual system was transformed into the much more complicated systems of Lithuanian, Latvian, and Slavic.

Matthew L. Juge is an Associate Professor of Linguistics at Texas State University. His research primarily addresses morphological and semantic factors in the development of verb systems, especially suppletion and grammaticalization. He is working as co-editor on a volume on the diachrony of suppletion and is writing a monograph on the typology and development of morphological mismatch. His most cited publications concern the development of suppletion in verbs (1999, 2013), the development of the Catalan periphrastic preterit (2006, 2008), and the role of metaphor and teleology in grammaticalization (2007).

Darya Kavitskaya is an Associate Professor in the Department of Slavic Languages and Literatures and in the Linguistics Department at the University of California, Berkeley. Her main research interests are the preservation and loss of contrast and opacity. The phonological phenomena she has worked on most extensively are palatalization and vowel harmony. Areally, her work is concentrated on the phonological issues in Slavic, Turkic, and Uralic. Her work in phonology is connected to other linguistic fields such as historical linguistics, phonetics, and language acquisition.

Adam McCollum (Ph.D. 2019, University of California San Diego) is an Assistant Professor in the Department of Linguistics at Rutgers University. Drawing on experimental, computational, and fieldwork methodologies, his research program examines the representational structure and complexity of phonological patterns.

H. Craig Melchert is Distinguished Professor of Linguistics and A. Richard Diebold Professor of Indo-European Studies Emeritus, University of California, Los Angeles. He previously taught at the University of North Carolina, Chapel Hill, for 29 years. His research centers on historical linguistics, especially the Anatolian sub-family of Indo-European. His interests include the modeling of language change and the application of modern linguistic

formal theories to ancient Indo-European languages and their prehistory. Major publications include *Anatolian Historical Phonology, A Dictionary of the Lycian Language*, and (with Harry A. Hoffner Jr.) *A Grammar of the Hittite Language*.

Lev Michael is a Professor in the Department of Linguistics at the University of California, Berkeley. His research focuses on the documentation and description, typology, and historical and contact linguistics of Amazonian languages.

Marianne Mithun is Professor of Linguistics at the University of California, Santa Barbara. Her interests range over the areas of morphology, syntax, discourse, prosody, and their interrelations; typology; language contact and language change, especially the development of grammar; language documentation; languages indigenous to North America, especially Iroquoian (Mohawk, Cayuga, Tuscarora, Seneca, Oneida, Onondaga), Pomoan (Central Pomo), Yupik-Inuktitut-Unangan (Central Alaskan Yup'ik), Dene (Navajo), and Siouan (Dakota/Lakota, Tutelo); and Austronesian languages (Kapampangan, Hiligaynon).

Mary Paster, Professor of Linguistics and Cognitive Science at Pomona College, is a phonologist and morphologist whose research is both theoretical and descriptive. She is best known for her contributions to the theory of the phonology–morphology interface and for her descriptive work, primarily on languages of sub-Saharan Africa. Her work appears in journals such as *Natural Language and Linguistic Theory*, *Phonology*, *Studies in African Linguistics*, *Yearbook of Morphology*, and *Word Structure* (of which she currently serves as Editor).

Clare S. Sandy is a lecturer in Linguistics and Language Development at San José State University. She received her Ph.D. in linguistics from University of California, Berkeley. Her primary research areas are phonology and morphology, and her dissertation focused on the prosody of the Karuk language. She has also worked on documentation and description of the Tupí-Guaraní language Omagua. An ongoing interest is making historical linguistic documentation accessible and usable for research and revitalization.

Justin Spence is Associate Professor of Native American Studies at the University of California, Davis. His research focuses on Indigenous languages of California, drawing on data from archival sources and original fieldwork with contemporary communities. He uses descriptive materials to explore historical relationships in the region, both over relatively long timespans and more recent developments due to language and dialect contact. He is also active in efforts to bridge the gap between academically oriented linguistic research and the needs of language revitalization practitioners, primarily through the development of digital resources that facilitate learning for diverse audiences.

Donca Steriade is the Class of 1941 Professor of Linguistics at the Massachusetts Institute of Technology. She studies the phonological influences that lexically related words exert on each other, the effects of perceptual similarity on phonology, and the representation of rhythmic units. Her contribution to this volume is a third attempt to understand reduplication in Sanskrit, after Steriade 1982, 1988.

Lucy Thomason began helping with documentation of the easternmost Salish languages, Bitterroot Salish and Pend d'Oreille, in 1982. She began helping with documentation of a Central Algonquian language, Meskwaki, in 1992. Since 1998 she has been an employee

of the Smithsonian Institution's National Museum of Natural History. Her dissertation (University of Texas at Austin, 2003) discussed the five different kinds of third persons in Meskwaki and their use for purposes of clarity, perspective, pathos, and drama. She has published various editions and translations of Meskwaki stories and assorted articles about aspects of Salish and Meskwaki linguistics. She co-authored, with Ives Goddard, *A Meskwaki-English and English-Meskwaki Dictionary Based on Early Twentieth-Century Writings by Native Speakers* (2014), and co-edited, with David Costa and Amy Dahlstrom, *Webs of Relationships and Words from Long Ago: A Festschrift Presented to Ives Goddard on the Occasion of his 80th Birthday* (2021).

Carol K. S. To is Associate Professor and Director of Clinical Education in the Unit of Human Communication, Development, and Information Sciences at The University of Hong Kong. Carol is a qualified speech-language pathologist. She researches and teaches in the areas of speech and language acquisition and disorders in Chinese.

Maziar Toosarvandani is Associate Professor of Linguistics at the University of California, Santa Cruz. His primary research interests are in syntax and semantics, and he has current projects on the principles governing the creation of meaningful discourse and the grammatical representation of reference. He has collaborated with speakers of Northern Paiute in eastern California and Zapotec varieties from the Sierra Norte of Oaxaca to develop and share knowledge about their languages with their communities.

Yao Yao is Associate Professor in the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University. She is interested in language change and variation, heritage phonological acquisition, and speech processing in bilinguals. She has worked on multiple Chinese languages (Mandarin, Shanghainese, Cantonese, etc.) and English.

Alan C. L. Yu is the William Colvin Professor of Linguistics and the College at the University of Chicago. His research focuses on language variation and change, particularly from an individual-difference perspective. He is the author of *A Natural History of Infixation* (2007, OUP) and the editor of *Origins of Sound Change: Approaches to Phonologization* (2013, OUP) and co-editor, with John Goldsmith and Jason Riggle, of *The Handbook of Phonological Theory, 2nd Edition* (2011, Wiley Blackwell). He is co-General Editor of *Laboratory Phonology, the Journal of the Association of Laboratory Phonology*.

PART I

RECONSTRUCTING THE PAST

1
The fall and rise of vowel length in Bantu

Larry M. Hyman

1.1 Introduction

The goal of this chapter is to trace the history of vowel length in Bantu. I will discuss the stages and processes by which the long vowels reconstructed in Proto-Bantu (PB) (Meeussen 1967: 82) have been shortened and ultimately lost in many Bantu languages, most of which innovate new vowel lengthening, hence the title "fall and rise of vowel length in Bantu." Since the status of vowel length can be rather different in one vs. another Bantu language, the survey I will present will also document some of the potential functions of vowel length not only in Bantu, but in language in general. My starting point will be to see to what extent vowel length can fulfill all the functions that have been documented for tone (§2), after which I will turn to consider length in PB (§3) and successively in different Bantu languages (§4, §5). I conclude with a return to the question of what vowel length can do in language, based on the demonstration from Bantu (§6).

1.2 The functions of tone

As stated in Section 1.1, a general concern is to determine how vowel length stacks up with respect to tone, which has been claimed to be able to be the most versatile of phonological properties:

> tone can do everything segments and non-tonal prosodies can do, but segments and non-tonal prosodies cannot do everything tone can do.
>
> (Hyman 2011: 214; 2018: 699)

A quick example of something that only tone can do is seen in (1) from Giryama [Bantu E72a] (Volk 2011: 17).[1]

[1] Bantu languages will be cited with their Guthrie (1967–1971) referential letter and number classification as further developed by Maho (2009). Throughout this study, length is indicated by doubling the vowel.

Larry M. Hyman, *The fall and rise of vowel length in Bantu*. In: *The Life Cycle of Language*. Edited by: Darya Kavitskaya and Alan C. L. Yu, Oxford University Press. © Larry M. Hyman (2023). DOI: 10.1093/oso/9780192845818.003.0001

(1) a. ni-na-mal-a ku-gul-a ŋguuwo 'I want to buy clothes'
b. a-na-mal-a ku-gul-a ŋguúwo 'he/she wants to buy clothes'
H

In (1a) all of the tone-bearing units (TBUs) are underlyingly toneless, pronounced with default L(ow) tone. In (1b) there is a H(igh) tone on the penultimate mora, marked with an acute accent. As can also be observed, the only grammatical difference in (1b) is the subject prefix, which introduces a H tone in (1b). As seen, the underlying H of /á-/ 'he/she (class 1)' shifts two words to the right, landing on the penultimate mora of the phonological phrase.[2] No other phonological property has the ability to shift long-distance in this fashion.

We also know that tone can have different functions: lexical, morphological, syntactic, pragmatic. The Luganda [JE15] examples from Snoxall (1967) in (2) show a lexical tone contrast on verb roots synchronically analyzed as /H/, as in (2a) or toneless, as in (2b) (Hyman & Katamba 2010: 70):

(2) a. /-bál-/ 'bear fruit' b. /-bal-/ 'count'
/-síng-/ 'surpass' /-sing-/ 'pledge, bet'
/-búul-/ 'propose' /-buul-/ 'open eyes after birth'

The morphological function of tone is seen in examples from Noni, a Bantoid language of Cameroon, where the plural of the following L tone singular class 9 nouns is expressed by H tone (Hyman 1981):

(3) singular:	jè	còn	gvùw	bìè	nsààn	mbàsè
plural:	jé	cón	gvúw	bíé	nsáán	mbásé
	'road(s)'	'hut(s)'	'chest(s)'	'fish'	'friend(s)'	'vegetable(s)'

An example of the syntactic function of tone can be seen in Kilega [D25] (Meeussen 1971: 20), where (in certain tenses) a H tone marks a verb followed by direct object:

(4) a. be-ko-bolot-a tɔŋgɔ 'they are pulling also'
b. be-ko-bolót-á mózígi 'they are pulling the rope'
H

As seen, the verb is entirely toneless when followed by an adverb in (4a), but acquires a H tone when followed by an object. Finally, the pragmatic function of tone can be seen in Lusoga [JE16; Uganda] (personal notes), marking the difference between an imperative utterance which is a command vs. a suggestion:

[2] Volk (2011) also considers the possibility that the H shifts to the final mora, but is subsequently pulled back to the penult, a position of prominence. As we will see in Section 1.5, many Bantu languages automatically lengthen a phrase-penultimate vowel.

(5) a. gùl-à è-kì-tàbò 'buy a book!' (a command)
b. gùl-à è-kí-tábó 'buy a book!' (a suggestion)
H

When the imperative is a command, as in (5a), the toneless noun *e-ki-tabo* 'book' is realized all L; when it is instead a suggestion, e.g. as an answer to the question, 'what should I do?,' a H% boundary tone is realized on the last three TBUs of the noun. The weaker imperative in (5b) corresponds tonally to an ordinary declarative utterance which also takes the H%: /a-gul-a e-bi-tabo/ → *a-gúl-á é-bí-tábó* 's/he buys books.'

The above examples illustrate the versatility of tone, the various functions tone can have within Bantu and in language in general. To repeat, "tone can do everything segments and non-tonal prosodies can do, but segments and non-tonal prosodies cannot do everything tone can do," raising the question of whether vowel length can do ALMOST everything tone can do? I will come back to the extent to which the same functions can be fulfilled by vowel length in §6, after we have considered the rather disparate manifestations of vowel length in the Bantu family of ca. 500 languages.

1.3 Vowel length in Proto-Bantu

Proto-Bantu (PB) is generally assumed to have had a contrast between long and short root vowels (Meeussen 1967, 1969, 1980). Examples from Bastin et al. (2002) are given in (6).

(6)	*tín-	'cut'	*vs.*	*tíːn-	'fear, run away'
	*kód-	'grow up'	*vs.*	*kóʊd-	'pull out'
	*tók-	'abuse'	*vs.*	*tóʊk-	'come from'

However, Meeussen (1979[1954]) pointed out that many of the long vowel reconstructions are verb forms, some of which may have had an internal CV-VC- bound root+suffix structure:

(7)	*dìik-	'bury'	<	**dì-ik-	cf.	*dìam-	'sink, be in earth'
						*dìʊd-	'pull up (out of ground)'
						*dìʊk-	'come out of the ground'
	*dúuk-	'take off (clothes)'	<	**dú-ʊk-	cf.	*dúat-	'wear'
						*dúɪk-	'clothe'
	*tóʊd-	'put down (a load)'	<	**tó-ʊd-	cf.	*tóad-	'carry (on the head)'
						*tóɪk-	'put on head, give to carry'

While the bound roots do not exist on their own, Meeussen (1979: 4) proposed the glosses **di-* 'in, into or out of ground,' **dú-* 'on or from body,' and *tʊ́-* 'on or from head' (cf. **-tʊ́è* 'head'). The examples in (7) point to a major source of long vowels: juxtaposition of vowels (V+V) (de Chene & Anderson 1979; Kavitskaya 2002; Myers & Hanson 2005; and others). However, Meeussen also recognized that other cases of long vowels "do not permit any decomposition," e.g. **dóot-* 'dream,' **pèep-* 'blow (wind),' including nouns, e.g. **-béédè* 'breast,' **dèèdó* 'today.'

What's clear is that whole regions of Bantu languages contrast long and short vowels, often with minimal pairs, as in Lulamogi [JE10] (Hyman 2017: 66):

(8)	a.			b.		
		ó-ku-siβ-á	'to tie'		ó-ku-siiβ-á	'to fast'
		ó-ku-sen-á	'to draw (water)'		ó-ku-seen-á	'to become thin'
		ó-ku-tum-á	'to send'		ó-ku-tuum-á	'to jump'
		ó-ku-hol-á	'to lend (money)'		ó-ku-hool-á	'to differentiate between'
		ó-ku-many-á	'to know'		ó-ku-maany-á	'to pluck'

In addition, while not present in PB, many Bantu languages like Lulamogi also support length contrasts in post-root position, as in (9a)[3]:

(9)						
	a.	ó-ku-lagir-á	'to command'	vs.	ó-ku-tamíír-á	'to become drunk'
		ó-ku-lekér-á	'to cease'		ó-ku-tegéér-á	'to know'
		ó-ku-sitúk-á	'to stand'		ó-ku-sihúúk-á	'to fade'
		ó-ku-tolók-á	'to run away'		ó-ku-tolóól-á	'to go around'
		ó-ku-sigál-á	'to stay'		ó-ku-liráán-á	'to become near'
	b.	ó-ku-leekán-á	'to make noise'	vs.	ó-ku-leekáán-á	'to shout'
		ó-ku-siiβúl-á	'to wave'		ó-ku-siimúúl-á	'to wipe'
		ó-ku-suuβíl-á	'to hope'		ó-ku-suuβíízy-á	'to fade'
		ó-ku-kaambúh-á	'to be fierce, scary'		ó-ku-kaaβúúk-á	'to go around'

This produces the possibility of more than one long vowel in the verb stem, as in (9b). On the other hand, as seen in Figure 1.1 from Guthrie (1967: 66), many Bantu languages have lost the inherited vowel length contrast.

The question is how? What I will now show is that they didn't just merge long and short vowels across the board, rather there were intervening steps. In the following discussion, I will distinguish four "types" of Bantu vowel length systems:

[3] Although rarely explicitly stated, I believe the general view is that vowel length did not contrast in pre- or post-root position in PB. Thus, Guthrie (1967–1971) speaks only of "the disappearance of the distinction *VV, *V > V in first position" (vol. 2, p. 56, §51.31). Non-etymological vowel length contrasts generally result from the loss of an intervocalic consonant. Thanks to Thilo Schadeberg for discussion of this issue.

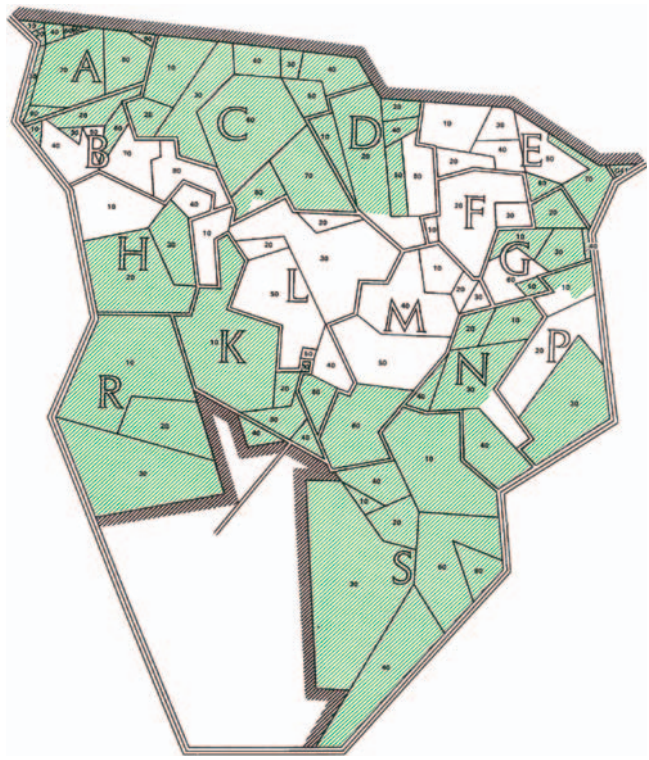


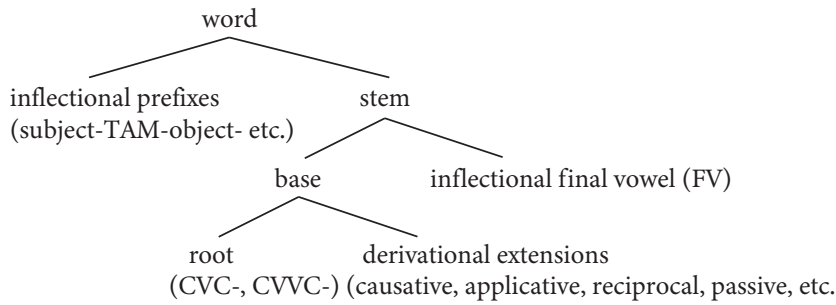
Figure 1.1 Languages which have lost the PB vowel length contrast (shaded).
Source: Guthrie (1967: 66).

(10) Type 1: the *V/*VV contrast survives and is extended without restrictions on where long vowels can occur (as just seen in Lulamogi)[4]
Type 2: the *V/*VV contrast survives with restrictions on where long vowels can occur
Type 3: the *V/*VV contrast is lost (with or without creation of new long vowels)
Type 4: the *V/*VV contrast is lost with predictable penultimate lengthening being introduced

To consider these types, we recognize the following traditional structure of the Bantu verb (Meeussen 1967).

[4] I know of no Bantu language that contrasts vowel length only on the root syllable and allows the length to surface without restriction. Cf. the discussion of type 2 systems in Section 1.4.

(11)



Type 1 languages are those which have preserved the *V/*VV contrast on roots and have typically extended the contrast to other positions as well. Again, Lulamogi [JE10] can serve as a typical example supporting long vowels in all positions:

(12) a. prefixes: tw-aaká-βál-á 'we have just counted'
b. root: ó-ku-huum-úl-á 'to rest'
c. extensions: ó-ku-lir-áán-á 'to become near, be close'
d. final vowel: ó-ku-tiis-y-áá =ku 'to frighten a little'

The long vowel in (12a) is from the gliding of the /u/ of /tu-aka-/ with the following /a/ undergoing compensatory lengthening (CL). While it is common, not all Bantu languages accompany gliding with CL. The root length in (12b) is as we saw in (8b). In (12c) we see a derivational suffix of the shape *-aan-* which likely comes from the loss of an intervocalic consonant, possibly a reduplicated *-an-an-* or *-agan-* (earlier *-*a(n)g-an-*) which otherwise exists as the reciprocal extension, e.g. *ó-ku-βon-ágán-á* 'to see each other.' Lastly, in (12d) the inflectional final vowel is long as a result of the gliding of the causative *-i-* suffix accompanied by CL.[5] In the following sections we will consider first type 2 systems (Section 1.4) and then types 3 and 4 (Section 1.5).

1.4 Type 2: Systems with vowel length restrictions

A number of Eastern and Western Bantu languages maintain the PB vowel length contrast on roots, but do not allow the length to be realized if the root is followed by too many syllables within the word or phrase. As seen in (13), depending on the language, vowel length may be restricted to occurring only in penultimate position, or it may be allowed if it is either penultimate or antepenultimate:

(13) a. penultimate syllable only, e.g. Cokwe [K11] (van den Eynde 1960: 17)
ku-huul-a 'peel off' *vs.* ku-hul-il-a 'to peel off for/at' (-il- 'applicative')

[5] The infinitive in (12d) is shown accompanied by the class 17 enclitic *=ku* 'a little' since the final vowel would otherwise undergo final vowel shortening in word-final position: *ó-ku-tiis-y-á* 'to frighten.'

b. antepenultimate or penultimate syllables only, e.g. Lunda [L52]
(personal notes from work with Boniface Kawasha)
ku-kwáat-a 'to hold, arrest'
ku-kwáat-ish-a 'to make hold, arrest' *vs.* ku-kwát-ish-il-a 'to make hold for/at'

Such restrictions exemplify three different general tendencies in languages. The first is the tendency for there to be more contrasts and contrast-preservation in "strong" or prominent positions, e.g. root and (ante-)penultimate syllables. In Punu [B43] not only are long vowels restricted to the root, but this also is the only position where the five vowels /i, ɛ, u, ɔ, a/ contrast (Kwenzi Mikala 1980: 8; Hyman 2019: 67–69). The second tendency is the widely reported process of "compensatory shortening" studied mostly in European languages: as more syllables are added, the stressed syllable is shortened, both at the word level (*speed* vs. *speedy*, *speedily*) and in syntactic combinations *(speed kills*) (Lehiste 1972). Punu is also susceptible: *u-wé:l-a* 'to marry' vs. *u-wél-↓án-a* 'to marry each other,' *mi:la* 'rivers' vs. *milá mya:mi* 'my rivers.'

The third tendency is the targeting of phrase-level shortening on the head noun or verb in specific syntactic contexts. A good case in point comes from Kimatumbi [P13] which has contrastive vowel length but two rules that shorten long vowels depending on position. The first, which Odden (1996: 157–162) terms "stem shortening," affects long vowels in pre-antepenultimate position within the word. This first process is consistent with "compensatory shortening":

(14) a. penultimate VV vs. pre-antepenultimate V
káat-a 'cut' kát-anik-a 'be cuttable'
nóol-a 'sharpen' nól-eyelw-a 'be sharpened up'
b. antepenultimate VV vs. pre-antepenultimate V
búund-ik-a 'store bananas' búnd-ikiy-a 'store bananas to complete ripeness'
chíɪl-ɪy-a 'be late' chíl-ɪkɪy-a 'be late for'

In addition, there is a second syntactically conditioned rule of phrasal shortening (Odden 1996: 218–233) which (among other things) affects nouns when followed by modifiers. The examples in (15) show such shortening occurring before a possessive pronoun, an adjective, a relative clause, and a determiner:

(15) a. ki-kóloombe 'cleaning shell' ki-kólombe chaángʊ 'my cleaning shell'
b. mi-kaáte 'loaves' mi-katé mikólʊ mikóló 'large loaves'
c. lu-kaámba 'string' lu-kambá lwalúpʊwáaniiké 'string which broke'
d. m-boópo 'machete' mbopó ye 'the machete'

The last example shows that shortening is not by syllable position, since the long vowel in the input /m-boópo ye/ should otherwise be allowed in phrase-antepenultimate position. The process is thus syntactically determined. In the case of nouns, it is clear that the "modifier" has to be within the same noun phrase, as in (16a), where *ki-kóloombe* 'cleaning shell' is shortened before the modifying adjective *kikółó* 'large':

(16) a. n-aa-m-péi ki-kólombe kikółó 'I gave him a large shell'
I-PAST-OP1-gave shell large

b. n-aa-m-péi ki-kóloombe kikółó 'I gave the large one a shell'
I-PAST-OP1-gave shell large

In (16b), on the other hand, the same word *kikółó* occurs in its own (headless) noun phrase and *ki-kóloombe* is not shortened.[6]

Verb stems also shorten their vowel when followed by an object (Odden 1996: 225–231). Prefixal length is not affected[7]:

(17) a. n-aa-kálaang-ite 'I fried'
b. n-aa-kálang-ite chóolyá 'I fried food'
I-PAST-fry-APPL.PFV food
c. n-aa-n-kálaang-iile 'I fried for him'
d. n-aa-n-kálang-ile lí 'I didn't fry for him'
I-PAST-OP1-fry-APPL.PFV NEG

In (17a) we see that *n-aa-kálaang-ite* 'I fried' has both a prefixal and stem long vowel. (17b) shows that only the latter is shortened when the object *chóolyá* is added. In the corresponding applicative verb form in (17c), there are two long stem vowels, both of which shorten in (17d). The fact that phrase-antepenultimate *-iile* is affected before the negative marker *lí* again shows that the process targets a non-final verb independent of the syllable position in which the long vowel appears. Crucially, as Odden (1996: 222) points out, only the head N of an NP, as in (16), or the head V of a VP, as in (17), can undergo shortening. Thus, in (18), the long vowel of the adjective *ki-keéle* 'red' is not affected, since it is not the head of the (zero-headed) noun phrase:

(18) a. ki-keéle chaángɔ 'my red (thing)'
red my
b. ki-keéle ki-kółó 'large red (thing)'
red large

[6] As seen, a class 7 noun (phrase) that refers to an animal optionally takes class 1 agreement, here the class 1 object prefix (OP_1) (Odden 1996: 32).

[7] Odden (1996: 225n) suggests that this is because the long vowel of *n-aa-* derives from *ni+a-* (1sg + past), with vowel coalescence counterfeeding the shortening rule. In the above and elsewhere OP1 stands for object prefix class 1, APPL = applicative, PFV = perfective.

This obligatory head-targeting property guarantees that there can be only one occurrence of phrasal shortening per immediate XP. Thus while *ki-kóloombe* undergoes vowel shortening in (19), neither *ki-keéle* nor *yaángʊ* do.

(19) a. ki-kólombe ki-keéle chaángʊ 'my red shell'
shell red my
b. i-kólombe yaángʊ yanaachímá 'my many shells'
shells my many

While the verbal examples show that prefixal length is not affected by shortening in Kimatuumbi, prefixes do undergo shortening in Safwa [M25] when followed by three or more moras (Voorhoeve, n.d.):

(20) a. a-gaa-gúzy-a 'he can sell'
b. a-ga-buúzy-a 'he can ask'
c. a-ga-buzy-aág-a 'he may ask'

These facts suggest a succession of changes as summarized in Hypothesis 1:

(21) *Hypothesis 1:* Positional effects and categorical VV > V started out at the stem level and only later generalized to the word and "tight" head+modifer/complement constituents.

According to this view, Kimatuumbi represents an earlier stage, which Safwa takes one step further by shortening vowels in the prefix domain. The last stage is to generalize to less tight phrasal configurations. In support of this direction of change, Kifuliiru [JD63] can be cited, where "any long vowel is shortened if it is followed by three or more morae within the domain of the phonological word," whether that word is phrase-final or not (van Otterloo 2011: 59):

(22) a. kú-húúmb-à 'to dig up something'
b. kú-húúmb-ír-à 'to dig up sth. for someone'
c. kú-húmb-írír-à 'to dig up intensively'

As seen, the long vowel of the root *-húúmb-* 'dig up' is maintained in penultimate (22a) and antepenultimate (22b), but not preantepenultimate position (22c). The same facts are seen in (23), where the length appears in the prefixal domain on /-gáá-/ 'distant future':

(23) a. à-gáà-ly-à 'he will eat'
b. à-gáá-hík-à 'he will arrive'
c. à-gá-bàlàm-à 'he will travel'

Van Otterloo (2011: 60) is unequivocal concerning the domain in which shortening occurs: "The fact that this rule does not apply across word boundaries

in phrases shows that this rule operates over the domain of the word, at the word-building (lexical) stage."

Further evidence for the stem domain comes from Ngangela [K12b], where a vowel can be long only if all of the vowels that follow it up to the penult are also long (Maniacky 2002: 20):[8]

(24)	a.	-tééta	'cut'	-teetááŋga	'share'
				-teetaaŋgééni	'share! (pl.)'
	b.	-vuulwííθa	'recall, remind'	-vulúka	'remember'
		-taambwííθa	'distribute'	-tambúla	'receive'
		-ʃaambwííθa	'infect, contaminate'	-ʃambúka	'be contaminated'
	c.	-púla	'cut with a knife'	-pulááŋga	'cut into slices'
		-holóka	'cool, calm (intr.)'	-holwééθa	'cool, calm (tr.)'
		-áʃa	'throw, launch'	-aʃááŋga	'reach several times'

In (24a) we see that the long vowel of *-tééta* 'cut' is maintained in the related verbs to the right, since all the vowels are long up to the penult. This contrasts with (24b) where the root syllable has a long vowel in the forms on the left, which has to be shortened in the forms on the right because the penult is short. That there is no rule of lengthening of prepenultimate vowels is seen in (24c): the root vowels in the forms on the left remain short in the verbs to the right where the penultimate vowel is long. The interpretation I would like to give to the Ngangela facts is that vowel length cannot occur in a less prominent position (pre-penultimate) without it also occurring in a more prominent position (penultimate). Otherwise the long vowel shortens as in the forms to the right in (24b).[9]

The above Ngangela facts support the key idea of Hypothesis 1 that shortening begins as a stem- then word-level process. It can however potentially expand, as it optionally does to noun + possessive constituents, which have a particularly tight bond in other Bantu languages: *ŋgóombe yáaŋge* or *ŋgómbe yáaŋge* 'my cow' (Maniacky 2002: 20). In other words, what is systematic at the word level is transitional within the noun phrase.[10] This brings us to the second hypothesis:

[8] That the penultimate condition holds only at the stem level is seen from the fact that prefixal long vowels can surface without meeting the requirement, e.g. /tu-éé-ku-món-a/ → *tw-ée-ku-món-a* 'we see' (1PL-PRES-INF-see-FV) (Maniacky 2002: 135).

[9] This idea of prominence would have to be scalar: the further away a syllable is from the penult, the less prominent it is. An alternative would be an ad hoc constraint against a CVV.CV.CVV sequence, i.e. a short vowel flanked by long vowels—in other words, a CVV.CV sequence would only be permitted if the long vowel is in the penultimate syllable. Under either interpretation a pre-penultimate long vowel must be followed by another long vowel.

[10] This is opposite to the view I took in Hyman (2013), where I assumed that vowel shortening begins at the phrase level and gradually narrows down.

(25) *Hypothesis 2:* Positional restrictions are subject to being generalized, ultimately leading to the loss of vowel length contrasts in all positions, as in many Bantu languages.

We now turn to consider what happens in such languages.

1.5 Types 3 and 4: Systems which have lost the PB vowel length contrast

I identify as Type 3 those languages which have lost the PB vowel length contrast with or without introducing new long vowels via consonant deletion and vowel assimilation. These are distinguished from type 4 systems which have also lost the PB vowel length contrast, but have introduced phrase-level penultimate lengthening. As seen in Figure 1.2, type 4 systems cluster in Eastern and Southern Bantu (the dark grey squares), while the type 3 systems are found further to the west (the light grey diamonds).[11] In other words most of the zone D-S Bantu languages which have lost the vowel length contrast also have phrase-level penultimate lengthening (PUL), e.g. Shona [S10], where the length on the root in (26a) is non-contrastive:

(26) a. ku-té:ng-á 'to buy'
b. ku-téng-é:s-á 'to sell, cause to buy' (causative -es-)
c. ku-téng-és-é:r-a 'to sell for/at' (applicative -er-)
d. ku-téng-á za:nze 'to sell fruit'

As seen in (26b, c) when the causative *-es-* and applicative *-er-* suffixes are added, the suffix vowel in penultimate position is lengthened. The example in (26d) shows that PUL is a phrase-level process: only the last word in the phrase undergoes lengthening. However, the size of the phrasal domain and the application of PUL vary considerably across Bantu languages. This brings us to Hypothesis 3:

(27) *Hypothesis 3:* PUL started out as an intonational property of utterances and then underwent "boundary narrowing": *Utterance* > *Intonational phrase* > *Phonological Phrase* > *Word*.

[11] I would like to thank Guillaume Segerer for producing Figure 1.2 for publication in Hyman (2013: 312). While there is an outlier far to the west, I am still unclear as to whether (and which) Myene [B10] languages of Gabon may have penultimate lengthening. In my 2013 study I was careful to consider only languages where the lengthening was phonological, i.e. involving the insertion of a mora. In most such languages the lengthening is quite noticeable and often affects the tone patterns as well, as in Shekgalagari [S311] (see (28) below).

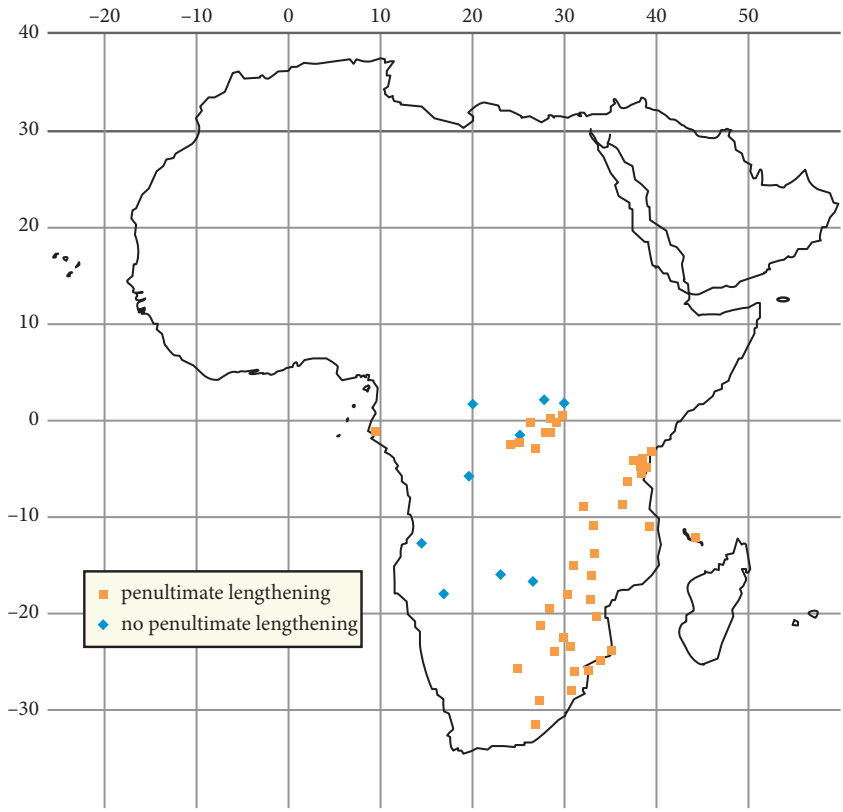


Figure 1.2 Languages which have introduced penultimate lengthening (Dark gray squares).

The idea is that PUL begins as an intonational property of utterances and only later "narrows" to smaller domains—exactly the opposite of vowel shortening.[12]

The first piece of evidence is that there are Bantu languages where PUL clearly is intonational. In Shekgalagari [S3111] PUL is a property of declarative utterances (Hyman & Monaka 2011), clearly visible in citation forms such as (28a) and sentences such as (28b).

(28)	a.	ri-nâːrí	'buffalos'
	b.	a-bal-a ri-nâːrɪ	'he is counting buffalos'
vs.	c.	ri-nárí	'buffalos?'
	d.	a-bal-a ri-nárí	'is he counting buffalos?'

[12] Although starting from opposite domains (stem vs. utterance), a reviewer points out that the two historical developments reveal a broader generalization: Both vowel shortening and PUL eventually wind up characterizing the word domain, consistent with Myers and Padgett's (2014) artificial language study showing that "learners are biased toward word-based distributional patterns" (p. 399).

In (28a, b) the underlying form /ri-nárí/ 'buffalos' undergoes both PUL and a tonal change from H-H to HL-L.[13] In (28c, d), however, we see that PUL does not apply to yes no-questions. There also is no penultimate lengthening in the environments in (29).

(29)

WH questions			
ri-nárí zhé ↓ríhí	'which buffalos?'	ányí a-bɔ́n-á ri-nárí	'who sees the buffalos?'
Imperatives			
bal-á	'count!'	bal-á ↓rí-nárí	'count the buffalos!'
Hortatives			
á ↓hí-bál-ɛ	'let's count!'	á ↓hí-bál-ɛ ri-nárí	'let's count the buffalos!'
Vocatives			
mʊnaká	'Monaka!'	ntó, Gabalʊxóŋ	'come here, Ghabalogong!'
Exclamatives			
á ↓ʃí-xóló	'what a situation!'	á ↓ʃí -ʧʊ́ʧʊ ʃá mʊ́-khyʊ	'what an idiot of a person!'
Monosyllabic words			
ri-nárí ʒé	'these buffalos'	a-bat-a ʃé	'he wants this one'
Ideophones	(with final devoicing)		
y-á-rɪ bílʊ̥	'it appeared suddenly out of water'	a-rɪ bítsɪ̥	'he left in a hurry'
Paused lists	(with final lengthening)		
a-bal-a ri-namaː	... rí-nawáː ... lí ri-nâːrɪ	'he's counting meats ...	beans ... and buffalos'

It is interesting to note that the declarative is pragmatically unmarked, but prosodically marked (by PUL) in Shekgalagari, while the environments in (29) are pragmatically marked, but—except for ideophones and paused lists—prosodically unmarked. It is also not necessary to raise the pitch in questions, which end without PUL and with the underlying tones of the final word unmodified. Yes-no questions can therefore be said to lack intonation (see Hyman & Monaka 2011).

While intonational PUL is rather restricted in Shekgalagari, other Bantu languages allow it in more, ultimately all clause types, as seen in the table in (30) (Hyman 2013: 314).

[13] The observed lengthening cannot be attributed to the tone change, since other patterns undergo PUL without creating a contour tone: /mʊ-lɪmi/ → *mʊ-lɪ:mi* 'farmer,' /ma-rumé/ → *ma-ru:mé* 'greetings,' /mʊ-nóna/ → *mʊ-nó:na* 'man' (Hyman & Monaka 2011: 271).

(30)

	Shekgalagari [S311]	Sesotho [S33]	Kinande [JD42]	Ikalanga [S16]	Ndebele [S44]	Chichewa [N31]
Declaratives	+	+	+	+	+	+
Yes–No Q	–	–	–	+	+	+
WH Q	–	–	–	+	+	+
Imperatives	–	+	+	+	+	+
Hortatives	–	+	+	+	+	+
Vocatives	–	±	+	+	+	+
Exclamatives	–	–	+	+	+	+
Monosyllables	–	+	+	+	+	+
Ideophones	–	–	–	–	+	+
Paused lists	–	+	+	–	+	+

With thanks to Malillo Machobane and Katherine Demuth (Sesotho), Joyce Mathangwane (Ikalanga), Ngessimo Mutaka (Kinande), Galen Sibanda (Ndebele), and Sam Mchombo and Al Mtenje (Chichewa).

While starting as a property of utterances and the intonational phrase, in some Bantu languages PUL has been narrowed to considerably smaller phonological phrases. A particularly striking case occurs in Simakonde [P23] (Manus 2018), from which non-contrastive PUL can be observed in the following citation verb infinitive and noun forms:

(31)

a.			b.		
	kú-lúúma	'to bite'		lí-ngéela	'mango'
	kú-lúmúúla	'to cut'		vi-loôngo	'pots'
	kú-lúmúláánga	'to cut into small pieces'		i-pooso	'present'
	kú-lúmúlángííla	'to cut into small pieces for s.o.'		sí-lóólo	'mirror'

As seen in (32), both a head noun and most modifiers undergo separate PUL and are hence analyzed as two phonological phrases (φ) (Manus 2003: 114):

(32)

noun + adjective	lí-ngéela	lí-kúmeêne	'big mango'
noun + numeral	vi-loôngo	vi-viíli	'two pots'
noun + genitive	lí-ngéela	lyá nkoôngwe	'the woman's mango'
noun + relative	vi-loôngo	vyá ŋgúsúmiile	'the pots that I bought'

Although the possessive pronouns generally phrase with the head noun, as in (33a), an alternative appositional version is also available in (33b), with separate

phrasing, where the pronoun has the same tone pattern it takes in an independent noun phrase (cf. *yáangu* 'mine'):

(33) a. i-posó yaángu 'my present'
b. i-pooso yáangu 'my present' (= 'present mine')

Demonstratives, on the other hand, obligatorily phrase with the head noun, which they also require to be all H tone:

(34) a. í-pósó aiilá 'that present' (cf. i-pooso, with all L tone)
b. ví-lóngó aviilá 'those pots' (cf. vì-loôngo, with L-LHL-L tone)

As seen in the table in (35), where 1φ and 2φ indicates one vs. two phonological phrases, Makonde dialects differ considerably in how they phrase noun modifiers (Rolle & Hyman 2019: 3):

(35)

Source	Dialect	POSS	DEM	ADJ	NUM
Leach (2010)	Plateau Shimakonde	1φ	1~2φ	2φ	2φ
Devos (2004)	Makwe	1φ	1~2φ	2φ	2φ
Manus (2003, 2018)	Zanzibar Simakonde	1~2φ	1φ	2φ	2φ
Kraal (2005)	Chinnima	1φ	1φ	2φ	2φ
Liphola (2001)	Coastal Shimakonde	1φ	1φ	1φ	2φ
Odden (1990a, b)	Chimaraba	1φ	1φ	1φ	1φ
Odden (1990c)	Chimahuta	1φ	1φ	1φ	1φ

All but the last two dialects studied by Odden show a contrast between modifiers which phrase with the head noun vs. those which don't. The generalizations from the above comparison are that possessive pronouns and demonstratives tend to form a single phonological phrase with the head noun, while adjectives and numerals tend to phrase separately, with numerals being the most prosodically independent noun modifier. The example in (36) shows that an NP can potentially consist of several phonological phrases, each undergoing PUL:

(36)

NOUN	ADJ	GEN	NUM	
(vi-loôngo)φ	(ví-kúmeêne)φ	(vy-á naáswe)φ	(vi-viíli)φ	'two big white pots'
CL8-pot	CL8-big	CL8-GEN white	CL8-two	

However, all of the 1φ/2φ dialects exhibit cases of "prosodic smothering" (Bennett, Harizanov, & Henderson 2018): A 1φ modifier that targets the head noun to form a phonological phrase "entraps" intervening 2φ modifiers (Rolle & Hyman 2019). An example again comes from Simakonde (Manus 2003, 2018), when a

demonstrative is added to the ADJ-GEN-NUM sequence in (36) where modifiers were seen to phrase separately:

(37)

NOUN	ADJ	GEN	NUM	DEM	
(ví-lóngó	ví-kúméné	vy-á náswé	ví-vílí	aviilá)φ	'those two big white pots'
CL8-pot	CL8-big	CL8-GEN white	CL8-two	cl8.DEM'	

Because of the requirement that a demonstrative phrase with the head noun, this overrides the separate phrasing property of the ADJ, GEN, and NUM which intervene, and PUL applies to the one φ. Note that the demonstrative, which tends to come last, also requires all of the preceding words in the NP to be all H tone, thereby confirming that a single phonological phrase has been formed.

Another case of prosodic smothering is found in Coastal Shimakonde (Liphola 2001), e.g. when a 1φ adjective follows a 2φ numeral:

(38)

a.	(NOUN	ADJ)φ			
	(má-pápájá	má-ngúlúguuma)		'round papayas'	(1φ)
b.	(NOUN)φ	(NUM)φ			
	(ma-papáaja)	(ma-taátu)		'three papayas'	(2φ)
c.	(NOUN	NUM	ADJ)φ		
	(má-pápájá	má-tátú	má-ngúlúguuma)	'three round papayas'	(1φ)

In this dialect, adjectives phrase with the head noun, as in (38a), while numerals phrase separately, as in (38b). However, when the adjective follows the numeral, as in (38c), a single phonological phrase is formed: the 1φ requirement of the adjective has overridden the 2φ requirement of the numeral.

Finally, concerning the verb phrase, PUL is closely integrated into expressing differences in information structure, in what is known as the conjoint-disjoint distinction in Bantu (van der Wal 2017), e.g. as in the present tense in Simakonde (Manus 2017: 246, 249):

(39)

a.	conjoint (1φ)	:	(á-tót-á	sí-júulu)φ	'she is sewing a HAT'
			CL1.SUBJ-sew-FV	CL7-hat.	
b.	disjoint (2φ)	:	(a-nku-tóót-a)φ	(sí-júulu)φ	'she is sewing a hat'
			CL1.SUBJ-PRES.DJ-sew-FV	CL7-hat	
c.	utterance-finally	:	(a-nku-tóót-a)φ		'she is sewing'
			(*a-toot-a)φ		

As seen in (39a) the conjoint form of the verb is used when focus is on the postverbal element, and both the verb and object are phrased together. In (39b), which has more "even" focus, the disjoint form of the verb is used and the two constituents are phrased separately. (39c) shows that the conjoint verb must have something

after it with which it phrases. It cannot occur at the end of a main clause utterance where the disjoint form of the verb must be used.

To summarize, there is considerable evidence that PUL was originally intonational occurring at the clause level but has a tendency to be narrowed to smaller phonological phrases. In this context it should be noted that there is no evidence for word-level penultimate prominence in early Bantu that could have been "widened" to phrasal prominence. Cases in Eastern Bantu where nouns and/or verbs show a bisyllabic minimality effect are clearly innovative, e.g. Kinande [JE42], where the imperative consists of the bare verb stem (*tum-à* 'send!') unless the stem is monosyllabic, in which case the second person singular subject prefix is required (*u-swa* 'grind!') (Mutaka & Hyman 1990: 112; Mutaka 2018: 174–175). In fact, I am aware of only one language that has extended PUL to the word level, namely, Komo [D23] (Paul Thomas, pers. comm.). It would be hard to explain the limitation of Shekgalagari PUL to declarative utterances as coming from word-level penultimate stress with other factors suppressing it in non-declaratives. Finally, there is other evidence, e.g. tonal, for phrase-penultimate prominence.

In this context recall the Giryama example in (1), where the H tone shifts to the penultimate mora of the phonological phrase, which also undergoes PUL. In Haya [JE22], which doesn't have PUL, tone changes apply at the end of an intonational phrase (IP), e.g. prepausally (Byarushengo, Hyman, & Tenenbaum 1976: 201–202; Hyman 1999: 155). As seen in (40a), an IP-penultimate H tone becomes a HL falling tone triggered by the L% phrasal boundary tone:

(40) a. e-m-búzi → e-m-bûzi 'goat' b. o-mu-tí → o-mú-ti 'tree'

H H L% H H L%

(cf. e-m-búzi yange 'my goat') (cf. o-mu-tí gwange 'my tree')

(40b) shows that the same L% causes an utterance-final H to shift to the penult. The forms in parentheses show that these processes do not occur phrase-internally. Finally note in (41) that the L% boundary tone is clearly related to information structure:

(41) a. base sentence: a-ba-kázi ni-ba-bal-íl-a ó-mw-ána é-m-bûzi]$_{IP}$

H H H HL%

'the women are counting the goats for the child'

b. nested IPs: ni-ba-zi-mu-bal-îl-a]$_{IP}$ á-ba-kâzi]$_{IP}$ ó-mw-âna]$_{IP}$

HL% HL% HL%

é-m-bûzi]$_{IP}$

HL%

'they are counting them for him, the women, the child, the goats'

In the base sentence in (41a) each of the four words has an input penultimate H tone, only the last of which receives the L% boundary tone. In (41b), on the other hand, where the three nouns 'women,' 'child,' and 'goats' are right-dislocated (with the pronominal marking *-ba-zi-mu-* referring to them, respectively), each constituent receives the L% boundary tone. Such tonal examples provide further evidence for the innovative nature of penultimate marking and indirect support for the intonational origin of PUL as per Hypothesis 3.

1.6 Discussion

To summarize, we have seen that although Proto-Bantu had a vowel length contrast on roots, many of the daughter languages have introduced significant changes. In some of the languages which keep the contrast, long vowels may be shortened in either word- or phrase-(ante-)penultimate position or on head nouns and verbs when followed by a constituent within their XP. In other languages which have lost the contrast, some have introduced phrase-level penultimate lengthening with different functions, e.g. marking declaratives in Shekgalagari, phrasing and focus in Makonde, word demarcation in Komo. As seen in Kimatuumbi and Simakonde, the head noun or verb is often prevented from phrasing separately, hence from having a long vowel. This can be viewed as an instantiation of a general linguistic process where the head noun or verb is a locus of prosodic unmarkedness, in these cases either losing length or not gaining penultimate length—just as heads are targeted for deaccenting and tonal mergers (cf. Selkirk 1984; Gussenhoven 2006; Harry & Hyman 2014; McPherson 2014; Rolle 2018, among others). Again, Haya [JE22] examples are instructive (Hyman & Byarushengo 1984: 57, 69):

(42) a. o-mú-ti 'tree' *vs.* o-mu-ti gwaa= Káto 'Kato's tree'
H Ø H

b. ba-jún-a 'they help' *vs.* ba-jun-a Káto 'they help Kato'
H Ø H

In (42a) the stem H of /-tí/ 'tree' is deleted before the possessive noun phrase gwaa= Káto 'of Kato,' while in (42b) the H of the final vowel /-á/ of the head verb 'they help' is lost when followed by an object noun phrase Káto.[14] In both cases the outer (dependent) trigger affects the inner (head) target, something which Rolle (2018: 5) terms "the outer dominance principle."

Which brings us back to tone. I began by referring to the different functions of tone: lexical, morphological, syntactic, pragmatic. Clearly length can match tone in these functions:

[14] These final Hs are realized on the penultimate syllable before pause.

(43) a. lexical: e.g. Lulamogi examples in (8)
b. morphological: e.g. Tiene in (44) below
c. syntactic: e.g. Haya in (45) below
d. pragmatic: e.g. Shekgalagari declarative intonation in (28)

The lexical and pragmatic functions of vowel length have already been seen. As an example of the morphological function, Tiene [B81] marks the applicative by lengthening the vowel of a root ending in a coronal consonant (Ellington 1977; Hyman 2010: 147), which we can refer to as "length ablaut":[15]

(44)

bót-a	'give birth'	→	bóot-ɛ	'give birth for'
bel-a	'speak'	→	beel-a	'speak to'
kas-a	'fight'	→	kaas-a	'fight on behalf of'
sɔ́n-ɔ	'write'	→	sɔ́ɔn-ɔ	'write for'
koŋ-a	'nibble'	→	kooŋ-ɛ	'nibble for'

As an example of a syntactically conditioned length fact, the long vowel of the Haya [JE22] today past tense marker /-áa-/ seen in (45a) shortens when the verb is non-final, as in (45b) (Hyman 1999: 160)[16]:

(45) a. y-áá-léet-a 'he brought'
b. y-a-leet-a Káto 'he brought Kato'

What's important is that shortening is not an automatic consequence of phrasing since it is limited to this tense and, as seen in (45b), the length of the root is not shortened.

It is thus clear that vowel length can have a variety of functions in Bantu, as it can in language in general. While vowel length can fulfill the four functions in (43), it still can't do everything tone can do—it would be thus be quite surprising if there were a parallel case to Giryama in (1) involving vowel length: If the subject prefix has an underlying long vowel, the length shifts to the penultimate syllable of the phonological phrase. In fact, as was said, nothing but tone can do this. Another thing length cannot do is harmonize. While such features as front-back, round, height, ATR can participate in vowel harmony, a process of length harmony such as in (46) is unattested:

(46) a. /lim-il-e/ → lim-il-e
b. /liim-il-e/ → liim-iil-ee

If we assume that features "assimilate by spreading" (Hayes 1986), vowels cannot assimilate in length because length is not a feature—there is no [+long] that could assimilate, rather a long vowel has two moras (vs. a short vowel which has one).

[15] Roots ending in a non-coronal take an infix with /l/, e.g. yók 'hear' → *yólek-ɛ* 'listen to.'
[16] The H tones are also reduced, as per the process seen earlier in (42).

I know of only one case of apparent length agreement in Leggbó [Cross-River; Nigeria], which affects verb roots (Hyman & Udoh 2007: 79).

(47)

		+ -ɛ̄ 'him, her'	+ -ɔ́ 'you sg.'
ff ìn-à	'touch'	ff ììn-ɛ̄ɛ̄	ff ììn-ɔ̀ɔ́
tùm-à	'stop'	tùùm-ɛ̄ɛ̄	tùùm-ɔ̀ɔ́
mān-ā	'hold'	māān-ɛ̄ɛ̄	māān-ɔ̀ɔ́

The verbs in the left column consist of a CVC root with a lexicalized /-a/ suffix. As seen, when one of the two object pronouns having the shape -V is added, fusion takes place: /a+ɛ/ → ɛɛ, /a+ɔ/ → ɔɔ. When (and only when) this happens, the vowel of the root is also lengthened. The question is whether this should be seen as a process of length harmony. Unfortunately the language conspires against testing whether the process is iterative (unbounded), since it does not provide appropriate inputs. While Hyman & Udoh (2007) consider several different analyses of the above root vowel lengthening, the interpretation I'd like to give to it is similar to the interpretation given of the Ngangela data in (24): Vowel length cannot occur in a less prominent position (here, suffixal) without it also occurring in a more prominent position (root). Otherwise, the root vowel lengthens. Whereas Ngangela shortens a vowel in weak position to avoid such a conflict, Leggbó lengthens a vowel in strong position. We thus once again note the versatility of vowel length which can mark positional prominence, clause types, syntactic headedness, and linguistic and paralinguistic intonation by itself. In the Bantu case, what is especially striking is the shift from the paradigmatic function of tone distinguishing lexical morphemes to various types of syntagmatic functions. Whereas segmental features cannot as readily do everything that length can do, the functions of length line up more along the lines of what tone and stress can do. While each has its own properties and limitations, length is probably second only to tone.

Acknowledgment

Earlier versions of this chapter were presented at the University of California, Berkeley (February 11, 2019), the Philological Society Meeting at SOAS, University of London (February 15, 2019), the 14th Annual Joshua and Verona Whatmough Lecture, Harvard (April 29, 2019), and the Laboratoire de Phonétique et Phonologie, Paris 3 (May 24, 2019). I am grateful to those audiences for their helpful questions and comments.

2

The rise and fall of rounding harmony in Turkic

Darya Kavitskaya and Adam McCollum

2.1 Introduction

It has been proposed that vowel harmony in general arises through the phonologization of vowel-to-vowel coarticulation (e.g. Ohala 1994; Hyman 2002; Przezdziecki 2005; Barnes 2006). In a similar manner Johanson (1979a) argues that the evolution of rounding harmony in Turkish is attributable to the reduced phonetic quality, [ə], of [+high] suffixes. Given these claims, the null hypothesis is that the loss of phonological harmony would affect the domain of harmony as a whole, resulting in phonetic vowel-to-vowel coarticulation. More generally, this predicts three kinds of languages relevant for vowel harmony, (1) those with coarticulation, which sows the seeds for harmony, (2) those with harmony, and (3) those with coarticulation as the residue of lost vowel harmony. In this chapter, drawing on nineteenth-century texts and our own fieldwork, we argue that both the emergence and decay of rounding harmony in numerous Turkic languages crucially involves stages between these endpoints.

In fact, if harmony is lost via a one-step change from iterative harmony to phonetic coarticulation we can collapse (1) and (3) above, predicting that there should only be two types of languages—those with harmony and those without. However, this claim is immediately falsified by the dialects of Crimean Tatar, which exhibit three different stages of the decline of rounding harmony. In the southern dialect, rounding harmony iterates throughout the word. In the central/standard dialect, rounding harmony affects a single syllable after a triggering round vowel. In the northern dialect, harmony is absent, and rounded vowels are licensed only in the first syllable, with the occasional loss of rounding even there (Samoilovich 1916; Sevortjan 1966; Kavitskaya 2010). These dialects suggest the need for intermediate stages between fully functioning harmony and coarticulation.

In tandem with the differences in the harmonic domain seen in Crimean Tatar, the literature on Turkic rounding harmony has repeatedly noted the lexically specific nature of harmony in the family (Johanson 1978–1979b; Anderson 1996;

Darya Kavitskaya and Adam McCollum, *The rise and fall of rounding harmony in Turkic*. In: *The Life Cycle of Language*. Edited by: Darya Kavitskaya and Alan C. L. Yu, Oxford University Press.
DOI: 10.1093/oso/9780192845818.003.0002