Historical Linguistics and the Comparative Study of African Languages

Historical Linguistics and the Comparative Study of African Languages

Gerrit J. Dimmendaal
University of Cologne

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Dimmendaal, Gerrit Jan.

Historical linguistics and the comparative study of African languages / Gerrit J. Dimmendaal.

p. cm.

Includes bibliographical references and index.

1. African languages--Grammar, Comparative. 2. Historical linguistics. I. Title.

PL8008.D56 2011

496--dc22 2011002759

ISBN 978 90 272 1178 1 (Hb; alk. paper)

ISBN 978 90 272 1179 8 (Pb; alk. paper)

ISBN 978 90 272 8722 9 (Eb)

John Benjamins Publishing Company • P.O. Box 36224 • 1020 ME Amsterdam • The Netherlands

John Benjamins North America • P.O. Box 27519 • Philadelphia PA 19118-0519 • USA

Table of contents

Preface

Although the comparative method has been applied to a range of African language families, these have received little coverage in general textbooks. The present monograph was written firstly in order to make the literature on African language families available to a wider public, and to give these languages the proper place they deserve in the comparative study of human language. Unfortunately, many students in African countries do not have access to more recent developments in historical-comparative linguistics or to the literature published on language families in their home countries, due to the sometimes deplorable financial situations at the institutions where they are studying. Furthermore, students interested in the historical comparison of African languages and the comparative method often have to use textbooks focusing on other parts of the world. Although this is useful in itself, these African languages also have their own story to tell, because of specific genetic and areal properties. This constitutes the second raison d'être for the present monograph. If the historical-comparative study of African languages were not to reveal certain novelties, one might still question the usefulness of another textbook on the comparative method.

An attempt has been made here to present a critical assessment of some current views in historical linguistics and to formulate a kind of research agenda which spells out the specific needs, as seen by the present author, concerning the future investigation of (African) languages. Readers no doubt at times will disagree with the position taken by the present author on certain matters. The interpretation of so-called "mixed languages", "creoles" or aspects of current grammaticalisation theories is no doubt controversial and, to some extent, runs against the (apparent) communis opinio. By explicitly stating my own position and by forwarding arguments for and against alternative positions, the reader should be able to take a stance especially on such controversial matters. Hopefully, these explicit statements in the current volume will also help to stimulate (rather than to kill) the academic debate.

Some readers may observe (or complain) that there is a particularly strong focus on Eastern and Central Africa in this textbook. This is due above all to the primary research experience of the author. Although an attempt has been made to include as many interesting phenomena from other parts of Africa as possible, it is not feasible to give detailed information about each and every language family on the African continent in a study of this size.

The present monograph not only grew out of a need for the production of a textbook on the comparative method that takes Africa as an incredibly rich and diverse area seriously, but also out of a psychological need to prove to myself that I am still capable of being creative in some sense. While being immersed in heavy teaching loads, numerous administrative duties and the supervision of theses, as well as providing a service to the academic community as a reviewer of articles, research proposals, or monographs produced by colleagues in the field, one sometimes wonders whether this is all there is to an academic position.

I would like to thank the Faculty of Arts at the University of Cologne for allowing me to spend part of an inspiring sabbatical between January 2006 and April 2006 at La Trobe University, Melbourne, where a first draft of the present monograph was completed. The support from the Institute for Advanced Study and the Research Centre for Linguistic Typology, La Trobe University, Melbourne, during my stay as a distinguished visiting scholar is also gratefully acknowledged here. R. M. W. (Bob) Dixon and Alexandra (Sasha) Aikhenvald (now at James Cook University, Cairns) deserve special thanks for being such wonderful hosts and for their relentless support over the years. This meant a lot to me, especially at times when I started questioning my academic capabilities. Moreover, the numerous other colleagues I interacted with during my stay in Australia should be thanked here – in particular Keith Allan, Kate Burridge, the late Michael Clyne, David Fleck, Randy LaPolla, Frank Lichtenberk, Simon Overall, Tonya Stebbins, Sheena van der Mark, David Watters, Liejiong Xu, Ghil'ad Zuckermann, Siew-Peng Condon, Gilah C. Leder and Julia Anderson. Thanks also to other colleagues in Australia – in particular Nick Evans, for his inspiring academic contributions and for being a mate.

Although I finished a first draft during my stay in Melbourne in 2006, I was not satisfied with the result. Of course, most of us never are, but one has to decide at which point one has reached an acceptable academic level. It took some more time to bring the manuscript to an academic level that was satisfactory to me, but also to reduce the monograph to a publishable size. The initial plan was to write an introduction to comparative linguistics with a special focus on Africa. Due to the fact that the original manuscript had to be reduced considerably, the end result was a much more dense presentation of information on different topics, which also requires a more extensive basic knowledge of linguistics. As a result, the text is probably no longer suitable as an introduction for undergraduates. Instead, it has become a textbook for more advanced students of linguistics and colleagues working on language families outside of Africa with an interest in the current state of research on African linguistics as seen by the present author.

I would like to express my gratitude to my mentor Paul Newman, who taught me comparative linguistics when I was an undergraduate student at Leiden University, in the Netherlands. It was his inspiring teaching which triggered my intellectual curiosity in the subject in the first place.

Monika Feinen produced the various maps for the present volume. Unfortunately, many of the beautiful maps she created could not be included for reasons of space. I am particularly grateful to Steffen Lorenz, who spent numerous hours solving formatting problems and correcting references. Silke Focke, Meikal Mumin, and Nico Nassenstein helped me in various other respects, in order to get the book published. To all of them I express my deepest gratitude.

I would also like to thank the various (former) colleagues at the African Studies Institute at the University of Cologne for helping to create an academic atmosphere where research is still important.

Last but not least, I would like to express my deeply felt gratitude to Bob Blust, Tucker Childs, Sigrid Newman, Bonny Sands, Gertrud Schneider-Blum, Anne Storch, and Elly van Gelderen for reading the manuscript and for providing numerous insightful and critical comments, and to Kees Vaes, Martine van Marsbergen, and Jan-Kees van Oord (John Benjamins Publishing Company) for their relentless support. It is only because of their encouragement and patience that I managed to bring this project to an end.

Cologne, 2010

Figures

Maps

Tables

Part I

The comparative method

This monograph is about the historical and comparative study of African languages. Moreover, it deals with the so-called "comparative method", which was developed initially with respect to the historical study of Indo-European languages, but which has since established itself as a basis for the comparative study of other language families as well. Although as a method it has probably become more sophisticated over the past 150 years or so, specific classical assumptions are still held to be valid by many scholars today. These basic principles of the comparative method will be discussed in Part I. They involve the establishment of lexical and grammatical cognates, the reconstruction of their historical development, techniques for the subclassification of related languages, and the use of language-internal evidence, more specifically the application of so-called "internal reconstruction".

Whereas Part I is intended as an elementary introduction to the comparative method, based on data which is primarily, though not uniquely, from African language families, Part II addresses language contact phenomena. Part III then deals with the relationship between comparative linguistics and other disciplines.

Although the present monograph is intended as an advanced course book in historical linguistics, it was decided not to add questions at the end of each chapter, as is sometimes done. After having started formulating such questions, it was concluded by the present author that readers (both teachers and students) should decide for themselves which issues are central or peripheral to them. Also, no exercises were added, mainly for reasons of space. It is hoped that the examples presented in different chapters present a sufficient basis as an introduction. With respect to these examples it should be noted (for example, in the typological survey of common sound changes in Chapter 2), that it is not possible in a monograph of the present size to repeat the evidence various authors have given for their reconstructions. Readers are referred to the original sources quoted in these various sections in order to be able to check for themselves whether they find the actual evidence convincing. Additional exercises and comparative data may be found in Cowan (1971), Anttila (1989) or Crowley (1997). Furthermore, extensive data sheets on different African language families that could be used as exercises can be found in, for example, Boyeldieu (2000), Capo (1991), Cloarec-Heiss (1978), Hedinger (1987), Jakobi (To appear), Moñino (1995), Nanfah (2003), or Sambiéni (2005).

Chapter 1

Explaining similarities

Historical linguistics as a concept has come to be used in a number of ways. First, in referring to a discipline investigating the development of single languages from an earlier to a later point in time, e.g. two or more stages of Ancient Egyptian, the language of Pharaonic times, whose written records date back four and a half millennia, and which probably became extinct as a spoken language (known as Coptic) several centuries ago.

A second way in which the term historical linguistics has come to be used is as a short hand for historical-comparative linguistics – more specifically the comparative study of genetically related languages, hence the alternative term genetic linguistics. This is the way in which the term is used in the present study. In historical-comparative linguistics, or simply comparative linguistics, one reconstructs "upstream", i.e. one works backwards from today's languages in order to establish genetic relationships and in order to reconstruct earlier stages by studying collateral relationships. Even though as a discipline historical (or diachronic) linguistics grew out of philology, one does not necessarily depend on written documents or texts in order to be able to reconstruct historical changes in languages, as should become clear from the following chapters. What is needed to begin with are solid analyses of individual languages.

This chapter sets out to list a number of reasons why languages may manifest similarities in their lexical and grammatical structures. After a discussion of chance, sound symbolism, borrowing and shared inheritance, basic principles of the comparative method are explained. As a first step, these involve setting up sound correspondences between cognate roots in languages assumed to be genetically related, and reconstructing historical changes in these forms.

1.1 The introduction of historical-comparative linguistics in an African context

The 17th century saw the beginning of systematic synchronic accounts of African languages. In 1635, the Italian Franciscan priest Arcangelo Carradori, who was based in Girgeh, Upper Egypt, produced an Italian-Nubian lexicon containing more than 7000 words from Kenuzi Nubian. As early as 1659, the first grammar of a Bantu language, Kongo, was produced by Brusciotto. In subsequent years, the first grammars of other African languages appeared, e.g. a grammar of the Ethiopian Semitic language Ge'ez by Ludolf (1661), and a grammar of Amharic by the same author (Ludolf 1698). However, it was only towards the end of the 18th century that in Europe the fashion arose of gathering wordlists in languages for comparative purposes. For example, the glossary in Latin of vocabularies in two hundred Asian and European languages, *Glossarium Comparativum linguarum Totius Orbis* (1787–1789), compiled by Peter Pallas, dates from this period. The project was sponsored by the Empress Catherine the Great of Russia, and appeared in revised editions in Russian between 1790 and 1791. The glossary also contained information from 30 African languages. A collection of vocabularies of some 24 languages of West Africa was compiled by Oldendorp in the West Indies as early as 1777 (Hair 1963: 13).

Von Schlözer is often credited with having recognised, and named, the Semitic family in 1781. However, as observed by Ruhlen (1987: 77), the affinity of Hebrew, Arabic, and Aramaic had been recognised for centuries by Jewish and Islamic scholars, and this knowledge was published in Western Europe as early as 1538 (Postel 1538). Ludolf, who had written grammars of Ge'ez and Amharic, recognised the extension of the Semitic family into East Africa. A number of other language families as we know them today were established during the 18th century, e.g. Finno-Ugric (or Finno-Ugaritic), which was recognised as a linguistic entity by Gyărmathi (1799).

The so-called **comparative method** is often assumed to have been initiated by Sir William Jones, a British judge who was based in Calcutta in the British colony India. In a famous lecture he pointed towards the systematic relation between Sanskrit and the better known languages Latin and Greek in terms of their lexicon and grammar; his hypothesis on the common historical origin of these languages was published in *Asiatic Researches* (1788).

The 19th century saw the commencement of more extensive comparative studies of African languages. Balbi was probably the first scholar to aim at an overall classification of African languages, connecting the so-called Bushman languages with Nama. *Mithridates, oder allgemeine Sprachenkunde* by Adelung and Vater, which appeared in four volumes between 1806 and 1817, contained the first hints at possible linguistic comparability and the relationship between several African languages.

With respect to the comparative study of African languages south of the Sahara, the German missionary Sigismund Wilhelm Koelle is probably to be credited for being one of the founders of this academic discipline through the publication of his *Polyglotta Africana* (1854). Koelle came to the (then) British colony of Sierra Leone in 1847 as an agent of the Church Missionary Society. He was next sent to the headquarters of this missionary organisation in Freetown, where he stayed between 1850 and 1852. Koelle left Freetown in 1853, and moved to London, where he completed a 1200 page manuscript for the printer. As observed by Hair (1963: 13), Koelle may have taken the idea for his title from a study *Asia Polyglotta*, which was published in 1823 by Klaproth. Hair (1963: 13) further notes that a number of other authors published short wordlists of West African languages during this era, e.g. Kilham (1828), Appleyard (1847), and Clarke (1848). However, none of these studies could compare in depth, size and thoroughness to the 156 languages covered by Koelle, who collected an average of 300 words for each of these. Koelle had the opportunity to work with speakers from so many different African languages which was due mainly to the fact that the neighbouring country of Liberia had become the (sometimes temporary) new home for liberated slaves dumped off by the British in the area. The remainder of the speakers interviewed by Koelle (who also collected their life histories) were mainly traders or seamen.

On the basis of a systematic comparison of the lexicon, the grammar and simple sentences of these languages, Koelle was able to find intersecting patterns of phonological similarity, and arrived at the conclusion that several of the languages investigated were genetically related. The twelve units postulated by him are repeated here for convenience (whereby the twelfth group in actual fact consists of languages for which no obvious genetic links could be established at the time):

I. North-West Atlantic Languages ("distinguishing themselves, like those of South Africa, by Prefixal Changes or an Initial Inflection")
II. North-Western High-Sudan Family or Mandén'ga Family of languages
III. Upper Guinea Languages or Middle-Coast Languages
IV. North-Eastern High Sudan Languages
V. Niger-Delta Languages
VI. Niger-Dshadda Languages, or Núpē Group
VII. Central African Languages
VIII. Atam Languages
IX. Mókō Languages
X. Kōn'gó N'góla languages
XI. South-Eastern Languages
XII. Unclassified and Isolated Languages

A sample of languages from Koelle's *Polyglotta Africana* classified as members of Group IV, North-Eastern High Sudan languages, is presented in Table 1.

Table 1. Data on languages from Koelle's Group IV

English	'one'	'two'	'three'	'head'	'hair'	'face'	'forehead'
Mőse	*yímre & yímbo*	*yĭ́wo*	*tá′wo*	*zúru (mamzúru, fozúru)*	*kőawdo*	*nḗn′ga*	*dḗ′ri′*
Dsélana	*n′ínera*	*báli*	*Báta*	*zoh*, pl. *zī*	*zĭ́uh* pl. *Zīr*	*núnur*	*sébōadar*
Gúrēsa	*wúnyi & n′kála*	*béia & nlḗa*	*báta & ńta*	*zu & zúma*	*su* pl. *sű́sa*	*nímbīe*	*díre*
Gúrma	*le*	*lē & nlḗ*	*tā & ntā́*	*yű́li*	*tí-yūdi & kő́vi-ti*	*kánun′ga*	*yūtúni*
Lḗgba	*kúdēm*	*nā́bela*	*púmbōárāra & ḿbōárāra*	*nyóro*	*nyős*	*ḗsa*	*tógū*
Káure	*kúdum*	*nā́le*	*nādéso*	*Nyóro*	*nyős & nyősi*	*ésa*	*tĭ́dīe & tĭ́de*
Kīámba	*kúdom & kúlom*	*néwale*	*néodóso & nóodóso*	*kúdyō (má gudyō, nyá gudyő)*	*nyőzi (mányōz, nyányōz)*	*ésāndá*	*nyĭ́ere*
Kóāma	*ndĭ́an′*	*lḗa*	*tḗre*	*nyún′*	*nyípose*	*sĭ́a*	*túle*
Bágbālan′	*dĭ́an′*	*lḗa*	*tőre*	*ńyo′ ményo′, ńyo′*	*nyúpon*, pl. *nyúpōze*	*sĭ́n′a*	*tű́la*, pl. *tű́lezi*
Kásm	*kálo*	*ńle*	*ńta*	*yíru*	*ĭ́ye*	*ĭ́ya*	*tíle*
Yűla	*kálo*	*ńle*	*ńto & ńtōa*	*yúru (ámu yuru, ḿmu yuru)*	*yű́a*, pl. *yű́e*	*yĭ́a*	*túli*, pl. *túlēa*

So what were Koelle's intuitions based on? When comparing words between two randomly selected languages, one may always find a similarity in a number of forms. The classic case sometimes mentioned among Africanists (probably going back to Gregersen 1977: 84) is that between Hausa and German.

Hausa	**German**	
leebèè	*Lippe*	'lip'
hannuu	*Hand*	'hand'
karfii	*Kraft*	'strength'

Most of the time, so-called similarities are in the same category as the presumed common words between Hausa and German. Given the finite number of sounds we find cross-linguistically, one will often come across a few such similarities between the six thousand plus languages still spoken today if one looks at larger data sets. This is due

to **chance**.[1] Once one goes beyond this brief list and starts comparing other words, e.g. for body parts, numerals or verbs like 'eat', 'drink', 'go', one does not find any similarity in form and meaning between Hausa and German:

Hausa	**German**	
jìkii	*Körper*	'body'
shaà	*trinken*	'drink'
biyu	*zwei*	'two'

Apart from chance, usually involving a small number of lexical items, there is a second reason why arbitrarily chosen languages may show some similarity in specific words without any historical reason, namely **sound symbolism.** Onomatopoetic words belong to this category of words whose form/meaning relationship is motivated, i.e. is not arbitrary. These may be names for animals producing specific sounds, or verbs describing the sounds particular animals make. Swahili *kuku* for 'fowl, chicken', or the verb form *kokola* for the sounds this animal may produce (*cackle* in English), belong to this category. Also, words like *papa* or *mama* in English and their parallels in Swahili *mama* or *baba* belong to this category. A third reason why two or more languages may manifest similarities is due to **borrowing**. Borrowing and **shared inheritance** involve historical processes to be accounted for by way of the comparative method. An important methodological issue consequently involves distinguishing borrowing from genetic inheritance.

Linguistic departments around the world occasionally receive letters (or e-mails these days) from laymen who claim to have found some interesting similarities between language X, for example in South America, and language Y in a different part of the world. It is usually enough to point out the parallel situation between Hausa and Germanic, i.e. to refer to chance resemblances. However, if one finds large numbers of forms which are similar in form and meaning between two or more languages, and if these words are part of so-called basic vocabulary, other explanations are required.[2] Borrowing would be one possible explanation, but this usually occurs in "culture words" and less commonly in fundamental vocabulary. Due to the fact that concepts like eating or drinking are basic, the general assumption is that most languages will have a word for these, and so there is no immediate reason for borrowing the corresponding words from another language – although as we shall see below in Chapter 8, there are sometimes problems with this assumption.

---

1. See Ringe (1992) on the factor of chance in language comparison.

2. Note that in the examples from Hausa and German above the orthography of these languages has been used. In modern comparative linguistics we are not comparing letters, but sound units (phonemes) and sounds.

For many words in languages, the relation between form and meaning is arbitrary (non-iconic). Thus, when learning the word for 'head' in Hausa we find *kâi*, whereas in Swahili it is referred to by way of a rather different form, *kichwa*. When two or more languages show similarities in form and meaning in a range of basic words, an explanation other than borrowing is needed. Koelle noticed that in various basic words in African languages which were compared by him, a strong similarity in form and meaning could be observed. The list of words compared by him included numerals, body part terminology, and attributive concepts like 'good' or 'bad'. Grammatical features, in particular involving bound morphemes which show similarities in form and meaning, are even better indicators of **genetic relationship** or **genetic affinity** because they are not easily borrowed, as we know today. Interestingly, Koelle (1854) also included grammatical comparisons in his comparative survey, e.g. verbal conjugations in different languages such as 'I go', 'I come', 'I lie down' etc. or phrases like 'I drink water', 'I eat rice', 'I do not play'. This we may interpret as an initial attempt to discern grammatical (next to lexical) similarities between languages.

Research in the decades following Koelle's seminal publication in 1854 confirmed his conclusions that the lexical and grammatical similarities observed between different groups of languages were due neither to chance nor to borrowing. Languages classified as Group IV are now considered to be part of the Gur language family (also referred to as Voltaique by French-speaking Africanists); Köhler (1964) gives a detailed account of the Gur languages identified by Koelle (1854). Moreover, as other scholars discovered during the 19th century, not only do we find similarities between languages that are genetically related, i.e. belong to the same language family, we can also establish systematic correspondences between these languages.

From the brief description above we may summarise the most important methodological considerations as follows. Arbitrariness of sound-meaning connections in words, except in nursery words and words involving sound symbolism, is a fundamental property of linguistic signs. If there are considerable numbers of resemblances between two or more languages in spite of this arbitrariness of sound-meaning relations, a historical interpretation is necessary: Either borrowing or a common genetic origin is involved. Resemblances in fundamental vocabulary and grammatical resemblances in form and meaning are of particular significance in this respect. Apart from the number of resemblances, there is another quantitative aspect involved. Not all sound-meaning similarities are equally important. The longer formerly similar words are, the less likely it is that chance is involved. Moreover, if we find similar suppletive forms in two or more languages, the chances are high that these are a reflex of a genetic relationship (Greenberg 1957). For example, languages classified as Germanic within Indo-European have a suppletive (and therefore unpredictable) alternation for one adjectival pair expressing comparison (superlative): *good/better* in English, *goed/beter* in Dutch, *gut/besser* in German, or *god/bättre* in Swedish.

1.2 Sound correspondences and reconstructions

When comparing vocabularies as well as grammatical features between several languages, one may be able to detect a historical relationship, as Koelle did in his pioneering work. However, there is more to this exercise than identifying "look alikes". If we take the Gur language family as a basis, or more specifically one group of languages within this family showing a particularly close affinity with each other in their lexical and grammatical structure, it turns out we are able to set up *systematic* correspondences between these languages. Sambiéni (2005) undertakes this task for a group of Gur languages called the Oti-Volta-Oriental cluster. As a first step, the author presents an inventory of the vowel and consonant phonemes of these languages. In Table 2, adapted from Sambiéni (2005: 33), shaded blocks indicate differences in consonant inventories between these languages.

Once a synchronic inventory of structural sound units (phonemes) and their phonetic realisation in individual languages has been made, the next step would be to set up a systematic comparison of basic vocabulary, in order to arrive at an inventory of all **sound correspondences** occurring between the set of languages.

Table 2. Consonant inventories in Oti-Volta Oriental (Gur)

	Biali	Ditammari	Naténi	Waama
p	+	+	+	+
b	+	+	+	+
t	+	+	+	+
d	+	+	+	+
c	+	+	+	+
k	+	+	+	+
kp	–	–	+	+
m	+	+	+	+
n	+	+	+	+
ɲ	+	–	–	+
ŋ	+	+	–	–
ŋm	–	–	–	+
f	+	+	+	+
s	+	+	+	+
g(ɣ)	+	–	–	–
h	+	–	+	–
l	+	–	–	–
r	+	–	–	–
w	+	+	+	+
y	+	+	+	+

These items with identical or similar shape and meaning, identified in particular through a comparison of basic vocabulary, are called **cognates.** They are conveniently ordered in such a way that correspondences for segments (or suprasegmental units for that matter) become obvious. Compare, for example, the consonants in stem-initial position in Table 3.

Table 3. Cognate sets in Oti-Volta Oriental (Gur)

	Biali	**Ditammari**	**Naténi**	**Waama**
year	*b*yēn-ī	di-*b*ē-nī	*b*ēñ-ní	*b*īn-dē
attach	–	*b*óó-ø-	*b*ō-ø-á	*b*óm-bóó-mā
arm	*b*ā-yɔ̄-hū	kū-*b*ā-kù	*b*āà-kū	*b*ā-kú
goat	*b*wōm-wā-yɔ̄	tā-*b*ūānīī-tā	*b*ɔ̄-nīī-tá	sáá-*b*ɔ́tɔ̄-fā
ask	*b*ēē-sé	*b*é-kā	*b*é-kā	*b*í-bí-mā
child	*b*íí-gɔ̄	dā-*b*ī-rā	*b*íī-tá	*b*í-kā
grain	*b*íī-m	*b*īī	*b*īī-	*b*ī-mā
try	*b*ē-í	*b*ē-nā	*b*ɛ̄ń	–
friend	*d*wó-p-ū	*d*á-p-ō	*d*óó-b-ō	*d*áá-s-ō
quiver	*d*wá-y-ī	*d*ī-dôò	*d*ō-kē-rí	*d*ō-kú
sing	–	dīèn-tà	*d*ēèñ-ø-	*d*é-rí-mā
get used to	*m*ān-s-á	*m*ān-tā-rā	*m*ān-dá	–
build	*m*ē-í	*m*áá	*m*ā-lí	*m*áá-rí-mā
excrement	*m*ín-dū	tī-*m*ḯn-tì	*m*íṅ-dī	*m*í-nā
iron	*m*á-tū	kū-*m*ā-rì-kù	*m*áà-kū	–
grass	*m*wó-hṹ	kū-*m*ū-rí-kù	*m*í-kū	*m*ɔ́-kú
fight	–	*m*ā	*m*ɛ̀ā	*m*áá-tí-mā
thin	*m*yè-sī	–	*m*é-hī	*m*ɛ̄ɛ̄-sí-dā
glanders	*m*yá-tū	tī-*m*ī̋ɛ̄-tì	*m*é-tī	*m*ɛ̄-nā
swallow	*n*íím	*n*ī-tóō	*n*í-tá	–
cattle, ox	*n*ā-fɔ̄	fā-*n*àà-fà	*n*à-fā	*n*ák-ō
mouth	*n*ū-ī	dī-*n*ùù	*n*ù-rī	*n*ɔ́-rē
eye	*n*wān-fɔ̄	fā-*n*ùɛ̀n-fa	*n*òṅ-fā	*n*ún-fā
bird	*n*ē-gē	tā-*n*ùɛ̀-tà	*n*ɛ̀-tā	*n*ɔ́nɔ́ɔ́-fā

It is useful from a methodological point of view to order lexemes for comparison according to the syntactic category to which they belong. It could be the case that a language uses only suffixes for nouns, whereas in verbs prefixes and suffixes are allowed. This is a common pattern, for example, in Cushitic languages. The alternating positions

for root-initial consonants for nouns, as opposed to verbs, is sometimes relevant for the conditioning of specific sound shifts. In the case of the Gur languages above, this apparently was not relevant, and accordingly lexemes were not ordered on the basis of categorical distinctions by Sambiéni (2005).

As the list above further shows, it is important to understand the morphological structure of the languages one intends to compare in order to be able to identify the common lexical items, or roots, for systematic comparison. Not only do languages change their sound structures over time, their morphological (and syntactic) systems may also be subject to restructuring. Consequently, one first tries to identify the lexical root, i.e. the smallest meaningful unit stripped of morphological elements like affixes and clitics. These morphological elements can also, and should also, be compared in a systematic matter. However, this is done best once phonological changes in lexical roots are reasonably well understood.

Let us examine some of the preparatory "handwork" in order to see what kind of work is needed for the establishment of regular sound correspondences, as in the examples from Gur languages above. For an effective comparative project it is important to understand the phonological structure as well as the phonetic realisation of the phonemes and the morphological structure of individual languages. In practice, one has to be prepared for a number of potential problems. For example, not all relevant data may be available for a specific language, or a particular lexeme may not be cognate because it replaced the original form. This is what the empty blocks marked with a hyphen in Table 3 represent. On the basis of cognate roots presented in Table 3, recurring regular sound correspondences can be set up; the corresponding starred form marks the reconstructed proto-form (based on the discussion in Sambiéni 2005: 184–190).

Table 4. Correspondence sets in Oti-Volta-Oriental (Gur)

	Biali	**Ditammari**	**Naténi**	**Waama**
**b-*	*b-*	*b-*	*b-*	*b-*
**d-*	*d-*	*d-*	*d-*	*d-*
**m-*	*m-*	*m-*	*m-*	*m-*
**n-*	*n-*	*n-*	*n-*	*n-*

The easiest hypothesis for the first correspondence set, involving voiced bilabial stops, is of course to assume that no changes occurred in the initial consonant position for these cognates. One may call this the economy principle. From a methodological point of view it is always good to start, if possible, with correspondence sets where no variation occurs, i.e. where the sound units compared are identical. The correspondence sets for the bilabial and alveolar nasals also suggest that these remained unchanged.

Once lexemes have been ordered according to different correspondence sets for segments, as in the examples above, it may turn out that one has a partially overlapping correspondence set, i.e. a set sharing one phoneme in language A, but two in a related language B. From the inventory in Table 2, we can already see that these four Gur languages have slightly different consonant inventories. We observe, for example, that Biali and Waama also have a palatal nasal, whereas Ditammari and Naténi do not. Given the fact that all four languages also have a palatal approximant (written as *y*), the palatal nasal is either an innovation in Biali and Waama, or it was lost in Ditammari and Naténi. A comparison of lexical cognates should provide an answer to this puzzle (data derived from Sambiéni 2005: 201–202).

Table 5. Correspondence sets for palatal consonants in Oti-Volta Oriental (Gur)

glosses	Biali	Ditammari	Naténi	Waama
drink	*ɲãã*	*yã*	*yãa*	*yãã-ma*
know	*ɲá*	*yã*	*yã-la*	*yĩm-ma*
head	*ywo-i*	*di-yuu-ø*	*yuu-ri*	*yan-ta*
hair	*ywo-hu*	*ku-yuri-ku*	*yu-ku*	*yɔ-si*

The principle "majority rules" can sometimes be applied to cases where only one of the languages deviates. However, as some changes are very common, languages may undergo these independently of one another. As the cognates for 'head' and 'hair' above show, there are cognate roots where we find a palatal approximant (or glide) in root-initial position in all four languages. At the same time, we find a second correspondence set, as in the entries for 'drink', and 'know', where Biali has a palatal nasal and the other three languages have a palatal approximant. We are thus left with two potential explanations: The system as found in Biali either reflects a more archaic stage, i.e. *ɲ* constitutes a **retention**, or it is due to an **innovation**. It turns out that the presence of the palatal nasal in Biali can be predicted historically (though not synchronically) from the presence of nasality on a following vowel (or sometimes a nasal consonant following this vowel, a situation essentially retained in the other three languages). It follows that the nasal consonant *ɲ* in Biali developed through anticipation by the palatal approximant of nasalisation on the following vowel: **yṽ* > *ỹṽ* > *ɲv* (Sambiéni 2005: 178).[3] The symbol > expresses "historically changed into", whereas the symbol < expresses "historically derives from". As we shall see later (in particular in Chapter 13),

3. Sambiéni (2005: 201) points out that the correspondences sets for these consonants are irregular in Waama (though not in the examples above), the second language with a palatal nasal, which is why this language is not further discussed here. The question why and how such irregularities may come about in languages is addressed in Chapter 2.4.

this kind of change involving alternations between nasalised vowels and neighbouring oral consonants is very common in West African languages. In other words, *ɲ* was innovated in Biali through a **split,** whereby **y* shifted to *ɲ* when nasality occurred on a following segment, whereas it remained unchanged elsewhere. The alternative process whereby two sound units are no longer phonologically distinctive is called **merger**.

The basic steps in the application of the comparative method may now be summarised as follows:

- Select the most probable cognates (lexical roots) in order to be able to set up sound correspondences.
- Set up correspondence sets, also for those cases where there is no apparent variation between languages.
- Start out from the most straightforward cases with identical sound units in cognates. These can be assumed to represent (unchanged) **reflexes** of the original sound unit.
- Specify the position within the root (initial, medial, final) for the consonants and vowels being compared, as this may turn out to be relevant for the conditioning of specific sound changes.
- Order these correspondences according to phonetic and phonological similarity.
- Check whether distinct reflexes occur in identical environments, or, alternatively, whether they are in complementary distribution with some other sound unit.
- Reconstruct phonemes for different positions within the root.

These sequences represent the logical steps in the reconstruction of the **proto-language**, the common ancestor from which a group of related languages emerged, and should not necessarily be applied mechanically in just this order. In practice, one often moves between these stages. Also, a number of heuristic principles play a role. As we shall see in the next chapter, specific sound changes are very common ("natural"). This experience often helps us to develop a first intuition about the direction of change. However, the most important point is that the arguments for specific reconstructions are pursued logically.

There are potential pitfalls when working out sound correspondences and innovations. For example, one may be missing conditioning factors for specific sound shifts, thereby multiplying the number of reconstructed contrasts. Another problem encountered in comparative work concerns gaps in the material, either because the forms are not clearly cognate, or because they could be cognate but do not seem to fit in with well-attested, regular sound correspondences. These irregularities or exceptions to established sound changes, however, can be approached in a systematic way, as shown in Chapter 2.

The observation that sound correspondences between genetically related languages are regular was an important methodological advance initially derived from the study of Indo-European languages, and later confirmed through the study of other language families, regardless of whether or not these languages were written. This regularity allows one to reconstruct historical stages. The comparative method was developed by the so-called Neogrammarians (*Junggrammatiker* in German, also sometimes referred to as the methods of the Leipziger School, after the name of the university town in Germany, Leipzig, where various scholars were based).[4] Their methods are central to the first seven chapters of the present monograph.

The most important innovation in the application of the method in the 20th century concerned the distinction between phonetics and phonology. Before embarking upon this important distinction (in Chapter 2), let us have a closer look at another group of languages assumed to constitute a language family, namely Bantu. This will help us acquire some additional insights into the comparative method, and also help us understand the relevance of the distinction between phonetics and phonology for the application of the method. In the following examples we will also introduce some additional, complicating factors, involving more dramatic sound changes, morphophonological complications, (minor) semantic changes, and a first identification of borrowings.

1.3 Additional methodological considerations

When Wilhelm Bleek (1827–1875), the German-born missionary and linguist, started his activities in Natal, South Africa, he also became aware of systematic similarities between languages spoken in the area. For these languages, he coined the term "Bantu", after the common word for 'people, human beings' in these languages.[5] Bantu languages typically use noun-class prefixes (rather than suffixes, as in Gur), and Bleek (1862, 1969) concentrated on Bantu languages in Southern Africa. However, research on other languages since has shown that the Bantu family stretches out over a vast area between Cameroon in the West and Kenya in the East, and the area south of this zone. Table 6 presents some lexical cognates among a number of Bantu languages. Data for Bobangi are based on Whitehead (1899), for Kinyarwanda on Coupez et al. (2005), for Luba-Kasai

---

4. The influential 20th century linguist Leonard Bloomfield, who was a student in Leipzig at one point, introduced these concepts into American structuralism. He applied the comparative method to Central Algonquian languages in North America; see Chapters 18–27 of his classic, *Language* (Bloomfield 1933).

5. Note that earlier investigators had already been alluding to a common ancestry for some languages now considered to be part of Bantu; see Nurse and Philippson (2003a) for a historical survey of Bantu studies.

on Van Avermaet (1954), for Nyamwezi on Maganga and Schadeberg (1992), for Swahili on Johnson (1971), and for Zulu on Doke, Malcolm, Sikakana, and Vilakazi (1990).

Because of the high number of cognates in basic vocabulary, a relatedness hypothesis (with an ancestral language as the common proto-language) as well as a regularity hypothesis (with cognate sets and correspondence sets established on the basis of reflexes in the daughter languages) is the most plausible historical explanation for the structural similarities between these languages.

Table 6. Cognate sets in Bantu

		Bobangi	Kinyarwanda	Luba-Kasai	Nyamwezi	Swahili	Zulu
1	count	*-bal-*	*-bar-*	*-bal-*	*-βal-*	*(-hesabu)*	*-bal-*
2	shine	*-bāl-*	*(-angar-)*	*-bal-*	*-βal-*	*-wa-*	*-bal-el-*
3	singe	*-bab-ọl-*	*-bab-*	*-bab-*	–	*(-unguz-)*	*(-hamul-)*
4	courtyard	*-banza*	*-baandza*	*-banza*	–	*-wanja*	*-bandla*
5	be(come)	–	*-b-*	*(-di)*	*-β-*	*-wa*	*-b-*
6	break	*-bun-*	*-vun-*	*-vung-* 'fold'	*-βɪɪn-zik-*	*-vun-jik*	*(-phul-)*
7	rain	*-bula*	*-vura*	*-vul-*	*(-phul-)*	*-vua*	*-vula*
8	ask	–	*(-saba)*	*(-eel-)*	*-βʊʊj-*	*-uliz-*	*-buz-*
9	mould	*(-y*m)*	*-buūmb-*	*-bumb-*	*-βʊʊmb-*	*-umb-*	*-bumb-*
10	kill	–	*(-iic-)*	*(-ship-)*	*-βʊlag-*	*-u-*	*-bulal-*
11	excrement	–	*-vyi*	*(-fi)*	*(-fii)*	*-vi*	*-vi*
12	thigh	–	*-biero*	*-belo*	*(-taango)*	*-weo*	–
13	stone	–	*-buye*	*-bwe*	*-we*	*-we*	*-tʃʼe*
14	bad	*-bẹ*	*-bi*	*-bi*	*-βii*	*(-baya)*	*-bi*
15	become rotten	*-bɔl-*	*-bor-*	*-bol-*	*-βol-*	*-oz-*	*-bol-*
16	two	*-bālë*	*biri-*	*-bidi*	*-βɪlɪ*	*-wili*	*-bili*
17	arm	*-bokō*	*-boko*	*-boko*	*-βoko* 'hand'	*(-kono)*	*(-khono)*
18	boil up	–	*-bir-*	*-bil-*	–	*(-chemka)*	*-bɪl-*
19	body	*(-zöt*)*	*-biri*	*-bidi*	*-βɪlɪ*	*-wili*	*(-zimba)*
20	see	*-bon-*	*-bon-*	*-mon-*	*-βon-*	*-on-*	*-bon-*

When comparing the root for 'count' above, it is immediately clear that the Swahili form does not fit in. First, unlike most Swahili verbs it does not end in *-a*, which is already indicative of it being a borrowing, more specifically a borrowing from Arabic; also, its tri-consonantal root structure is reminiscent of its Semitic origin. Because the lexical roots in the related languages are cognate, and because the sound correspondences involved in these forms are regular, they are still included in the correspondence sets in Table 7. Sometimes it is not immediately clear whether specific

forms in two or more languages are cognate, as with the root for 'shoulder' in Bobangi (Table 8). Once a more extensive corpus has been investigated, however, it usually becomes clear on the basis of recurring sound correspondences whether we are dealing with a cognate root or not. The quality of the data of course is also a crucial issue from a methodological point of view. There are times, however, where words in a specific language simply are not available to the comparativist, or where the data are not up to modern standards (as with the Bobangi source above).

On the basis of the cognate forms above, we can begin to set up regular sound correspondences between these languages, whereby Swahili turns out to be the language with the most dramatic historical sound changes. Let us therefore concentrate on this language below. The set of cognates in Table 6 allows us to set up the correspondences for root-initial consonants summarised in Table 7.

Given the fact that at least four languages have a bilabial stop in initial position in the various cognate roots shown in Table 7, the original sound unit was probably a stop as well. There is also internal evidence from morphophonemic alternations in languages like Swahili that the original consonant was a stop. In this language, **b* was retained after a nasal (as shown by the examples in Table 8), whereas it was modified in other positions.

Table 7. Correspondence sets between Bantu languages for bilabials

	Reconstructed form	Bobangi	Kinyarwanda	Luba-Kasai	Nyamwezi	Swahili	Zulu
1, 2, 3, 4, 5, 12, 13, 15, 16, 17, 19	**b-*	*b-*	*b-*	*b-*	*β-*	*w-*	*b-*
6, 7, 11	**b-*	*b-*	*v-*	*v-*	*β-*	*v-*	*v-*
8, 9, 10, 15, 20	**b-*	*b-*	*b-*	*b-* (*-m-* if a nasal follows)	*β-*	*ø-*	*b-*

Table 8. Bilabial stops after nasals in Swahili

		Bobangi	Kinyarwanda	Luba-Kasai	Nyamwezi	Swahili	Zulu
1	two		*-**b**iri*	*-**b**idi*	*-**β**ɪɪlɪ*	*m-**b**ili*	*-**b**ili*
2	dog	*-**b**wa*	*-**b**wa*	*-**b**wa*	*b**w**aa, -**w**aa*	*m-**b**wa*	*-**dʒ**a*
3	goat	*-**b**ọli*	*(-hene)*	*-**b**uʒi*	*-**b**ɔli, -**β**ɔli*	*m-**b**uzi*	*-**b**uzi*
4	shoulder	*-bēkī (?)*	*-**b**ega*	*-**b**eya*	*-**β**ega*	*m-**b**ega*	*(-hlombe)*
5	penis	-	*-**b**oro*	*-**b**olo*	*-**β**olo*	*m-**b**oo*	*-**b**ol*
6	mosquito	-	*-**b**u*	*-**b**u*	*-**b**ʊʊ*	*m-**b**u*	*(-miyane)*
7	unripe	*-**b**ẹsu*	*-**b**isi*	*-**b**iʃi*	*-**β**ɪsɪ*	*m-**b**ichi*	*(-ngavu-thiwe)*

Table 9. Correspondence sets between Swahili and Nyamwezi for vowels

Swahili		Nyamwezi	
	i ~ i(i)		
-zima		*-zimá*	'extinguish'
-vimba		*-βiimba*	'thatch'
-vimba		*-βiḯmbá*	'swell'
-linda		*-liinda*	'guard'
	i ~ ɪ(ɪ)		
-lia		*-lɪla*	'cry'
-imba		*-ɪmbá*	'sing'
-pima		*-pɪɪma*	'measure'
	e ~ e(e)		
-cheka		*-seka*	'laugh'
-lea		*-lela*	'bring up'
-leta		*-leéta*	'bring'
	a ~ a(a)		
-anza		*-andya*	'begin'
-washa		*-βachá*	'burn, light'
-fa		*-faá*	'die'
	o ~ o(o)		
-ona		*-βoná*	'see'
-poa		*-polá*	'become cool'
-ota		*-looóta*	'dream'
	u ~ ʊ(ʊ)		
-tuma		*-tʊmá*	'send'
-umba		*-βʊʊ́mbá*	'mould, create'
-kua		*-kʊlá*	'grow'
	u ~ u(u)		
-uma		*-lumá*	'bite'
-nuka		*-nuuɲha*	'stink'
-futa		*-futa*	'wipe, rub'

Before the central vowel *a* and the front vowel *e* (as in 2, 4, 12), **b* changed to *w* (**b* > *w*). Synchronically, *b* and *w* cannot be treated as variants (allophones) of one and the same phoneme in Swahili, since they occur in identical environments, as in the word *saba* 'seven' (a borrowing from Arabic) versus *sawa* 'equal, the same'. In other words, a split occurred whereby **b* shifted to *w* and merged with another existing sound unit (phoneme) before the central vowel *a* and the front vowel *e*.

Before a back vowel *o* **b* shifted to *ɸ* (i.e. was lost), as shown in Examples 15 and 20 (Table 6). The cognate forms in 8, 9 and 10 suggest that this rule extended to another back vowel, *u*. However, the cognate roots in 6 and 7 contradicts this hypothesis, at least at first sight. These forms suggest that a shift **b* > *v* occurred before a high back vowel *u*. A closer look at cognate forms in, for example, Nyamwezi shows that in this language we only find a vowel *u* in the cognate forms 6 and 7, whereas in the cognate roots in 8, 9 and 10 Nyamwezi has a vowel *ʊ*. As additional cognates show (illustrated in Table 9 above), Nyamwezi has retained a vowel distinction between *u* and *ʊ* which was lost in Swahili.

The seven vowels of Nyamwezi thus correspond to five vowels in Swahili; moreover, whereas Nyamwezi has a distinctive vowel length, Swahili does not. (Data on Swahili is based on Johnson 1971, and that for Nyamwezi is based on Maganga and Schadeberg 1992.) There are two (alternative) hypotheses that could explain the observed structural differences between the five-vowel and the seven-vowel system in these two Bantu languages. One possibility is that a merger occurred in Swahili, as a result of which two open high vowel units (**ɪ* and **ʊ*) shifted to and merged with **i* and **u*, respectively. Consequently, the situation in Nyamwezi would represent the older situation. If a split had occurred in Nyamwezi, e.g. whereby **e* split into *e* and *ɪ* and **o* split into *o* and *ʊ*, one should be able to specify an environment (e.g. from neighbouring segments). In that case, the Swahili system would represent the older situation. How do we decide between the two logical options, an original five-vowel or a seven-vowel system? As a general principle, if a split occurs, we should be able to find a conditioning factor. The alternative (and first option), that of a merger, does not require such a conditioning factor. The third option, namely that neither of the two represents the original situation should only be invoked once it is clear that the first two cannot account for the existing situation. Applying this economy principle follows from an old scientific principle, known as Occam's Razor, after the medieval philosopher Occam who stated that entities in an argument are not to be multiplied beyond necessity (*entia non sunt multiplicanda praeter necessitate).*

Since it is not possible to find a conditioning factor for a so-called split in Nyamwezi, the alternative, namely a merger of vowel distinctions in Swahili, is the simplest alternative, hence:

**i*			**u*		
	>	*i*		>	*u*
**ɪ*			**ʊ*		

Whereas vowel length is not distinctive before nasals, it is contrastive elsewhere in Nyamwezi. Historically, no conditioning can be found for vowel lengthening in these other positions in Nyamwezi; the distinction between short and long vowels again must be a retention, whereas Swahili lost the contrast. The conclusions arrived at,

namely that the seven-vowel system as well as vowel length in Nyamwezi reflect the older or more archaic stage, is confirmed by evidence from other Bantu languages; the issue of vowel length in Bantu is further discussed in Meeussen (1979). The structural aspect of the historical change is reflected by the fact that both the front and the back high vowel shifted, and also that vowel length was lost for all vowels in Swahili.[6]

We are now, finally, in a position where we are beginning to understand the changes which root-initial **b* underwent historically in Swahili. Apparently, **b* shifted to *v* when followed by a high close vowel (**i* or **u*) in Swahili, as in 'break' and 'rain' (Table 6) or 'thatch' (Table 9) above, but not when followed by an open high vowel (**ɪ* or **ʊ*), as in 'ask' (8), 'mould' (9) or 'kill' (Table 6). Although it is not immediately obvious from the root for 'excrement' (11 in Table 6) that the initial consonant was also a voiced bilabial stop historically, a systematic comparison of additional roots with a high close front vowel in these Bantu languages shows that the Swahili reflex is regular, as shown, for example, by the cognate roots for 'swell' and 'thatch' in Table 9 above. Swahili is a language marking stress on the penultimate vowel in a word, whereas Nyamwezi is a tonal language with two registers, high (marked with ´) and low (unmarked in the examples above). The cognate forms in Nyamwezi and Swahili strongly suggest that this tonal contrast goes back to their common ancestor, and that this system was replaced by a stress system in Swahili. This preliminary conclusion, which is confirmed by comparative evidence from numerous other Bantu languages, is based on the fact that no conditioning factor can be found for its emergence historically in Nyamwezi, whereas the situation in Swahili can be explained as a case of neutralisation of such a contrast historically; tonal phenomena are discussed in more detail in Chapter 2.

Due to the merger of the (front and back) open and close high vowels, both *b* and *v* now occur before *i* and *u*. Consequently, *b* and *v* are distinct phonemes in Swahili. This is a phenomenon we will observe again and again below; what starts as allophonic variation for specific consonants or vowels may become unpredictable (i.e. may phonologise) as a result of subsequent changes in neighbouring segments.

Sound change usually starts with the development of allophonic variation in segments in specific environments. For example, in Nyamwezi intervocalic *β* alternates with *b* after a nasal; the Nyamwezi root for 'rain' occurs after a nasal prefix (*m-bulá*), and thus is part of a regular correspondence set. This also applies to Swahili, where the root *-wili* after a vowel alternates with *-bili* after a nasal. A shift towards a fricative in prevocalic or intervocalic position, as in Nyamwezi, is a common sound change cross-linguistically. In Kinyarwanda, the root-initial consonant in the examples above

6. Meinhof (1899, 1906) is one of the pioneer studies of comparative Bantu, next to Guthrie (1967–1971) and Meeussen (1980); the latter was published posthumously, but had been circulating among Bantuists for several years before it appeared.

is also realised as a voiced bilabial fricative. Because it is in complementary distribution with a voiced stop elsewhere (e.g. after a nasal), Coupez et al. (2005) treat the bilabial fricative in Kinyarwanda as a variant (allophone) of the latter, and, consequently, use the symbol *b* in order to represent this structural sound unit. In Zulu, the bilabial stop is slightly implosive intervocalically; after a nasal, however, it is realised as a plosive. In other words, implosion is a redundant rather than a contrastive phonetic detail of the stop. Nevertheless, such information can be important since sound changes usually start with allophonic variation in individual sounds.

It is fundamentally important to have information on the realisation of phonemes, more specifically their allophonic distribution, also for another reason: These may reflect earlier phonetic properties of the segments involved or, alternatively, they may reflect phonological contrasts which were lost. Implosion is the phonetic norm for the bilabial stop in geographically distant Bantu languages like Zulu (South Africa) or Tunen (Cameroon). Consequently, they may already have been slightly implosive in Proto-Bantu, but there is no evidence that Proto-Bantu had a phonological distinction between plosive and implosive stops.

For obvious methodological reasons we start our comparison with forms that have identical or near-identical meanings. If we were to initially allow ourselves too much freedom in terms of acceptable semantic variation when comparing forms, we could easily get lost in a plethora of forms and corresponding meanings. Also, semantic change is a poorly studied domain of language change, as we shall see in Chapter 5, and so our knowledge about what constitutes a plausible semantic shift is still limited. Once correspondence sets have been established, however, additional material may be brought in. Thus, the Swahili word for 'breast', *maziwa* (also 'milk'), does not appear to be cognate. There is, however, a form which fits in quite well also in terms of its semantics, namely *mbele* 'in front of'.

As these above examples show, reconstructing historical scenarios can be intricate. The recurrence of a particular correspondence set is of crucial importance in such endeavours. How many witnesses are required to ensure that a presumed correspondence set is regular? There is no straightforward answer to this, other than "the more the better". The great Indo-Europeanist Antoine Meillet (1937: 340) formulated "the three witness" requirement. A correspondence set which is attested in three or four lexical roots belonging to a basic lexicon obviously is less well-established than one which occurs fifteen times. Also, the length of words plays a role since the chances of finding similar non-cognate forms between two languages in words of three of four syllables is smaller than in monosyllabic forms.

The fact that in Swahili the two vowels **ɪ* and **ʊ* shifted to and merged with the high vowels **i* and **u* suggests that the former vowels were high, rather than mid vowels, originally. This is also the assumption commonly held by Bantuists, who have developed a special set of symbols to express this:

Guthrie (1967–1971)		**Meeussen (1967)**		**Schadeberg (2003)**	
Close	open	close	open		
ɨ	*ʉ*	*ɨ*	*ʉ*	*i*	*u*
u	*u*	*i*	*u*	*ɪ*	*ʊ*

Formal similarity between lexical roots in related languages is sometimes misleading. The crucial point is to establish recurrent, and thereby regular, correspondences, not just similarities between forms. An elegant illustration of this comparative pitfall is formed in comparative data sets for Chadic. Newman and Newman Ma (1966) and Newman (1977) applied the comparative method to this language family, and established a range of regular sound correspondences as well as innovations. From these, it also became clear that a number of lexical forms which look similar in specific Chadic languages in fact are not cognate. Thus, Hausa *wutaa* and Ga'anda *waata* 'fire' are formally similar, but they are not cognate. For similar reasons, because the forms do not fit in with regular sound correspondences between these languages, French *feu* and German *Feuer* for 'fire' are not cognate. The latter is cognate with forms in other Germanic languages like Dutch (*vuur*), or more distantly related Indo-European languages like Greek *pur* and Armenian *hur*. The form in French has cognates in Spanish (*fuego*) or Italian (*fuoco*), i.e. it goes back to a different lexeme with a word-medial velar consonant. These latter languages thus have words for 'fire' which are cognate with *feu* in French.

According to Newman and Newman Ma (1966), the following roots in two Chadic languages are cognate because they can be related to each other by way of regular sound correspondences.

Hausa *soya* 'roast' **Kanakuru** *wuri* 'roast'

The crucial point here is that there is a regular correspondence between Hausa *s* and Kanakuru *w*. Similarly, the fact that Margi *psar* and Tera *wuzən* 'grass' are cognate (i.e. go back to a common ancestral form) may not be immediately obvious. "However, if one looks at these languages plus a dozen other Chadic languages such as Bole, Gude, Kotoko and Ngizim, and, for contrast, at words from a few non-Chadic languages such as Kanuri and Fulani, intersecting patterns of lexical similarity emerge such that the relationship between Tera and Margi becomes evident" (Newman 2000a: 263). The forms are historically related (cognate) because they can be derived through regular sound shifts from a common proto-form. Sound units do not change overnight, and so what we observe between Hausa and Kanakuru, or between Margi and Tera is the endpoint of a sequence of changes. Language changes normally occur through a succession of restricted steps, namely subtle sound changes. The end result, however, may be correspondence sets between rather distinct sound units whose historical link only becomes obvious once other related languages are taken into account.

The standard classification of African languages as proposed by Greenberg (1963) is based, not on the comparative method, but on mass comparison, a method further discussed in Chapters 3 and 14. According to Greenberg, the languages spoken on the African continent can be grouped into four phyla. Considerable work using the historical-comparative method, however, has been carried out on smaller units all of which were assumed by Greenberg (1963) to be part of his four hypothesised language phyla. These latter groups, rather than the larger phyla proposed by Greenberg, play a central role in the following chapters. For ease of reference, these are listed in Appendix I, together with a range of linguistic isolates. We will begin by addressing the question of why some sound changes are more common than others.

Chapter 2

Explaining sound change

When specific languages are thought to be genetically related, because of systematic and recurrent form-meaning similarities in a large number of lexical roots in particular involving basic vocabulary as well as grammatical morphemes, it is usually possible to set up regular sound correspondences between segments in cognate forms. Moreover, as illustrated in the preceding chapter, specific sound changes may be reconstructed as a next step. Experience has shown not only that similar principles of regular sound change apply to language after language, but also that similar types of sound changes occur again and again in different parts of the world. Some of these more widespread structural changes in sound systems are illustrated next.

2.1 Some common types of segmental changes

In the comparative data from Oti-Volta (Gur) languages above, we saw that a root-initial approximant became a nasal whenever a nasalised vowel followed in the same root. Experienced comparative linguists would immediately state that such a change, whereby a consonant changes its manner of articulation before another segment is a "natural sound change". The characterisation "natural" is somewhat of a tautology, because if such a change occurs in human languages, it is "natural" by definition; what

one is actually saying is that nature tends towards the natural. However, some sound changes are more common than others, and it is this experience presumably which has led scholars to talk about "naturalness" for some sound changes. An unnatural sound change, then, would involve a rare type, requiring a language-specific explanation. Let us have a look at the more common sound shifts first.

In the comparison of cognate roots for a group of Bantu languages in Chapter 1, it was concluded that the original root-initial consonant in roots like 'count', 'shine', 'two' and 'rib' was probably a bilabial stop, rather than a fricative. Although caution is in order, as a sound change needs to be established ("proven") through the comparative method, experienced researchers in comparative linguistics probably would immediately have gone for this option rather than any other, e.g. assuming a shift from a sonorant or fricative towards a stop. This is due to a common experience that the **weakening** of consonants between vowels, also known as **lenition**, is well attested cross-linguistically as a "natural" sound change. Westermann (1927: 181–85), for example, already pointed towards such processes for Mande:

**kun*	'head'	>	*kun*	Mandingo
		>	*xuni*	Susu
		>	*wu(n)*	Kpelle

The following instances of lenition involve a change in manner of articulation from total occlusion to friction, because the two articulators became separated; this, in turn may be followed by a further distancing between the two articulators (e.g. the lips, or the tongue and the alveolar ridge etc.). For the voiceless obstruents (i.e. plosives plus fricatives) voicing may occur as an intermediate step. Alternatively, or in combination with this voicing stage, stops may become affricates first (*p* > *pf, t* > *ts* etc.).

plosive	>	**fricative**	>	**approximant**
p		*ɸ / f*		*h*
b		*β / v*		*w*
t		*θ / s*		*h*
d		*ð / z*		
c		*ʃ /s*		
ɟ		*ʒ*		*j*
k		*x*		*h*
g		*ɣ*		*j*

Alternatively, a voiced alveolar stop may become a lateral (*l*), for example, a conversion known as **lambdacism**, or it may develop into *r*. Stops like *k*, *q* or *g* may also develop into a glottal stop *ʔ* (sometimes called **debuccalisation**, i.e. removal of activity from the mouth). Voiceless approximants are relatively rare or marked cross-linguistically, and

so again this historical sound change involving manner of articulation with voicing is very common in languages.[7] Approximants like *h*, *w* and *j* often represent the final stage in a change from a fricative (or stop) towards the loss of a segment.

Intervocalic voicing of voiceless stops and subsequent weakening of such consonants appears to be a permanent rule, i.e. a rule which re-occurs again and again at different points in time, e.g. in language groups such as the Eastern Sudanic branch within Nilo-Saharan. In the following examples from the Surmic branch within Eastern Sudanic, Baale constitutes the innovating language (Yigezu 2001: 245).

Murle	*kɛbɛrɛ*	'eye'
Didinga	*kɛbɛrɪ*	'eye'
Narim	*kɛbɛrɛ*	'eye'
Baale	*kɛɛrɛ*	'eye'

Whereas lenition processes are widely attested cross-linguistically, the inverse process nevertheless appears to be attested as well. Thus, in Lebanese Arabic, the original Arabic dental fricatives **θ* and **ð* shifted to and merged with *t* and *d* respectively. The latter process is presumably favoured by a dental (as against an alveolar) point of articulation for such stops. This inverse process, involving consonantal **strengthening**, is also known as **fortition**.

Sasse (1974/1975, 1979) reconstructs two distinct phonemes **d* and **z* for Proto-Eastern Cushitic. Whereas Proto-Eastern Cushitic (PEC) **d* remained unaltered in all branches of modern Eastern Cushitic, **z* was either changed into *d* or *z* in the modern languages: "PEC **d* and **z* have merged into a single *d* […] in most Saho dialects and Afar, Northern Somali, Baiso, Macro-Oromo […]", according to Sasse (1979: 15).

In Bantu languages, the **fricativisation** (or **spirantisation**, as it is often called in Bantu studies) is often conditioned by the close high vowels *i* and *u* (commonly written as **i̯* and **u̯* in comparative Bantu studies). These vowels also triggered a shift in articulation point for the preceding consonant in Bantu languages like Swahili, as we saw in Chapter 1. The articulatory motivation for this appears to be a shift in tongue position whereby the new point of articulation approaches that of the vowel. Compare the following historical sound shifts in a number of Bantu languages (based on Bostoen 2005: 202; morpheme boundaries added), e.g. involving a shift **pi > fi > si*.

7. Defining the notion of markedness in general terms is not so easy, but a number of criteria have been applied, such as the frequency of occurrence in a particular language or cross-linguistically; alternatively, or in addition, the relative age at which a particular sound is acquired by children may be used as a criterion. Languages like Ewe have both a bilabial and a labiodental fricative, but the latter is the most common one of the two cross-linguistically.

**-pígò*			'kidney'	(Proto-Bantu)
		m-pigo	'kidney'	(Kimbu)
	>	*figo*	'kidney'	(Swahili)
	>	*en-sigo*	'kidney'	(Luganda)
**-piga*			'firestone'	(Proto-Bantu)
	>	*ama-piga*	'firestone'	(Kimbu)
	>	*figa*	'firestone'	(Swahili)
	>	*es-siga*	'firestone'	(Luganda)

Similar processes have been reconstructed for related language families, e.g. Jukunoid, by Storch (1999).

If stops tend to be weakened intervocalically or before specific vowels, where do stops come from? One common source for single stops in languages appears to be from double stops that became simplified, e.g. **bb* > *b*, or **kk* > *k* etc. Examples of this may be found in the historical phonology of Cushitic languages; for example, in the following **degemination** processes reconstructed for West-Rift Southern Cushitic by Kießling (2002) in a classical contribution using the comparative method:

Pre-West Rift to Proto-West Rift

**follit*	>	**folit*	'dig'
**xayya*	>	**xaya*	'he comes'
**hhaddoo*	>	**hhadoo*	'sticks'
**sliffoo*	>	**slifoo*	'lips'

Kießling (2002) as well as Kießling and Mous (2003) present detailed historical accounts of the phonological development of Southern Cushitic languages in general, thereby illustrating a wide range of historical sound changes.

Cluster **reduction** may also involve consonants with different points or manner of articulation which become simplified. Thus, simplifications of labial-velar (or labio-velar) stops, such as *gb* > *b* or *kp* > *p,* are rather common in West African languages. Kastenholz (1996: 152) refers to such processes in Western Mande languages. In Looma, for example, the reflex of Proto-Southwest-Mande **kp* is *ɓ*, whereas in Mende the original sound unit has become voiced:

Looma	**Mende**	
-dɔ́ɓɔ́	*-dɔ̀gbɔ́*	'wilderness'
-duɓa	*-dìgbá*	'press'

Furthermore, where do double or geminated stops come from? They in turn may result from a kind of **assimilation** process involving adjacent consonants with different points or manner of articulation adapting or assimilating to each other. These assimilation processes may be partial or complete in nature. Such modifications can