7

Comparative African Linguistics

Gerrit J. Dimmendaal

7.1 Introduction

The historical-comparative study of African languages involves at least two dimensions: first, the identification of genetic relationships between languages; and second, the application of the comparative method to languages assumed to be genetically related, in order to establish historical changes between them. Section 7.2 addresses the successes and drawbacks of this continent-wide endeavour to date, while Section 7.3 discusses some of the results emerging from in-depth studies of well-established language families. Internal reconstruction and dialectology complement the historical-comparative method, as discussed in Sections 7.4 and 7.5, respectively. Distinguishing between borrowing and shared inheritance from a common ancestor is not always easy, as shown in Section 7.6, where different contact scenarios between genetically related as well as unrelated languages are discussed within an African context. While, traditionally, linguists have represented genetic relationships in terms of family trees, more recently network analyses have come into vogue. Whereas the former cladistic approaches reflect tree-branching or divergence, the latter rhizotic models visualize convergence between genetically related languages, as shown in Section 7.7. This chapter concludes with some observations (Section 7.8) on possible future directions for the field.

7.2 The Historical Comparison of African Languages

The gradual identification of different languages outside Europe, mainly a result of European colonialism during the 18th and 19th centuries, led to the publication of collections of vocabularies for African languages. For example, Simon Pallas's *Linguarum totius Orbis Vocabularia Comparativa*, a study published in two volumes in 1787 and 1789 and containing word

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

lists for Amharic, Fulani, Igbo, Khoekhoe, or Mandinka (and other African languages), was made possible with financial support from Catherine the Great, empress of Russia, who initiated this comparative endeavour (Fodor 1975:7). During her reign, word lists were distributed via Russian ambassadors, and subsequently also reached Dutch, British, and Spanish colonies.

Of course, European investigators were not the first scholars in the field. Scientific knowledge had declined in the Middle Ages in Europe, until the 12th century, when scholars in Southern Europe recovered ancient Greek science through successive translations from Arabic and Hebrew, often with commentaries from scholars speaking these latter languages. These contributions to mathematics, medicine, or astronomy are well known, but scholars from the Middle East were also among the earliest investigators in comparative linguistics. The awareness of a relationship between languages, resumed today under the label 'Semitic' on the one hand and Berber on the other, probably dates back to as early as the 10th century, when a physician by the name of Judah ben Quraysh, who lived in Morocco, had apparently already identified structural similarities between them (Becker 1984, quoted in Frajzyngier & Shay 2012a:4).

The French scholar Postel, who was familiar with Arabic and Hebrew texts, published a study in 1538 on the origins of the Hebrew language, thereby de facto identifying Semitic, although this label was not used as a common denominator for the family until the end of the 18th century. Today, we know that Semitic is part of a larger macro-family or phylum, called Afroasiatic, which includes not only Berber but also Chadic, Cushitic, Egyptian, and Omotic. As shown in the elegant survey by Hayward (2000), formal similarities, particularly involving case marking, conjugational features of the verb, personal pronouns, and plural markers point towards the common origin of these groups as Afroasiatic languages. Grammatical features tend to be more conservative (and less subject to borrowing) than lexical features, and so it should not come as a surprise that there are few obvious lexical cognates for Afroasiatic as a whole, given the time depth involved for this phylum; see also Chapter 10.

Similarities between two or more languages are particularly significant when they can be observed in basic vocabulary, involving words for body parts or verbs such as 'drink', 'eat', or 'walk'. Such lexemes are less likely to be borrowed, although borrowing is possible, as discussed in Section 7.6. By taking on a 'wide-angle' perspective, that is, by comparing a large number of languages at once rather than proceeding pairwise in a step-by-step fashion, it is usually easier to identify lexical (and grammatical) similarities, as the following example with the word for 'goat' in Swahili, a language of wider communication in East Africa, and Tima, a distantly related language spoken in the Nuba Mountains in Sudan, should help to illustrate.

(1)	Swahili	Tima	
	mbuzi	címìì	'goat'

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Multilateral (or mass) comparison of vocabularies between a range of languages helps to identify similarities that are not obvious when comparing just two languages. Only when comparing other languages closely or more distantly related to Tima (and classified as members of the Kordofanian branch of Niger-Kordofanian by Greenberg 1963), does it become obvious that the two words in Swahili and Tima are in fact related. The oldest generation of Tima speakers refers to 'goat' as *címìð*. This form is remarkably similar to the word for 'goat' in languages like Tegali and Tumali, which are also spoken in the Nuba Mountains.

(2) e-mbʊt 'goat' (Tegali)
yimbʊte 'goat' (Tumale)

The Swahili form is widespread across a range of languages now considered to form the Bantu family within Niger-Congo (compare the examples for 'goat' in (1)), and Meeussen (1969) reconstructs **-búdį* for their common ancestor, Proto-Bantu. In his contribution on the grammatical reconstruction of Proto-Bantu, Meeussen (1967:102) reconstructs a (homorganic nasal) prefix as well as a pre-prefix or augment **ji-* for the noun class to which this lexical root belonged. The emerging original Proto-Bantu form for 'goat', **ji-m-búdį*, is remarkably similar to the form in the Niger-Congo languages spoken in the Nuba Mountains in Sudan, illustrated in (2).

When one observes hundreds of roots and grammatical morphemes with similar forms and meanings, especially involving basic vocabulary, a borrowing scenario becomes extremely unlikely. Nevertheless, disentangling borrowing from shared inheritance is not always easy, as further discussed in Section 7.6.

Apart from genetic inheritance and borrowing, chance is another factor explaining the occasional lexical similarity between two languages. Sound symbolism (including onomatopoetic words) may also result in similarities, without necessarily relating to a common historical origin. As shown by Blasi et al. (2016), there are interesting and rather unexpected worldwide sound-meaning biases, for example the frequent occurrence of the vowel 'e' and the consonant 'l' in the word for 'tongue': the corresponding forms in Swahili, *u-leme*, and Tima, *kìlíŋìì*, confirm this.

It was on the basis of multilateral comparisons of basic vocabulary, together with grammatical patterns (for example verb paradigms), that Koelle (1854) published the most extensive historical-comparative study of African languages to that date, involving a comparison of more than 150 languages, based on interviews with liberated victims of the Atlantic slave trade in Sierra Leone and Liberia. Koelle (1854) identified 11 language families, alongside several 'unclassified and isolated languages'.

During subsequent decades and in the first half of the 20th century, a range of scholars aimed at continent-wide classifications of African languages. Among them was Lepsius (1880), whose contribution is characteristic of the era in which it was written. Lepsius distinguished between Bantu, constituting a so-called Southern zone, as a genetic group using

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

prefixes, and Hamitic and Semitic languages as languages of the so-called Northern Zone, with suffixes and gender distinctions. Based on the dominant hypothesis of those days that languages develop from morphologically complex into more reduced systems, Bantu languages were assumed to represent an original stage, whereas languages of the so-called Middle Zone were assumed to be a mixture of Bantu and Hamitic languages, so-called 'Mixed-Negro languages'. This Middle Zone contained 'proto-African' languages and infiltrating Asian languages, but also Bushman and Hottentot.

For reasons of space, it is not possible to discuss the contributions of other scholars during the 19th and 20th centuries here. The interested reader is referred to Childs (2003:19–54), Doneux and Rey (2003), Dimmendaal (2011:75–92, 307–331), and the companion volume to the present handbook (Wolff 2019) for detailed discussions.

It is probably fair to say that, with few exceptions, much of the classificatory work on African languages during the first half of the 20th century was characterized by a mixture of genetic and typological classifications and marred by racist connotations. It was the late Joseph Greenberg (1915–2001) who tried in a range of articles between 1949 and 1954, published as a monograph in Greenberg (1955), to break with this tradition, and to arrive at a continent-wide genetic classification using multilateral (or mass) comparison without the misleading 'noise' of typology and racial or cultural considerations. While he established 16 independent families in his earlier studies, he hypothesized four phyla (stocks, or macro families) in his follow-up study, Greenberg (1963): Afroasiatic, Khoisan, Niger-Kordofanian, and Nilo-Saharan. Tables 7.1–4 summarize this classification, as well the critiques and alternative hypotheses formulated by various scholars in response to Greenberg's seminal contribution. In addition, the tables provide information on the application of the comparative method (discussed below) to subgroups within these hypothesized macro-families.

As these tables show, there is disagreement among scholars about the proper classification of a range of groups. Nevertheless, Greenberg's (1963)

Table 7.1 The classification of Afroasiatic by J. H. Greenberg and subsequent views

Greenberg (1963)		Disputed affiliations	Comparative method
Afroasiatic	Berber Chadic Cushitic Egyptian Semitic	• The genetic affiliation of Omotic (called Western Cushitic in Greenberg 1963) is disputed by some scholars • The Eastern branch within Omotic is not part of Omotic, according to some scholars	• All subbranches, i.e., Berber, Chadic, Cushitic, Omotic, Semitic • Historical development of Egyptian known through historical documents; the latest stage of Egyptian, Coptic, is still used as a liturgical language

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Table 7.2 The classification of Khoisan by J. H. Greenberg and subsequent views

Greenberg (1963)		Disputed affiliations	Comparative method
Khoisan	Northern Khoisan Central Khoisan Southern Khoisan Sandawe Hadza	• Northern, Central, and Southern Khoisan assumed to constitute independent families by specialists • Hadza is a linguistic isolate • Sandawe probably forms a genetic unit with Kwadi and Central Khoisan	• Central Khoisan

Table 7.3 The classification of Niger-Congo by J. H. Greenberg and subsequent views

Greenberg (1963)		Disputed affiliations	Comparative method
Niger-Kordofanian (referred to as Niger-Congo in more recent studies)	West Atlantic Mande Gur Kwa Benue-Congo Adamawa-Eastern Kordofanian	• The genetic unity of Atlantic (Greenberg's West Atlantic), in particular the status of the Bijogo group, is disputed • Mande and Ubangian (called 'Eastern' in Greenberg 1963) may constitute independent families (although the Mba group within Ubangian may be part of Niger-Congo) • Dogon as an independent family rather than as a subgroup of Gur • Ijo (and Defaka) as an independent family rather than as a subgroup of Benue-Congo • Kadu (called 'Tumtum' in Greenberg 1963) as an independent family • Kordofanian probably consists of a number of distantly related Niger-Congo branches rather than forming a single branch	• Northern branch and the Bak branch within Atlantic • Gur • Mande • Kwa • Bantoid and other subgroups of Benue-Congo • Subgroups of Kordofanian

Table 7.4 The classification of Nilo-Saharan by J. H. Greenberg and subsequent views

Greenberg (1963)		Disputed affiliations	Comparative method
Nilo-Saharan	Songhai Maban Fur Saharan Chari-Nile Koman	• Songhai (Songhay) as an independent family • Koman (=Coman), as an independent family • Mimi of Nachtigal and Mime of Gaudefroy-Demombynes as isolates rather than as members of the Maban group • Gumuz as an isolate rather than as a member of Koman • Kuliak (treated as a branch of the Eastern Sudanic subbranch of Chari-Nile in Greenberg 1963) as an independent family	• Songhay • Maban • Central Sudanic • Eastern Sudanic as a branch of Nilo-Saharan, with sub-branches: Daju, Nilotic, Nubian, Surmic, Taman • Koman and B'ata

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

classification still forms the heuristic basis for conferences and workshops on African languages, as well as for the classification of monographs and subject indexes in university libraries. Presumably, all human languages are related; one cannot prove that they are not. The point is to present evidence that they are, and scholars disagree on what constitutes 'sufficient' evidence. Some historical linguists mention 12,000 years as an upper limit for the establishment of genetic relationships between languages, as much of the inherited lexicon and grammatical structure will have been replaced within this time span. A classification that treats, for example, Fur as part of Nilo-Saharan (as Greenberg did in his 1963 publication), rather than as an independent family (as in Greenberg 1955) in and by itself is not very interesting unless this helps to explain grammatical phenomena historically. As argued in Chapter 11 on Nilo-Saharan, several grammatical features in Fur do indeed find a natural historical explanation when comparing these with cognate forms in Kunama and other groups, all of which are hypothesized to belong to the same phylum in Greenberg (1963).

Over the past decades, several linguistic isolates, that is, languages without any obvious relatives, have been identified as well. These are listed in Table 7.5 (see also Map 7.1). Hadza, for example, was classified by Greenberg (1963) as a Khoisan language, but the comparative evidence for this claim remains rather weak.

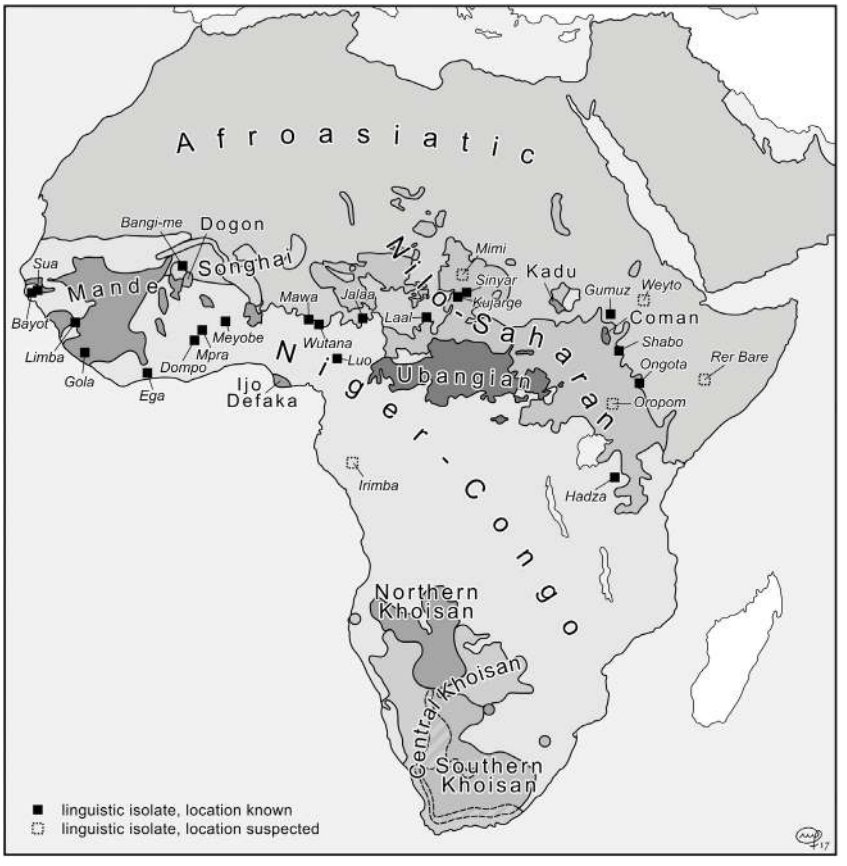
7.3 The Comparative Method

Apart from multilateral (or mass) comparisons, which were used in the past in order to come to grips with the classification of the roughly 2,000 languages spoken on the continent, the comparative method has been applied successfully to well-defined language families, which may or may not belong to larger phyla or stocks such as those hypothesized by Greenberg (1963). This classical method was first applied within Indo-European, but it

Table 7.5 Linguistic isolates

Bangi Me (Mali)	Mawa (Nigeria)
Dompo (Ghana)	Meyobe (Benin, Togo)
Ega (Ivory Coast)	Mimi of Decorse (Chad)
Gola (Liberia, Sierra Leone)	Mimi of Nachtigal (Chad)
Gomba (Ethiopia)	Mpra (Ghana)
Gumuz (Ethiopia, Sudan)	Ongota (Ethiopia)
Hadza (Tanzania)	Oropom (Kenya, Uganda)
Irimba (Gabon)	Rer Bare (Ethiopia)
Jalaa (Nigeria)	Shabo (Ethiopia)
Kujarge (Chad)	Sinyar (Chad)
Laal (Chad)	Sua (Guinea-Bissau)
Limba (Sierra Leone, Guinea)	Weyto (Ethiopia)
Luo (Nigeria)	Wutana (Nigeria)

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007



Map 7.1 Linguistic isolates

has also been applied successfully to a range of African language families, as summarized in the tables above. Some basic principles of this method and some of its problems are illustrated with examples from (Narrow) Bantu below.

Koelle (1854) had identified a group called 'Kōn'go N'góla languages' as one of 11 genetic groupings, but the systematic comparison of what would come to be called Bantu received an important impetus through Bleek (1862, 1869), who may be seen as the founder of comparative Bantu studies, as well as Meinhof (1899, 1906).[1] Whereas Bleek and Meinhof had to rely on a small set of languages, modern researchers can rely on hundreds of synchronic and diachronic studies on Bantu languages, as a result of which it has become the best studied African language family. This subgroup of Niger-Congo plays a central role in the next sections.

[1] Schadeberg (2003:144) points out that 'Bleek (1856) recognized not only the unity of the Bantu language family but also its relation to, what came to be known as, Niger-Congo and Kordofanian after Greenberg (1963); he names as speakers of related languages "the Otsi, or Ashantee" (Akan), "the Bullum, and the Timneh of Sierra Leone", "the wide-spread Fulah" (Fulfulde), "the Accra" (Ga), "the Wolof", "the Ukuafi" (Maasai, some of whom speak Bantu languages), and "the Tumale of Darfur" (speaking a Kordofanian language).'

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Table 7.6 Lexical cognates in some Bantu languages

Cognate lexeme		Swahili	Nyamwezi	Kinyarwanda	Luba-Kasai	Proto-Bantu
1	dog	m-bwa	bwaa, -waa	-bwa	-bwa	*-búa
2	goat	m-buzi	-bʊli, -ßʊli	—	-buʒi	*-budi̧
3	shoulder	m-bega	-ßega	-bega	-beya	*-bega
4	penis	m-boo	-ßolo	-boro	-bolo	*-bodó
5	mosquito	m-bu	-bʊʊ	-bu	-bu	*-bu

7.3.1 Structural Changes

Because, for most words in languages, the link between sound and meaning is arbitrary, their actual resemblance between two or more languages in a range of basic vocabulary words requires a historical explanation of common origin. The pioneers of Indo-European studies in the 19th century, who came to be known as the Neogrammarians, discovered that such formal resemblances are systematic, and can be covered by regular 'sound correspondences' between genetically related languages. From a modern perspective, it would be more appropriate to talk about 'phoneme correspondences', but phonology as a linguistic concept was only established in 1881, when J. Baudouin de Courtenay introduced concepts like 'phoneme' and 'phonetic alternation' in scientific publications in Polish. This conceptual innovation of phonology is illustrated below with examples from the Bantu branch within Niger-Congo, and specifically from Swahili.

In spite of the fact that Swahili has borrowed extensively from Arabic and Persian (Farsi), its core vocabulary was not affected in most cases, but shows clear cognates with other Bantu languages, as the following lexical comparisons with Kinyarwanda (in Rwanda) and Luba-Kasai (in the Democratic Republic of the Congo) illustrate. The reconstructed Proto-Bantu forms are derived from Meeussen (1969), who based his reconstructions on a systematic comparison of more than 400 Bantu languages.

The Swahili examples (Table 7.6) involve nouns preceded by a nasal class prefix. When adding further examples, it becomes clear that in intervocalic contexts, that is, where a different noun class prefix or a verbal prefix ending in a vowel preceded, the original bilabial stop underwent different changes. Synchronic alternations in Swahili, as with the lexical root for 'two', for example *m-bu m-bili* 'two mosquitos' and *wa-tu wa-wili* 'two persons', suggest that this conditioning was relevant historically.

Based on these lexical cognates, one can establish slightly complex correspondence sets (Table 7.7).

Because these correspondences are recurrent and regular, they must be the result of a systematic restructuring of phonological systems. Synchronically, *w* and *v* are distinct phonemes in Swahili, that is, sound units that cannot (or can no longer) be treated as variants (allophones) of *b*. They may occur in the same position within a word (e.g., as a result of borrowing from Arabic, as with the word for 'seven', *saba*, which forms

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Table 7.7 Cognate sets in Bantu

Cognate lexeme		Swahili	Nyamwezi	Kinyarwanda	Luba-Kasai	Proto-Bantu (Meeussen 1969)
6	shine	-wa-	-ßal-	—	-bal-	*-bád-
7	courtyard	-wanja	—	-baandza	-banza	*-bánjá
8	stone	-we	-we	-buye	-bwe	(-)
9	thigh	-weo	—	-biero	-belo	*-bédo
10	break	-vunj-ik-	-ßɪɪn-zik-	-vun-	-vung-	*-bʉnj-
11	rain	-vua	—	-vura	-vul-	*-bʉ́da
12	excrement	-vi	—	-vyi	—	*-bɟ́
13	mould	-umb-	-ßʊʊmb-	-buumb-	-bumb-	*-búmb-
14	become rotten	-oz-	-ßol-	-bor-	-bol-	*-bod-
15	see	-on-	-ßon-	-bon-	-mon-	*-bón-

Table 7.8 Correspondence sets between four Bantu languages

Relevant lexemes	Swahili	Nyamwezi	Kinyarwanda	Luba-Kasai
1, 2, 3, 4, 5	-b	-ß, -b	-b	-b
6, 7, 8, 9	-w	-ß	-b	-b
10, 11, 12	-v	-ß	-v-	-v
13, 14, 15	-ø	-ß	-b	-b (-m if a nasal follows)

a minimal pair with *sawa* 'like, equal'). However, in inherited Bantu roots they go back to one phoneme, the voiced bilabial stop **b*. The exact conditioning for these changes, including the loss of the original **b* before non–high back vowels (as in lexemes 13, 14, and 15), becomes clear only once correspondences for vowels have been worked out as well and once additional Bantu languages are brought in for comparison. The **b* >*v* shift involved a split whereby fricativization occurred before high close vowels (e.g., **-bʉnj-* > *-vung-* 'fold'). For Proto-Bantu, a seven-vowel system can be reconstructed. The initial stop is still found in Bantu languages that have retained the original seven-vowel system (**ɟ, *i, *ʉ, *u, *e, *o, *a*), for example Bobangi, where the root for 'break' is *-bun*. Swahili has five vowels, due to a historical merger of **ɟ* and **i*, and **ʉ* and **u*. As a result, **v* is no longer an allophone of **b*, as both can now occur before high vowels, thereby establishing their status as independent phonemes.

Fricativization of intervocalic vowels is a 'natural', that is, frequently occurring, sound change cross-linguistically; see Dimmendaal (2011:22–58) for a discussion of common sound changes, also showing that it is not only speakers' articulations that play a role in the historical reinterpretation of sound systems, but also the acoustic interpretations of hearers. As shown by the examples from Luba and Kinyarwanda in Table 7.8 above, these languages underwent the same fricativization process as Swahili (and they also have a five-vowel system). When two or more related languages

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

undergo the same change, whether phonological, morphological, or lexical in nature, this may be due to independent developments; alternatively, the innovation was shared, that is, it occurred before these languages emerged as separate languages. The more innovations two or more languages share, the more likely it is that they were shared, that is, that they occurred before these languages emerged as separate languages, thus pointing towards a closer genetic relationship. No velar nasal *ŋ can be reconstructed for Proto-Bantu (Meeussen 1969). Comparative evidence suggests that this was a shared innovation of languages classified as Narrow Bantu, setting the latter apart from closely related groups such as Ekoid (with which Narrow forms the Bantoid branch within Benue-Congo, the latter in turn forming a primary branch of Greenberg's Niger-Congo, which in turn forms a coordinate branch with Niger-Kordofanian in Greenberg 1963). Greenberg (1963:33–34) gives examples of cognates between Efik, a language belonging to the Cross-River branch within Benue-Congo, and Proto-Bantu, in order to illustrate regular tonal correspondences. But several of these cognates also illustrate the loss of the velar nasal as an apparently shared innovation of (Narrow) Bantu.

(3)	Proto-Bantu	Efik (Cross-River)	
	*-tuɟ	-tɔŋ	'ear'
	*-duɟ	-dɔŋ	'knee'
	*-naɟ	-naŋ	'four'

When comparing such lexical roots with cognates in Ekoid Bantu, one finds that the velar nasal still occurs in these languages, for example in Ejagham *-tûŋ* 'ear', *-rûŋ* 'knee' (data from Crabb 1965). The loss of the velar nasal is therefore a shared innovation of all the languages belonging to the Narrow Bantu branch within Bantoid (Dimmendaal 1988:233).

It is not always easy to distinguish shared innovations from areal spreading (i.e., borrowing) of innovations, in particular for Narrow Bantu (Schadeberg 2003:156–159). Bound grammatical (function) morphemes are less subject to borrowing, and consequently are sometimes more relevant for genetic classifications. The following example from Semitic illustrates this point. Like in Bantu, subgrouping on the basis of shared phonological innovations is notably difficult in Semitic, but shared morphological innovations provide a more clear-cut picture.

Akkadian manifests the highest degree of formal heterogeneity, with an initial 'k' for the first person subject marker on verbs, and 't' for the second person. In Arabic, paradigm levelling occurred, whereby the first person marker underwent a sporadic change *k >t by analogy with the second person, whereas in Ethiopian Semitic, the first person formed the basis for this levelling process, as shown by Hetzron (1976), who gives additional examples of shared morphological innovations, all pointing towards the following family tree for Semitic (Figure 7.1).

For a more extensive discussion of morphological reinterpretations, the reader is referred to Dimmendaal (2011:93–114, 290–297, 365–370).

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Table 7.9 Pronominal suffixes in Semitic

	Akkadian	Arabic	Ethiopian Semitic
1SG	- (a-)ku	-tu	-ku
2SG:M	- (a-)ta	-ta	-ka

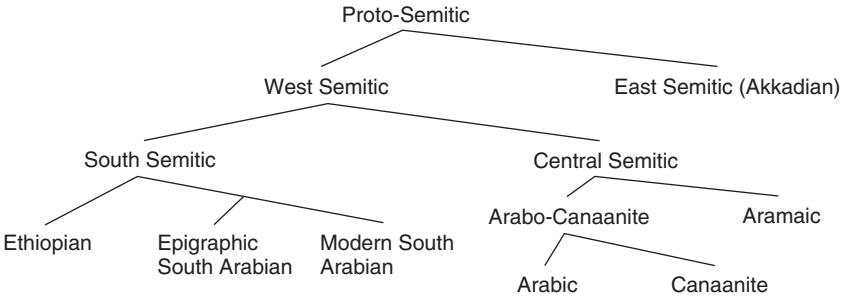


Figure 7.1 The subclassification of Semitic

While the historical-comparative study of genetically related languages in Africa has shown that regular sound changes can usually be established on the basis of cognate roots and affixes, it is also clear that irregularities occur for a number of reasons. One reason, analogical levelling, has already been illustrated for Semitic above. Another reason, involving so-called 'sporadic changes' in individual words with a high frequency of usage, is the common tendency to pronounce such words faster than words with a lower frequency, which in turn results in the loss of segments in the former. This is probably the main reason why, cross-linguistically, one tends to find irregular singular/plural alternations in basic vocabulary items.

Tabuization of specific words is another reason why sound changes (and thereby correspondences) are not always regular. As shown by Herbert (1990b), Bantu languages belonging to the Nguni group have replaced inherited Bantu consonants with clicks (which in turn were borrowed from neighbouring Khoisan languages) as part of a strategy known as Hlonipa, aiming at the omission of words sounding like the names of key figures in the society. Hence, the Proto-Bantu root for 'extinguish, be extinguished', reconstructed as **-dím* by Meeussen (1969), has a reflex *-cima* in the Nguni language Zulu, whereby 'c' (representing a dental click) replaced the original **d* in this particular root.

Frequency of usage is an important and therefore central parameter determining the (in)stability of both phonological and morphosyntactic phenomena in two ways. First, it affects speech rates (thereby triggering sporadic sound changes); second, frequency of usage may affect the stability of sounds or grammatical structures. Thus, irregular present/past alternations in Germanic languages (e.g., English *find/found*, German *finde/fand*, Dutch *vind/vond*) are extremely stable (in spite of the fact that

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

the approximately 160 irregular verbs in German, for example, have to be memorised by speakers). Because of their high frequency, they are acquired by children at an early age, and are thus remarkably stable historically.

Frequency of usage may also be reinforced by patterns of multilingualism. This is presumably the main reason why certain sounds are more common in some areas than in others. For example, labiovelars (or labial-velars) like *kp*, *gb*, and *ŋm* are 'marked' on a worldwide scale, but they are quite common in languages across West Africa and Central Africa, particularly in Niger-Congo and the Central Sudanic branch of Nilo-Saharan. Language contact and corresponding patterns of multilingualism have contributed to their spreading. Consequently, labiovelars also spread into neighbouring Chadic languages in north-eastern Nigeria (Greenberg 1983:6), for example in trade languages like Bacama.

Once languages (or their speakers) move into areas where such phonetic features are absent, such sounds may disappear. For example, in geographical outliers such as the Niger-Congo languages in the Nuba Mountains in Sudan, only Katla and the closely related language Julut have labial-velars. In Tima, also closely related, these have become labialised velar stops (**gb*> **g (w)*>*k (w)*).

(4)

Katla	Tima	
gb-ɔ́lâc	k-ʌ́lɨ̀h	'(kind of) spear'
gb-ə̀lana	kə̀-láánù	'elder'
gb-àjàŋ	k-àràŋ	'leopard'
g-àc	k-âh	'head'
g-û	k-úù	'dog'
g-únû	k-ɔ́nɔ̀	'ear'

Whether sound changes are blind to the overall phonological system of a language is a matter of dispute. Historical changes in African languages with so-called advanced tongue root (ATR) vowel harmony, for example, suggest that system preservation sometimes does play a role. Within the Eastern Nilotic branch of Nilotic, languages constituting the Bari subgroup have a classical system with five [–ATR] vowels and five [+ATR] vowels. In the so-called Non-Bari languages (forming the other primary branch of Eastern Nilotic), the tenth vowel **ə* shifted to and merged with **a* in roots, but in suffixes it shifted to and merged with **ɔ* or **o*, depending on the language, thereby maintaining a system of alternating suffixes (Dimmendaal 2011).

While loss of the central [+ATR] vowel *ɜ*, *ʌ*, or *ə* (depending on the language) is common in a range of Niger-Congo and Nilo-Saharan languages (where ATR harmony is endemic), another type of restructuring occurred in Western Nilotic Jumjum. As shown by Andersen (2006), the correspondence sets between Jumjum and the closely related language Mayak at first

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

sight suggest a 'flip-flop' or 'swopping' rule in one of the two languages, as the vowels in cognate roots have opposite ATR values:

(5)

Jumjum	Mayak	
wiil	wɪɪl	'tail'
buuy	bʊʊr	'shoulder'
pɪɪk	pii	'water'
lʊʊm	luum	'grass'

Based on lexical cognates and regular correspondences with other Western Nilotic languages, Andersen (2006) postulates the following regular sound changes (in the order presented below) for Jumjum (historical changes in Mayak are not further discussed here):

(6) *ɪ > i and *ʊ > u (raising of high [–ATR] vowels)
*e > ɪ and *o > ʊ (raising of mid [+ATR] vowels)

Apart from ATR harmony, there is much to be said about tone as another prosodic property shared by many languages south of the Sahara. The number of tonal registers ranges from two to five, with language families in the geographically peripheral zones (e.g., Atlantic or Omotic) also showing transitions between tone and pitch accent or stress systems.

Again, for reasons of space, these rather fascinating historical changes cannot be further discussed here; see Chapters 10 and 11 for more information.

Syntagmatic changes because of sound units assimilating to neighbouring segments in adjacent morphemes may result in new allomorphs, and thereby in morphological complications. The Perfective (or Anterior/Past) marker in Bantu is an example of this. Bastin (1983) reconstructs this suffix as *-*ɟide*, but also points out that in many Bantu languages it fuses with the lexical root or stem (i.e., root plus derivational suffixes), thereby causing 'imbrication' (or 'overlapping'). Hyman (1995a) gives a detailed account of the various factors influencing its formal realization in Bemba. Some of the allomorphs can be predicted on the basis of the root vowel and the oral or nasal character of the root-final vowel, but a further range of fusion processes occur in this language, whose conditioning is also morphological in nature (examples adapted from Hyman 1995):

(7)

Root		Perfective stem
cít-	'do'	cít-il-e
eb-	'tell'	eb-el-e
lim-	'cultivate'	lim-in-e
tém-	'cut'	tém-en-e
sákat-	'seize'	sákeet-
fyáal-	'give birth'	fyéel-
ceesh-	'cause to dawn'	cees-esh

Thus, with some verbs, a phonologically transparent form occurs next to an imbricated form, as with 'flash' and 'cause to have a child' below; with

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

others, for example 'want, find', the phonologically transparent form is the norm:

(8)

Root		Perfective stem
byaat-	'flash'	byeet- ~ byaat-il-
fyáash-	'cause to have a child'	fyéesh- ~ fyáash-ish-
fwaay-	'want, find'	fwaa-il (*fweey-)

Analogical levelling, as in the example 'want, find' above, is usually based on more productive and phonologically transparent forms.

When speakers no longer recognize the separate status of affixes, these may become incorporated into stems, as in English *child-r-en*, with the incorporated plural *-r* still found in cognate stems in Germanic languages such as German (*kind/kind-er*). This latter process, also known as increment, is common in Niger-Congo noun class systems, and appears to have occurred independently in languages belonging to different subgroups.[2] For example, the Proto-Bantu noun class prefix **ku-* is no longer used with nouns in Swahili, and only occurs as an infinitive prefix (*ku-/kw-*) with verbs. The prefix was incorporated into the noun stem with nouns formally occurring in this noun class, as in *kwapa*, from Proto-Bantu **ku-japa* 'armpit' (Meeussen 1980), which is treated synchronically in Swahili as a noun belonging to class 5, as shown by its agreement marker *li-*, rather than *ku-*.

Africa is often seen as a 'hotbed' of noun classes and verbal extensions. Interesting as these morphosyntactic phenomena are in and of themselves, it is also time to get away from clichés about the continent, in order to start investigating the historical development of a range of other fascinating morphosyntactic phenomena. These include such varied topics as the historical development of case systems (including split ergativity), converbs, numeral classifiers, verbal compounding, and the questions of how derivational affixes become inflectional morphemes and vice versa, and how such affixes develop out of compounds.

7.3.2 Semantic Change

While considerable progress has been made in our understanding of phonological and, to a lesser extent, morphosyntactic changes, semantic changes remain less well understood, because they are frequently associated with the cultural specifics of a speech community. The Proto-Bantu root **-bánjá* 'courtyard', in Table 7.7 above, for example, has been retained with the same meaning in a range of Bantu languages, but its derived meanings (historically) include 'family branch' in Bolia, and 'village where the chief resides', as in Mbala (both languages being spoken in the Democratic Republic of Congo). Specific semantic extensions may also

[2] This parallel development of features that were not present in the common ancestor is also known as drift, convergent evolution, or homoplasy (Dimmendaal 2011:365–370).

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

be common to areas manifesting phonological and morphological convergence, as Hayward (1991) shows for Ethiopia, where Cushitic, Omotic, and Semitic, as distantly related members of the Afroasiatic phylum, manifest identical lexicalization patterns:

(9)	ayət’	1. mouse rat; 2. biceps muscle (Amharic; Semitic)
	hantuuta	1. mouse rat; 2. biceps muscle (Oromo; Cushitic)
	ʔeč’ eré	1. mouse, rat; 2. biceps muscle (Gamo; Omotic)

(10)	k’ ädda	1. draw water; 2. copy, imitate (Amharic; Semitic)
	waraabe	1. draw water; 2. copy, imitate (Oromo; Cushitic)
	duuk’k’ídes	1. draw water; 2. copy, imitate (Gamo; Omotic)

These examples illustrate the two most common types of semantic changes: metaphorical and metonymic extensions, which may occur within a specific semantic field (intrafield changes) or between semantic fields (interfield changes). These extensions also play a role when lexemes develop a more grammatical function, or when one specific grammatical function is extended into another. Research within these latter domains has become particularly popular following the influential contribution of Heine et al. (1991), who show that there is a cross-linguistic tendency to move from concrete to more abstract meanings, for example from spatial to temporal meanings.

A central claim of grammaticalization theories is that grammatical markers tend to be drawn from specific lexical domains, and that their functional development is unidirectional historically. The important point to keep in mind from a historical-comparative point of view is that such claims on specific grammaticalization paths need to be supported by way of the comparative method; they cannot simply be assumed, otherwise such claims become circular.

The studies in Devos and van der Wal (2014) present elegant examples of ‘less beaten grammaticalization path[s]’, as the editors call them. While ‘come’ and ‘go’ are common source items for tense and aspect markers cross-linguistically, something which has been known for a few centuries, a wide range of other grammatical extensions are attested. Carlson (2014) discusses the verb *sa* ‘go’ in the Gur language Supyire, whose function has been extended into that of an intensifying discourse marker (Carlson 2014:257):

(11)	k’a	sà	a	pèè
	it PERF	INT	SCN	be.big
	‘it is really big’			

Other functional extensions of movement verbs include the use of the verb for ‘go’ as a focus marker in the Bantu language Shangaci, as a passive auxiliary in the Romance languages Italian and Romanian (where it may also be used as a copula), and as a necessive (modality) marker in Indo-Aryan languages (Devos & van der Wal 2014).

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

By using specific morphemes in new morphological or syntactic environments, additional meanings may develop. However, apart from polysemy resulting from their use in new constructions, paradigmatic oppositions (so-called relations in absentia) remain important for the development of grammatical systems, as the following example from the Omotic branch of Afroasiatic shows. Omotic languages usually make a formal distinction between a formally marked declarative mood and a zero-marked Interrogative mood (alongside other mood distinctions expressing illocutionary force) on the verb. Amha (2001:288) gives the following examples from Maale:

(12) ʔatsí ziginó mukk-é-ne
person-M:NOM yesterday come-PERF-AFF:DECL
'the man came yesterday'

(13) ʔatsí mukk-íya 'did the man come?'
person-M:NOM come-PERF:Q

In Sheko, which belongs to another subgroup within Omotic, a number of mood markers are found. They are not cognate with the marker *-ne* in Maale (their selection being determined by tense and modality), but the paradigmatic contrast is maintained (data from Hellenthal 2010:402, who points out that '[t]he absence of a modal marker thus distinguishes between declaratives and interrogatives').

(14) a. ṇ=māāk-ā b. ṇ=māāk-ā-m
1SG=tell-put[Q] 1SG=tell-put-IRR
'shall I tell?' 'I will tell'

The presence of the mood marker helps to focus on the correctness of a statement. As shown in the studies edited by Crass and Meyer (2007), there is a wide spread etymological link between demonstratives, copulas, and focus markers in Ethiopian languages, a phenomenon that is also well known cross-linguistically. The use of the same morpheme for semantically related functions, that occur in different syntactic contexts (for example after a noun or verb), is known as heterosemy. According to current grammaticalization theories, there is always a unidirectional change from copula to focus marker (see Heine & Kuteva 2002:95–96). However, as Sasse (2007:48) points out based on evidence from Eastern Cushitic, '[c]hanges or extensions of meanings within a semantic "landscape" are never unidirectional. The fact that "normally" focus markers come from copulae does not preclude the possibility that it may sometimes be the other way around.'

7.4 Internal Reconstruction

Languages with rich morphophonological alternations, as manifested in alternative realizations of the same morpheme (from a semantic point

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

of view) in different morphological environments, usually provide additional means of reconstructing older stages of languages. The Surmic (Nilo-Saharan) language Baale, in southwestern Ethiopia, presents such a case. Baale does not allow the phonetic realization of stops in word-final position. However, once suffixes are added, otherwise unpredictable consonants show up, conditioning the exact shape of, for example, number suffixes for nouns.

(15)

Singular	Plural	
mɛɛlé<k>	mɛɛlɛ-k-ká	'axe'
keyí<c>	keyi-c-cá	'udder'
agá<t>	aga-t-tá	'tongue'

The root-final 'floating' stop (represented as <c> in the examples above) also manifests itself when other bound morphemes (suffixes or enclitics) are added, such as a possessive linker *-a-* followed by a pronominal possessive (*-naandí* 'my' in the example below):[3]

(16) mɛɛlé-k-a-naandí [mɛ̀ɛ̀lɛ́ɣànààndí] 'my axe'

The latent final stop is still pronounced in closely related languages like Didinga: *mɛɛlɛk* 'axe'.

As shown by these examples from Baale, simplifications at the phonetic level (loss of final consonants) result in morphological complications (an increase in the number of allomorphs). Phrased differently, syntagmatic simplifications (e.g., phonetic erosion) frequently lead to paradigmatic complications. Analogical levelling may result, eventually, in the reduction of the number of allomorphs for a particular inflectional or derivational morpheme.

Semi-productive or unproductive morphological alternations in categories involving open sets (usually nouns or verbs) may also provide keys to earlier productive patterns of affixation.[4] Alamin (2012) describes such a situation for the Niger-Congo language Tima. In Tima, the noun-class system characteristic of other Niger-Congo branches (including other languages in the Nuba Mountains belonging to this phylum) has been reduced essentially to a singular/plural alternation involving one productive singular prefix, *kɪ-/ki-* alternating with one productive plural prefix, *ɪ-/i-*. The alternation between *ɪ* and *i* in these prefixes is conditioned by ATR harmony. However, a number of singular/plural alternations involving closed sets of nouns point towards a more elaborate former noun-class system.

(17)

Singular	Plural	
t̪-ɔ̀ndɔ	y-ɔ̀ndɔ	'road, path'
t̪-ɛ́lmɛ̀	y-ɛ́lmɛ̀	'aardvark, ant-eater'

[3] The liaison or linking of otherwise hidden consonants is also attested in languages such as French or Turkish.

[4] Of course, it is not universally true that verbs constitute open classes in languages. In Saharan languages like Kanuri, for example, event structures are expressed by basic (underived) verbs in less than 10 percent of the predications, remaining predications being expressed by a coverb and the light verb 'do, say'.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Additional details on internal reconstruction as a method can be found in Dimmendaal (2011:141–151).

7.5 Dialectology

Speech communities are rarely homogeneous, and it is from pools of dialectal (and idiolectal) variation that language change is drawn. Once specific innovations accumulate in number, be they phonological, lexical, or grammatical in nature, different lects may emerge. The geographical boundaries for specific variation between dialects are called isoglosses (corresponding to isobars, as lines on a map connecting places of equal atmospheric pressure in geography). The more isoglosses coincide, the stronger the linguistic boundary. If there are clear geographical centres from which innovations spread, these may be called focal areas, with transitional and relic areas being less affected by these innovations; for further details on the analysis of language-internal variation, including the use of dialectometrical studies, the interested reader is referred to Dimmendaal (2011:153–178).

Once Indo-Europeanists had established sound correspondences and sound changes between major languages in Europe, usually involving a comparison between standardized national languages (German, English, etc.) or the languages known from ancient written texts (Latin, Greek, Sanskrit), they started looking for the implementation of innovations in dialects of these languages. Detailed dialectal analyses of Indo-European languages at the beginning of the 20th century led to a number of interesting discoveries, including, for example, that homonymic clash due to the phonological merger of words with different meanings may result in lexical replacement. Gilliéron and Roques (1912) describe historical sound shifts which occurred from Latin to the Romance language Franco-Provençal. These sound shifts also resulted in the phonological merger of certain words.

(18) *cattus> *gat ‘cat’
*gallus> *gat ‘cock’

Interestingly, in dialects where these words had merged (as a result of sound changes), the word for ‘cock’ was always replaced by another lexeme, such as *vicaire* or *faisan*. In other words, homonymic clash had occurred to an extent which presumably caused communicative problems, as both words referred to domestic animals whose status in a traditional southern French household rather differed. Consequently, a ‘therapeutic’ change occurred, that is, loss by collision and lexical replacement through semantic widening of a related term in the same semantic field.

Homonymic clash lies at the historical basis of click insertion in Bantu languages of Southern Africa. As shown by Herbert (1990b), among others, these clicks result from unadapted borrowing from neighbouring Khoisan

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

languages (as an areal grouping), but also from the systematic substitution of consonants in Bantu words by way of click sounds, in order to make such words sufficiently distinct from taboo words (as mentioned in the discussion of Zulu in Section 7.3.1).

Homonymic clash between grammatical (as against lexical) morphemes is less well known or studied. When different functions are involved for two morphemes with the same form, this should not interfere with comprehension. But what happens if these morphemes are in paradigmatic opposition to each other or if there is semantic affinity? Bostoen and Mundeke (2011) describe such a case for the Bantu language Mbuun, spoken in the Democratic Republic of the Congo. In Mbuun, the Proto-Bantu causative suffix *-*ici* merged with the applicative suffix *-*id* due to a convergent phonological evolution of the two morphemes. Synchronically, these derived verb stems are expressed by means of consonant alternation on the verb (*ka-lóm* 'leave, come out' / *ka-lómme* 'to remove, take off/out'; *ka-laam* 'to follow / *ka-léemme* 'to follow for'). This syncretism gave rise to a system whereby 'causative verbs which are not applicative in origin may adopt an applied object' (Bostoen & Mundeke 2011:216).

(19) taar o-a-lɛlle mo-an m-byeel
father SM_1-PRS.PROG-shrink (.APPL) NP_1-child NP_9-leg
'my father is checking his step (= is slowing down) for the child'

Another interesting phenomenon, involving dialect change due to areal contact, is illustrated in example (20) below. Swahili, as a major contact language, spread along coastal zones (hence its name, derived from the Arabic word for 'littorals, coasts, seashores', *sawāḥil*, with the prefix *ki-* expressing a way or manner of speaking). The Swahili examples presented in the discussions above are based upon Zanzibar Swahili, which formed the basis for Standard Swahili as the official language of Tanzania. Zanzibar Swahili is a Southern Swahili dialect, as becomes clear from the examples in (20) (derived from Nurse 1982).

(20)

Northern Swahili	Southern Swahili	Standard Swahili	
nd̪e	nje	nje	'outside'
nd̪aa	njaa	njaa	'hunger'
-t̪eka	-cheka	-cheka	'laugh'
-t̪aka	-chaka	-taka	'want'

In the case of words like 'outside', 'hunger', and 'laugh', the same forms are found in Standard Swahili. But the Standard Swahili form for 'want' is *-taka* rather than *-chaka*, probably as a result of borrowing from a northern dialect (possibly the variety spoken in Mombasa). Another isogloss distinguishing Northern and Southern Swahili concerns the expression of the perfective. Whereas Northern Swahili dialects have retained the perfective suffix (see the discussion in Section 7.3.1), Southern Swahili dialects (including Standard Swahili) have replaced the archaic Bantu aspect marker with a perfective prefix *-me-*.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

An interesting historical feature of the expansion of Swahili is the fact that during the past thousand years it spread as far south as the coastal zones of Mozambique, where it had a strong impact on the lexicon of southern Bantu languages such as Makhwa or Koti. During the 19th century, Swahili also spread into the African interior, as a result of which it is now spoken in different varieties in Tanzania, Kenya, and neighbouring countries such as Burundi, Rwanda, Uganda, and the Democratic Republic of Congo. The spreading of this lingua franca in the latter country is particularly interesting because the regiolects are strongly associated with major cities in the eastern and north-eastern zones of the Democratic Republic of Congo, Bunia, Goma, Kisangani, Kivu, and Lubumbashi. As argued by Nassenstein (forthcoming) in his survey of western Swahili dialects, a 'gravity model' is the most fitting template for the spread of linguistic innovations in this area. Such a model was first proposed by Trudgill (1974) in order to explain the diffusion of innovations in English dialects, namely from the largest urban centres to smaller towns and villages. As further shown in Nassenstein (forthcoming), contact with other languages in the area played a central role in the development of western Swahili dialects. The most dramatic restructuring as a result of language contact occurred in Bunia Swahili, where acrolectal, mesolectal, and basilectal varieties are used by speakers as fluid repertoires (Nassenstein, forthcoming). Whereas the acrolectal variety is similar to Standard Swahili, the basilectal variety superficially looks like a pidginised variety, due to its reduced bound morphology. In actual fact, however, it constitutes a replication of structures found in Central Sudanic languages in the area such as Lendu and Ngiti, which are also characterized by a low degree of bound morphology. For example, pronominal subject clitics (as shortened forms of independent pronouns) and tone (which is also a prosodic feature of the basilectal variety of Bunia Swahili) have replaced the typical Bantu prefixes (also for pronominal objects) and former stress patterns.

(21)	Basilectal Bunia Swahili	Standard Swahili
	mi-pa-li-on-o	si-ku-ona
	1SG:S-NEG-PAST-see-2SG:O	1SG:S:NEG-2SG:O-see
	'I did not see you'	'I did not see you'

7.6 Genetic Inheritance versus Borrowing

The second phenomenon of historical interest in comparative linguistics, alongside shared inheritance from a common ancestor, is borrowing, as already illustrated in the preceding section. The transfer of features from one language into another is the second reason why two or more languages may show formal similarities. It is not always easy to distinguish between shared inheritance and borrowing, especially when languages

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

are closely related; an irregular sound correspondence, as with the root for 'want' in Standard Swahili illustrated in the preceding section, is one such indication. Other rules of thumb for the identification of borrowing between two languages, including the direction in which the borrowing occurred, involve the identification of phonological and morphological features that are unique to one of the two languages involved, as shown by Greenberg (1960) in an illustrative contribution on borrowing in the Chadic (Afroasiatic) language Hausa from the Nilo-Saharan language Kanuri. The following two examples illustrate some of the diagnostic features involved:

(22)	Hausa	Kanuri	
	f̃ ùbùùtúú	rəvo-tə́	'write, writing'
	kàrààtúú	kə̀rà-tə́	'read, reading'

Since Hausa does not have voiced labiodental fricatives, it replaced the latter sound with a bilabial stop. This phonological adaptation, as well as the fact that in the Saharan language Kanuri there is a nominalizing suffix *-tə́*, strongly suggests that it was the donor language.[5]

The inverse process, folk etymology (whereby morphologically simplex forms are decomposed into meaningful units), is found in the Swahili word *vi-rusi,* from English *virus*; the corresponding singular in Swahili is *ki-rusi* (the addition of a final vowel being due to a preference for open syllables in this Bantu language).

Borrowing occurs to differing degrees, and may go way beyond lexical transfer to include grammatical borrowing, as shown by Thomason and Kaufman (1988) in their seminal study on language contact. Grammatical borrowing without lexical borrowing is also possible, but this requires a language ideology whereby speakers want to keep their language free from foreign influences. Whether typological distances between languages play a role when grammatical morphemes are borrowed is a matter of dispute. But the different structural effects of Bantu languages in Kenya on two neighbouring Nilotic languages, Kalenjin (which forms a dialect continuum) and Luo, for example, suggests that this factor does play a role. Both Kalenjin and Luo restructured their verb along a Bantu pattern by adding tense marking, something that could be accommodated easily since their verbal templates were already very similar. But Luo also developed a Bantu type of noun class prefixation system, whereas Kalenjin did not; Luo has only two number suffixes, whereas Kalenjin has over 20 different suffixation patterns; see Dimmendaal (2011:188–196), who also discusses the sociohistorical background to this kind of structural borrowing.

As shown by Thomason and Kaufman, grammatical borrowing (alongside lexical borrowing) and the transfer of structural features may come about (a) as a result of shift-induced interference from the primary language

[5] The form *kə́rà* 'reading' in Kanuri derives from Arabic *qaraʕa* with the same meaning.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

(L1) when shifting towards another, more dominant language (L2) or (b) without language shift (because the primary language is emblematic for the community), but with structural interference from a dominant 'interlanguage'. This latter process, which is now commonly referred to as metatypy, is attested in Greek as spoken in Turkey, but also in African languages. As shown by Yigezu (2005) for the Surmic (Nilo-Saharan) language Baale in Ethiopia, even basic vocabulary may be borrowed in such cases.

While mass comparison has yielded interesting first results on genetic groupings for Africa as a whole, the same method may also lead to misinterpretations, if one fails to identify extensive borrowing. According to Greenberg (1963), Northern Khoisan and Southern Khoisan are genetically related. But a study by Traill and Nakagawa (2000) reveals that a range of lexical similarities – these languages have little bound morphology that could be compared – is probably due to borrowing. The Kuliak languages in northern Uganda present a further example of intensive borrowing from Nilotic, which may have misled Greenberg (1963) to classify this group of languages as a subgroup of the Eastern Sudanic branch of Nilo-Saharan; see Chapter 11 for further details.[6]

Language contact phenomena may be divided into four prototypes: (a) borrowing; (b) pidginization and creolization; (c) 'language mixing'; and (d) language loss. These are not mutually exclusive, of course, as borrowing may occur with the second, third, and fourth phenotypes.

Some authors emphasize the (presumed) unique nature of creoles (e.g., Bakker et al. 2011, and *Ethnologue*, including the last updated version by Lewis et al. 2015). Like various other authors, the present author has argued against this so-called 'exceptionalism' (Dimmendaal 2011:213–236; see also Haspelmath 2013 for a succinct criticism). First, pidginization and creolization as restructuring processes come by degrees, depending on factors such as the number of primary language speakers and the intensity of their interactions in multilingual groups. Second, morphological simplification does not occur only when 'emergency languages' develop between multilingual communities; the same processes are attested with speakers of morphologically rich languages encountering speakers of languages with limited bound morphology; see the discussion of Bunia Swahili in Section 7.5.

'Language mixing', as a label describing a third type of contact phenomenon, tends to be avoided by Africanists due to its association with scholarly contributions from the end of the 19th century and the beginning of the 20th century, and specifically those involving claims about the link between language, race, and culture. Moreover, as shown by Velupillai

[6] Dixon (1997) claims that such lexical and grammatical diffusion processes typically occur during so-called equilibrium periods. These follow so-called punctuation, whereby languages develop into different varieties over a relatively short time span as a result of geographical expansion (due, for example, to technological innovations). The interested reader is referred to Dimmendaal (2011:352–354) for a discussion.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

(2015), different contact situations can result in similar 'mixed' languages, and different 'mixed' languages may arise from similar contact situations. The term consequently does not refer to a proper taxonomic class. What is more, virtually all languages are 'mixed', that is, influenced by other languages. Dimmendaal (2011:237–252) refers to such intertwined languages as syncretic languages. An interesting African case of syncretism, involving the structural merger of two languages, is described by Souag (2015) for Tadaksahak, a variety of Northern Songhay. Tadaksahak, which is spoken by nomadic groups in Mali that are culturally close to Tuaregs (who speak a Berber language), shows a strong structural influence from Berber, as manifested in extensive lexical and grammatical borrowings. Historically, the syncretic structure is the result of a shift in language solidarity from specific varieties of Berber towards Northern Songhay. In Tadaksahak, for example, structural changes include the borrowing of a feminine plural ending *-in*, a centrifugal particle *-an*, a preposition *inžin* 'like', and the 'exclusive use of Berber diathesis morphology, resulting in systematic suppletion for non-Berber verbs' (Souag 2015:126–127), as illustrated in (23):

(23)	Verb	Causative	Passive	
	kós	š-ə̀ɣdəš	t-ə̀ɣdəš	'do'
	kaw	š-úkuš	t-úkuš	'take out'

Interestingly, the presence of different languages in a specific area (sometimes belonging to different, unrelated language families) does not necessarily lead to borrowing or convergence. The crucial point is whether bilingualism or multilingualism occurs. Apart from convergence (or 'Sprachbund') areas, there are a number of accretion (or residual) areas on the African continent, characterized by tremendous genetic diversity and typological disparity, such as southwestern Ethiopia and the Nuba Mountains in Sudan. In this latter area, more than 40 languages are spoken in an area roughly the size of Scotland, which belong to three different language families: Niger-Congo, Nilo-Saharan, and Kadu. Languages in this area present a microcosm of typological variation as found on the African continent as a whole. Languages in the Nuba Mountains vary in terms of constituent order, whether they prefer to mark syntactic relations on the verb (head marking) or rather on dependent categories (dependent marking), and whether they have extensive number-marking for nouns or (virtually) none at all. Whereas, in many parts of Africa, people switch between different languages on a day-to-day basis (a process frequently leading towards convergence between the languages involved), it is also clear that this does not apply to areas such as the Nuba Mountains. The use of Arabic as a contact language in this area these days, to an extent that it is gradually replacing local languages, is a fairly recent development. The highly restricted degree of bilingualism and localist language ideologies in the Nuba Mountains is the main reason for the absence of language union or convergence in this area (Dimmendaal 2015b:9–81).

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

One may also observe 'fragmentation belts', where language differences were maintained in the presence of extensive gene flows between people over an extended period of time, as shown by Veeramah et al. (2010) for the Cross River branch of Benue-Congo. While there are more than 60 Cross River languages (spoken mainly in south-eastern Nigeria), the different speech communities are fairly homogeneous in terms of Y-chromosomes and mtDNA markers. Genetic research may thus provide additional interesting insights into our understanding of language ecology. Currently, however, such research is interdicted in some countries (for example Ethiopia), as there is a fear that such findings may be instrumentalized politically.

7.7 Cladistic and Rhizotic Representations of Genetic Relationships

The classical family tree, as a metaphor, represents subgrouping based on the notion of shared innovations between related languages (as discussed in Section 7.3), and as such reflects historical divergence. However, language change also involves convergence, between both unrelated and genetically related languages, as illustrated above. Scholars have been aware of such contact phenomena ever since the first dialectological studies and contributions on creolized varieties of Indo-European languages appeared at the end of the 19th century.

While family trees (i.e., cladistic models) continue to serve as a (metaphorical) representation of genetic subgroupings for languages, more recently, network analyses have become popular in order to represent genetic relationships. The integration of stochastic methods originally designed for biological applications has also made it possible to analyse large quantitative datasets for languages automatically. McMahon and McMahon (2005:122–138) give a detailed discussion of such methods, but their contribution also contains cautions and caveats (which are shared by the present author; see Dimmendaal 2011:276–280). In Chapter 11, a 'classical' family tree is presented for the Eastern Sudanic branch within Nilo-Saharan. Schnoebelen (2009) presents an alternative rhizotic model or NeighbourNet graph of Eastern Sudanic languages, which is reproduced in Figure 7.2.

Here, shared similarities (which are not necessarily shared innovations in the classical, Neogrammarian sense) are visualized by way of a 'clouding' of genetic splits. Figure 7.2 shows, for example, that Eastern Nilotic Maasai is part of a 'cloud' containing Southern and Western Nilotic languages (two additional subgroupings that have been accepted by Niloticists for more than 50 years). As shown in the detailed historical-comparative studies by Voßen (1982) for Eastern Nilotic, there is a range of phonological, lexical and grammatical innovations clearly showing that Maasai

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

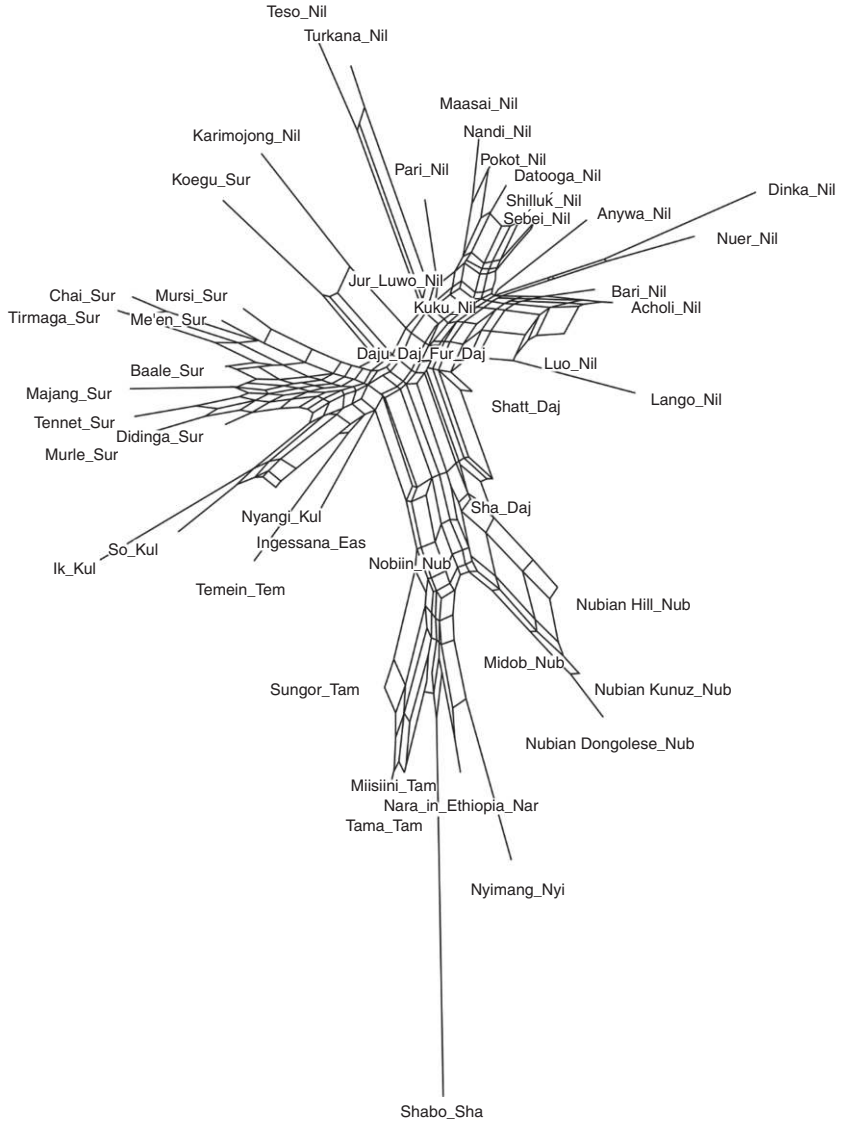


Figure 7.2 NeighbourNet graph of Eastern Sudanic languages

is an Eastern Nilotic language, closely related to the Lotuxo cluster, and forming a subgroup within Eastern Nilotic together with Teso-Turkana. But the network analysis in Figure 7.2 suggests something different. Network analyses are usually based on the counting of (presumed) cognates, although the actual calculations are far more complex than with traditional lexicostatistics.[7] According to the present author, adherents of NeighbourNet graphs are mixing up the genesis of a language – the

[7] For most comparative linguists, lexicostatistics (or glottochronology) constitutes an outdated method for the genetic classification of languages; see Dimmendaal (2011:71–74) for a discussion.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

processes contributing to the present structure of a language – with its genealogy – its inherited structures (an important terminological distinction going back to Manessy 1990). In this respect, such networks are to some extent 'pre-Darwinian' (like the botanical classification by Linaeus), focusing on superficial typological similarities which actually evolved due to areal contact (borrowing).[8]

7.8 Future Directions

The traditional view on language change is that it is a more or less gradual process usually taking place through 'the invisible hand' (Keller 1994), that is, without speakers using the language being aware of these changes. But we now know that this characterization is incomplete. Abrupt creolization processes or the emergence of syncretic languages (as discussed above) show that dramatic changes may occur within a few generations. Moreover, the emergence of urban languages in a range of African metropoles show that 'agency' and deliberate language change are important too. These latter phenomena do not just constitute recent developments, but instead are continuations of ancient traditions, as becomes clear from the survey of language manipulation by Storch (2011).

Current sociolinguistic investigations in an African context show that 'trans-languaging' (as discussed in some chapters in Part III of this volume) also results in new phonological and grammatical structures as a result of contact between African languages and 'colonial' languages. Nevertheless, the continued importance of grammars as abstracted and more or less bounded entities in the mind of speakers should be emphasized, as these are as much part of their competence as fluid practices or repertoires are. If such grammatical systems were just constructs in the mind of scientist, it would be impossible to apply the comparative method, and thereby to establish genetic relationships and to reconstruct regular historical changes between them.

After the numerous attempts at arriving at continent-wide classifications, and many empirically poorly founded claims about the unidirectionality of grammatical change, it is now time to go back to less ambitious – but ultimately intellectually more rewarding – applications of the comparative method to well-defined language families, in order to make progress in our understanding of language change. This would be in line with language typological research in the 21st century, which has also drifted away from large-scale comparisons and macro-areas, as such classifications do not do justice to the tremendous variation between languages.

[8] Biologists would refer to these as 'epigenetic features'. They involve modifications developed by plants and animals during their lives, rather than being part of those species' inherited gene package.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007

Modern technology allows for new quantitative approaches towards the historical-comparative study of languages by connecting huge databases, as with RefLex, a database aiming at the exploitation of existing lexical documentation. As shown, for example, by Segerer (2009), such computer-based techniques allow for the search for cognates through extensive comparisons of databases.

The 'transition problem', that is, understanding how a given variable (whether phonological, lexical, or grammatical) is selected as sociolinguistically significant and thus as generalized by groups of individuals, was one of the central questions in the research initiated by Labov in American cities. How and why linguistic innovations spread is a still largely unanswered question within the context of African linguistics, but initial fascinating results from social network analyses, such as Schreiber and Beyer (forthcoming), provide promising prospects.

As a result of progress in our understanding of typological variation between languages, whether these are genetically related or not, we now know that parallel grammatical systems may come about without areal contact necessarily being involved. An example is the emergence of ergative case systems through the reinterpretation of a peripheral instrumental phrase as a core phrase marking agentivity (and volitionality) in languages spoken in Africa, Asia, or the Americas. Homoplasy (as it is called in evolutionary biology, where phenomena like the independent development of wings or poison as defence strategies in different species are instances of convergent evolution) is therefore another fascinating phenomenon of human language, where the comparative method as applied to different language families and language typology may be combined into a model of dynamic comparison, in order to explain recurring patterns of phonological, morphological, and semantic remodelling in languages in Africa and elsewhere.

Acknowledgements

The author would like to thank Yvonne Treis and Ekkehard Wolff for useful comments and criticisms on an earlier draft, and Monika Feinen for the map and the improved representation of the NeighbourNet graph.

Downloaded from https://www.cambridge.org/core. Nottingham Trent University, on 06 Sep 2019 at 17:35:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108283991.007