THE CONTRIBUTIONS OF LINGUISTICS TO THE STUDY OF HISTORY IN AFRICA

BY DEREK NURSE

Memorial University of Newfoundland

The work of historical and comparative linguists has long interested African historians. By classifying languages into families, linguists provide models of their historical development that may point to historical events and processes that occurred among peoples speaking those languages. Once classified, linguists can then reconstruct earlier forms of present languages, thus providing direct evidence of words, their meanings and historical influences in the past. Finally, linguists seek to explain innovations that are revealed in their reconstructions by pointing to a combination of internal linguistic developments and different forms of contact that occurred among speakers of different languages.

Simple classification, based largely on counting cognate words in related languages (a technique known as lexicostatistics), is still a very common activity. However, and thus the one most historians rely on, but lexicostatistics gives only a very limited, and often deceptive, view of language history. Historians should thus be aware of its limitations as well as the potential of a number of important techniques now employed by linguists, including the Comparative Method, reconstruction of ancestral languages, and contact models.

European scholars have been concerned from the early nineteenth century with describing and arranging African languages and making hypotheses about their historical development. Classification of languages into groups of related languages (often presented via the familiar tree model) came first, while reconstruction of ancestral forms— which presupposed acceptance of the monogenetic model1— came second because linguists had to acquire adequate data and a classificatory framework before attempting to reconstruct earlier forms.2 In recent decades, larger numbers of professional linguists, more data and improved theoretical tools and insights have led to more sophisticated classification and larger and better sets of lexical reconstructions. As a result, existing classifications of language families have changed, and we

---

1 Monogenesis is the hypothesis that all languages in a given set developed from a single ancestor. This is often interpreted to mean that all the world's languages developed from a single original language, but it can also mean that all Bantu, Indo-European, or Afro-Asiatic languages each had a single ancestor.

2 Williamson has a good summary of the unfolding of two centuries of classificatory work on Africa's biggest language family, Niger-Congo, that is typical of that on other language families in Africa. K. Williamson, 'Niger-Congo overview', in J. Bendor-Samuel (ed.), The Niger-Congo Languages (Lanham, 1989), 3-8.
now have considerable lexical, phonological and morphological re-constructions for some of these.

Linguists in the nineteenth and early twentieth century commonly associated languages with self-contained ethnolinguistic groups, which were seen as relatively discrete and little changing. Thus 'the' Khoisan were and probably had always been hunters; 'the' Bantu were subsistence farmers who incorporated iron-working and new crops to migrate across Africa, sweeping aside those who stood in their way; while Nilotes, Cushites, Arabs, and other northeners came from the north carrying innovation to the south. Linguists have now modified or discarded these earlier views, and innovation is now seen as having multiple sources. Communities previously seen as living in relative isolation are now seen as co-existing and interacting with others. Assimilation of preceding populations has superseded their replacement. Evidence from analyses of gene flow and DNA has entered the picture. Images of massive migrations are being replaced by those of shorter migrations, local dispersal of people, and diffusion of languages and cultures among pre-existing populations. Comparative and historical linguists have thus become increasingly interested in processes of change and contact. Inspired by work on change in sociolinguistics, new models of change and contact are emerging. We see modern language communities always in contact and constantly evolving, and we project these processes into the past and seek to understand the results.

This emphasis on change and the diversity of evidence now available has made the task of African historians seeking to understand and use linguistic data more complicated and challenging. The following sections will discuss classification, reconstruction and contact - the three mainstays of historical and comparative linguistics - to seek to explain the methods involved, their strengths and weaknesses, their results and their possible uses for historians. Some parts will be familiar to linguists but perhaps less so to historiats, while others will present newer linguistic practice and show its relevance to historiats. A brief final section will discuss current interest among linguists in change.

---

4 Lexis has to do with words and their meanings; phonology with sounds, patterns of sounds and sound processes; and morphology with word structures.


6 A word is in order here on 'schools of thought' and theoretical positions among historical and comparative linguists operating today. Most linguists are interested in the same core set of topics, and they use the same core techniques, which have evolved and been refined through the last two centuries. Most innovations in historical and comparative linguistics in the twentieth century have consisted of incorporating findings from other subdisciplines of linguistics, including sociolinguistics, computational linguistics, phonetics, phonology, morphology and syntax. Differences among linguists are not national or geographical, but center on how far they can or want to incorporate these innovations, on how far they want to extend their horizons and methods, on the use or non-use of specific techniques, and on how far they want to engage in dialogue with non-linguists.

Some of these claims might seem surprising to historiats. African historians tend to think of the linguistic group around Guthrie at SOAS after the Second World War, but such groups no longer exist. Guthrie was remarkable for his rigor, energy, dedication,
While the following is often couched in linguistic language, technical terms are defined as they occur. Historians should be alert to the fact that when the names of languages or linguistic communities are mentioned, the reference is to the languages and not to the people who may have spoken them unless indicated otherwise. In many cases, we do not know what historical populations spoke which languages, and of course, pots do not speak, although many have made mistakes in trying to associate archaeological ‘cultures’ with specific languages or language families. The vexed case of Bantu expansion, discussed below, comes immediately to mind. Similarly, linguistic processes are not necessarily the same as historical ones. Thus, when one notes that languages split or different languages came into contact with one another, this does not mean that people speaking those languages had to move. In fact, languages often begin to differentiate in situ, and contact can occur among peoples living for long periods of time in close proximity to one another.

**LANGUAGE CLASSIFICATION**

**Methods**

*The comparative method*

Starting in the early nineteenth century, linguists gradually worked out a method which enabled them to assign most of the world’s languages to different language families. This came to be known as the Comparative Method and is widely accepted as the standard among linguists today. The Comparative Method was first applied in Germany to the Indo-European group of languages, but it has subsequently been applied to more and more language families, while the methodology itself has become considerably more sophisticated. The Comparative Method is at the center of comparative linguistics, which works backwards, or upstream, from today’s languages to establish genetic relationships among them and to reconstruct their ancestors using exclusively linguistic data. By contrast, historical linguistics charts the forward, or downstream, development of languages from an earlier to a later point and may draw on ancillary data from history and archaeology to reconstruct a plausible history of the language’s development.

The method follows a few relatively simple steps. First, the linguist, suspecting certain languages are related, compares vocabularies from each of them and picks out items with similar or identical phonetic shapes and

---

width of geographical horizons, productivity and advancing the frontiers of African studies, but in historical and comparative linguistic terms, he was reinventing the nineteenth-century wheel. Conversely, scholars such as Ehret or Vansina, well known to African historians, are historians using linguistic methods.

Some historical and comparative linguists are well known to other historical and comparative linguists, but are not well known outside that circle. That is because historical and comparative linguistics is itself outside the mainstream of contemporary linguistics. Most mainstream linguists may be interested in certain topics, such as change, that also interest us, but they are interested in these topics linguistically, not historically. And while Departments of Linguistics contain clusters of linguists, it is hard to think of any that have significant concentrations of historical or comparative linguists. Most such linguists thus work in relative isolation.

* The few dissenters consist of those who deny, or are agnostic about, monogenesis.
meanings (e.g. Swahili m-ti and Giryama mu-zi, both meaning ‘tree’). From these, one identifies regular sound correspondences: sound x in Language A regularly and frequently corresponds to sound y in Language B and sound z in Language C. Ideally, one is able to identify several hundred words with the same or related meanings across the languages, and the sound correspondences should occur as often as possible. Second, for each set of sound correspondences (e.g. Swahili m = Giryama m, Swahili t = Giryama h, etc.), the linguist posits a single original sound in the proto-language from which today’s sounds derive. Third, the linguist then organizes these original sounds into a sound system (consisting of vowels, consonants and accents or tones) for the proto-language. In this way, he constructs the phonetics and phonology of the proto-language. Fourth, the linguist returns to the original body of lexical data and establishes for each set of similar words whether they consist solely of sounds derivable from the proto-language. If so, the words in each each set are said to be cognate, and the linguist reconstructs a single reconstructed word in the proto-language for each set, slowly building up a vocabulary for the proto-language. The vocabulary need not consist only of words or stems, but it may also consist of grammatical elements, such as details of noun classes or tenses in verbs.

The fifth step consists of identifying phonological innovations shared by certain languages that indicate common developments in the related languages after the proto-community disintegrated. While these innovations are not strictly necessary to the reconstruction process, they help the linguist to make an internal classification of the target languages, ordering each relative to others within the set. Such phonetic innovations are later supplemented by lexical and morphological innovations to further establish internal relationships among the languages in the set. Step 1 thus produces an external classification ordering languages relative to others outside the set by including some languages and excluding others; steps 2-4 lead to reconstruction; and step 5 results in an internal classification.

While the Comparative Method is the ideal method of comparison, classification, and reconstruction, it has not been comprehensively and consistently applied to African languages. One reason for this is the sheer amount of data and work involved compared to the small number of linguists working in the field. Computers may help ease the task, but so far they have been of only limited use. The other reason, which has led some linguists to doubt the very applicability of the method to the African context, lies in the complicated and interwoven linguistic picture in Africa. Africa may have 1,200-1,500 languages (see Table 1), whereas Indo-European, the best known language family and the one to which the Comparative Method has been applied most extensively, has some two hundred. African language families also have a much greater time depth than Indo-European ones, and African languages have had many millennia in which to evolve and interact.

---

7 A proto-language is a reconstructed language ancestral to, and based on, a set of later, related, and usually attested, languages. Proto-languages may exist at different levels; a proto-language can not only be reconstructed for each valid language family but also for all intermediate, lower level, sub-branches within the family.

8 Two or more words are said to be cognate if (a) they are of similar or relatable meaning and phonetic shape, and (b) they can be demonstrated to have derived by direct oral transmission from a single item in a proto-language.
compared with just a few millennia for Indo-European. The result is that the African language mosaic is much harder to decipher than Indo-European. Finally, good data on African languages is limited, and few languages are sufficiently well described to apply the Comparative Method properly.

**Mass comparison**

Given the difficulties of applying the Comparative Method, Greenberg developed a second method, that of mass comparison, and applied it to the languages of Africa and the Americas.\(^8\) Whereas his classification of American languages has been controversial, his general framework for assigning African languages to four families has been universally accepted even as some of the details have had to be modified subsequently.

The method has two components. The first consists of comparing lexical data (i.e. vocabulary) from a large range of languages to ascertain their similarity.\(^9\) If significant lexical similarity is found among a sub-set of languages, this similarity is assumed to have resulted from shared descent from a common ancestor. The second component is to compare morphological similarities among the languages in question. Such similarities are thought to be more convincing if they are idiosyncratic, that is, not likely to derive from accident or universal linguistic tendencies. Thus, the pattern of good-better-best or the simple contrast between non-past and past (e.g. jump/jumped) are idiosyncratic innovations specific to the Germanic languages, while noun class systems are specific to Niger-Congo languages and thus confirm the validity of each family.

Greenberg himself thought that his technique ought to be used only as a preliminary to the standard Comparative Method, but his approach was remarkably successful in Africa at assigning all African languages to four macro-level language families and their larger sub-groups. The approach has been much less successful at the micro-level, however, because it lacks the rigor of the Comparative Method. On such a crowded linguistic stage as Africa, a fine tooth comb rather than a broad brush is needed to distinguish levels of relatedness below that of the family. Greenberg was also confident that it was always possible to distinguish similarities derived from shared inheritance from those resulting from contact with other languages. As more cases of extensive borrowing have surfaced, however, linguists today are no longer as confident that it is possible to distinguish between inherited similarities and those resulting from contact in all cases.

**Lexicostatistics**

A further attempt to overcome the problems of applying the Comparative Method is through lexicostatistics.\(^11\) Lexicostatistics is based largely on

---


\(^9\) Greenberg is often photographed beside his famous and compendious notebooks with word sets from many languages.

\(^11\) Lexicostatistics was developed largely by Swadesh in the 1940s and refined later by others. The most concise and up-to-date statement of the method is S. M. Embleton, *Statistics in Historical Linguistics* (Buchum, 1986).
A comparison of too- or 200-word lists of core vocabulary\textsuperscript{12} to calculate degrees of similarity among them, thus avoiding the detailed comparisons of phonology, grammar and vocabulary required by the Comparative Method. In its most rigorous application, a linguist first identifies regular sound correspondences between the relevant languages as in the first step of the Comparative Method, but in less rigorous versions, such correspondences are worked out only roughly, if at all. When done properly, however, one can then assess which words in the constituent languages are cognate. Once all the items on the word lists are compared and the cognates determined, the linguist can then count the total number of cognates in each pair of languages and calculate the percentage of cognation between them, thus providing a statistical measure of how closely the vocabulary of each language is related to that of the other members of the group. And once such degrees of comparative similarity are obtained, the languages can then be represented by the familiar tree diagram to create an overall genealogy of the language ‘family’. Lexicostatistics is thus most useful for internal classification because it shows how closely related each language is to every other member of its sub-group.

Lexicostatistics was widely used with African languages during the 1970s and 1980s, mainly with Niger-Congo and Banu, but also with Nilo-Saharan and Afrasian.\textsuperscript{13} While many practitioners have claimed this was only a preliminary to application of the Comparative Method, few have proceeded to the latter, more thorough analysis. This is unfortunate for lexicostatistics suffers from a number of serious drawbacks. Firstly, all lexicostatistical studies to date have drawn on only a sample of target languages, but studies now show that an analysis based on all the target languages frequently produces a different picture from that based on only a sample of them. Secondly, the linguistic data used has often been drawn from various sources

\textsuperscript{12} Core vocabulary consists of numbers, parts of the body etc. that are more resistant to borrowing and thus more likely to be inherited. It is thus thought to be a more accurate indicator of internal historical development than cultural terms that are more likely to be borrowed from neighboring languages.

and has not always been of sufficient quality to produce reliable results. Thirdly, claims of similarity do not always differentiate between strict cognates and mere look-alikes. Fourthly, as Vansina notes, lexicostatistics often forces one to posit binary splits in the development of related languages where they are not always appropriate. That is, if three languages are compared, the statistical method used always obliges the linguist to say that two of the languages are more similar to each other than either is to the third, even though in fact they might all be very similar. In a situation like this, a three-way split might be a more appropriate historical interpretation.

Fifthly, there is a disturbing correlation between lexical similarity and geographical proximity in several of the lexicostatistical surveys of Bantu. This suggests that languages that have long been adjacent to one another have transferred material so extensively from each other that they now appear to be more closely related genetically than they really are. This is because lexicostatistics measures the degree of similarity in inherited parts of vocabularies, but words are the most easily transferred component of language, and real people have a habit of modifying words so that such transfers are not always apparent to the inquiring linguist. For centuries, many people in Africa have been bi- or multi-lingual, and people who live in adjacent communities and speak both languages are often capable of transmitting phonetic material, making the appropriate changes as they do so, such that one can no longer recognize it as having been transferred in the first place. This distorts measures of lexical similarity and makes languages look more similar than they really are.

Finally, we need to clarify precisely what lexicostatistics measures. If one dialect of Swahili is compared to another, levels of lexicostatistical similarity upwards of 70 per cent are obtained, but most of these words are similar because they are inherited from a pre-Swahili or proto-Bantu stage and thus are of little help in distinguishing Swahili from that stage. A small handful of the 70 per cent, perhaps 5 per cent, however, are shared lexical innovations which entered the language at the proto-Swahili stage and thus serve to distinguish Swahili from its ancestors. Thus, lexicostatistics mainly


16 The Digo of south-east Kenya and north-east Tanzania have long been bilingual in Digo and Swahili, and their ability to transmute material phonetically has deceived at least one set of linguists quite successfully about their taxonomic status. How communities transmute words phonetically can be illustrated by the Swahili dialects spoken in Lamu and Sin in northern Kenya. People in these two communities have known each other for many centuries and are well aware of the regular phonetic differences between their dialects. They know, for example, that where Lamu has *t*, Sin has *c*. The phonetic change from original *t* to *c* in Sin took place centuries ago, and all words with original *t* were changed to *c*. To maintain the character of their speech, people in Sin take words entering from Lamu and change *t* to *c*. Thus the English words (football) ‘team’ and ‘tape’ (recorder) appear in Lamu as *timo* and *topi*, but as *chimpa* and *chepi* in Sin. This constitutes a ‘regular phonetic correspondence’ and would allow an unsuspecting linguist to reconstruct tape recorders and football teams for early Swahili society.

17 D. Nourse and T. Himmelsbach, *Swahili and Shindo* (Berkeley, 1993), 274, 83. Some practitioners have used methods similar to lexicostatistics that also measure similarity or retention. Thus Henrici measures rates of retention in different Bantu languages of words.
measures rates of retention and splits in tree diagrams based on lexicostatistics reflect this. But it is axiomatic in comparative linguistics that innovations indicate shared development better than retention does.\footnote{18}

**Glottochronology**

While lexicostatistics establishes a relative chronology of language development, as represented by the familiar tree diagram, glottochronology is a subset of lexicostatistics used to date those developments absolutely. Taking well-documented cases of language development, such as Chinese, Egyptian, Romance, Germanic or Slavic, it is possible to calculate the average rate of replacement of core vocabulary over a millennium. This average rate of replacement can then be applied to the lexicostatistical results to calculate how long the constituent languages have been leading separate lives and thus to determine when ‘daughters’ split from their respective ‘mothers’. Many historical linguists feel there are serious difficulties with glottochronology, however, because the rate of linguistic change can be quite uneven, proceeding in fits and starts, and is directly affected by external circumstances. And since we usually cannot know the historical circumstances for Africa, indeed they are frequently what we are trying to determine from the linguistic data, glottochronology is unlikely to provide us with reliable dates by itself. Glottochronological dates are much more useful when they can be compared with those obtained by archaeology or other means, but determining what languages were spoken by peoples occupying ancient archaeological sites is also very problematic, as we will see.

**The language families of Africa**

Through the nineteenth and first part of the twentieth century a gradual consensus emerged about how Africa’s languages were to be classified. The most recent stage started with Greenberg’s 1963 classification, whereby virtually all mainland African languages are assigned to one of four families (see Table 1). While there is wide agreement on Greenberg’s general picture, however, two points need to be made. One is that most of the world’s accepted language groupings were not arrived at by rigorously applying


\footnote{18} The truth of this claim may not be immediately obvious to historians because historians and linguists may view the relative values of innovation and retention differently. It can be exemplified by considering what has happened in various Bantu languages to the form *manu*, the reconstruction for ‘person’ in proto-Bantu. After some 5,000 years many Bantu languages still have that shape while others have changed to shapes such as *mitu*, *mutu*, *nashi*, *nada* etc. Linguists need the languages retaining *manu* because they provide a standard against which to measure subsequent innovation in other languages and provide a basis for reconstructing the proto-Bantu form. Such languages are otherwise less interesting, whereas languages which show related changes can be mapped and the data used to hypothesize about possible periods of shared evolution.
Table 1. Language families of Africa

<table>
<thead>
<tr>
<th>Family</th>
<th>Sub-group</th>
<th>Number</th>
<th>Area spoken</th>
<th>Population (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger-Congo</td>
<td>Kordofanian</td>
<td>30</td>
<td>Sudan</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Narrow Bantu</td>
<td>&gt; 300</td>
<td>S/C/E Africa</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>&gt; 600</td>
<td>W Africa</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>c. 1,000</td>
<td></td>
<td>c. 370</td>
</tr>
<tr>
<td>Afrasian</td>
<td>Coptic</td>
<td>4</td>
<td>Egypt</td>
<td></td>
</tr>
<tr>
<td>(Afro-Asiatic)</td>
<td>Omotic</td>
<td>n/a</td>
<td>Ethiopia</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Berber</td>
<td>n/a</td>
<td>NW Africa</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Chadic</td>
<td>125</td>
<td>W Africa</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Cushitic</td>
<td>30</td>
<td>NE Africa</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Semitic</td>
<td>20</td>
<td>N Africa</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>&gt; 150</td>
<td></td>
<td>c. 220</td>
</tr>
<tr>
<td>Nilo-Saharan</td>
<td></td>
<td>60</td>
<td>E/W Africa</td>
<td>20</td>
</tr>
<tr>
<td>Khoisan</td>
<td></td>
<td>10</td>
<td>S Africa</td>
<td>0.2</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>c. 1,250</td>
<td></td>
<td>c. 610</td>
</tr>
</tbody>
</table>

*Linguistic populations and the number of languages are notoriously difficult to state with any accuracy in Africa. The population estimates in C. Moseley and R. Asher, Atlas of the World’s Languages (London, 1994); J. Bendor-Samuel, The Niger-Congo Languages (Lanham, 1989); and B. Heine, T. Schadeberg and E. Wolff (eds.), Die Sprachen Afrikas (Hamburg, 1981) are out-of-date to varying degrees. I took these statements and massaged them upwards, but I was still short of The World Factbook (Washington, 1994) figures by over 50 million.*

The usual estimates of the number of languages in Africa (‘some 1500’) are equally problematic, partly because there is no satisfactory definition of the dividing line between language and dialect. Languages are conventionally said to be mutually unintelligible and/or to have a standard form and/or to be relatively dissimilar, while dialects are mutually intelligible, have no standard form, and are relatively similar. While these criteria are impossible to apply in Africa, the problem is exacerbated by the issue of ethnicity. Nearly all African languages, at least south of the Sahara, are tied to ethnic groups. While many Africans are willing to view other languages as dialects, few are willing to see their own as such.

The number of Bantu languages is usually placed between 400 and 600, but extrapolating from my own detailed knowledge of East Africa, I estimate there are slightly more than 300. M. Mann and D. Dalby, A Thesaurus of African Languages (London, 1987), estimate more than 450, but most of their additional languages are demonstrably dialects. If we add the figures in the table, we get at least 1,250 languages total, a number lower than the 1,500 usually quoted for Africa. This is the estimate of a ‘lumper’; a ‘splitter’ might estimate 2,000 or even more.

The Comparative Method in the first place. Rather, individual linguists hypothesized that certain languages or groups belonged together, and only later, sometimes decades later, was the hypothesis ‘scientifically’ confirmed.

This can be seen by considering briefly the history of Africa’s largest family, Niger-Congo. Earlier this century, Westermann established sets of lexical cognates using data from some Niger-Congo languages. They were in fact part of an attempt to establish a larger group, Congo-Saharan (Niger-Congo plus Nilo-Saharan); they did not cover all Niger-Congo; and he did
not provide regular sound correspondences.\textsuperscript{29} Greenberg then applied his method of mass comparison to a large selection of Niger-Congo languages. While he reorganized it, he did not question the unity of Niger-Congo, and he did not provide systematic sound correspondences either.\textsuperscript{20} In the meantime, various scholars produced modified classifications, while others reconstructed vocabulary or sound systems for parts of the family.\textsuperscript{21} Finally, Williamson provided a set of partly typological, partly genetic features for the group.\textsuperscript{22}

While most African linguists today thus accept the notion of Niger-Congo, largely because of the overwhelming general similarities between its members and the very specific and striking noun class system found throughout the family, the fact remains that no one has yet attempted a rigorous demonstration of the genetic unity of Niger-Congo by means of the Comparative Method. Similarly, strict demonstration of the existence of Nilo-Saharan as a genetic grouping is only now replacing thirty years of tacit acceptance by most linguists. Nor is there rigorous proof of the genetic status of Khoisan. The only family reasonably "proved" by conventional methods is Afrasi, and "proof" here is only a statistical probability.\textsuperscript{23}

The second point is that details of Greenberg's classification have changed in the last thirty years, and the affiliation of some marginal languages and sub-groups and the internal configuration of some of the groups are still matters of considerable debate. Thus, for instance, the exact border between Bantu and Bantoid (two sub-groups of Niger-Congo), the internal configuration of Bantu, the status of Kordofanian and parts of Niger-Congo and of course the status of Khoisan are all subject to dispute.

\textit{Caveats and pitfalls of linguistic classification}

\textit{Limits to time depth}

Currently, comparative linguists are mired in controversy over whether many of the world's language families can be further reduced to a few larger phyla and maybe to a single super phylum.\textsuperscript{24} Most linguists on both sides of the argument would like to be able to reduce today's families to a smaller

\begin{itemize}
\item[D. Westernmann, \textit{Die Südansprachen} (Hamburg, 1911) and \textit{Die Westlichen Südansprachen und Ihre Beziehungen zum Bantu} (Berlin, 1927).]
\item[Greenberg, 'Genetic relationship among languages', in Greenberg, \textit{Languages of Africa}, 35-43.]
\item[Williamson, 'Niger-Congo overview'.]
\item[A. Bomhard and J. Keräse, \textit{The Nostratic Macrofamily} (Berlin, 1994); C. Renfrew, 'World linguistic diversity', \textit{Scientific American} (Jan. 1994), 116-23.]
\end{itemize}
number, but they are at loggerheads about whether it is scientifically feasible. Central to this argument is the issue of methodology and data, which affects work in Africa also. Languages are assigned to language families by the Comparative Method. This relies on finding regular and frequent phonetic correspondences between sets of words in different languages, as we have seen. As languages pass through time, however, words are lost and replaced; phonetic changes pile up and each new change makes the effects of the previous ones harder to identify. The longer two or more languages or families diverge from a common ancestor, the greater the lexical loss and the more opaque the accumulation of phonetic change. Thus, many linguists would say that the loss of evidence is so great beyond a certain point that trying to reconstruct beyond it is as much an act of faith as of method.\textsuperscript{25}

The language family most affected by this is Khoisan. Is Khoisan a genetic grouping or just a typological construct? If genetic, then it is a family of related languages which have grown so far apart over thousands of years that it is hard to prove their common ancestry. And if it is a typological construct, then it is simply a label for a group of disparate languages which have been adjacent for so long that they have grown together and are now characterized by a set of shared features such as the famous clicks. No one has yet demonstrated, to the satisfaction of other linguists, the existence of the regular and frequent phonetic correspondences required to prove the genetic relatedness of the Khoisan languages, yet most linguists who have thought about it lean toward accepting Khoisan as a genetic language grouping. Nor is archaeology much help. While skeletons widely agreed to be morphologically Khoisan can be dated to 10,000 B.P. in southern Africa, anatomically modern humans date to 100,000 B.P. in the same region, and there is no consensus that the latter are the direct ancestors of the former.\textsuperscript{26} So the earliest antecedents of the people likely to have spoken Khoisan are unclear.

Family tree diagrams: what do they represent?

Linguists use family tree diagrams as a form of pictorial shorthand to illustrate the relationship and development of languages and groups of languages within families. Each node on a language family tree is supposed to represent the starting point of an innovation or set of innovations shared by all the languages below the node. Underlying this is the assumption that if an identical change has occurred in two or more related languages – especially if they are geographically close – that change is likely to have occurred only once rather than several times independently. The innovation or set of innovations are what defines these languages as a group vis-à-vis other languages. The most useful innovations are phonological or morphological as these elements are relatively stable and slow to change, while lexical innovations are less satisfactory because words spread so easily and quickly across language boundaries that they rarely remain specific to one

\textsuperscript{25} For a bold proposal to deal with this problem, see I. Nichols, Language Diversity in Space and Time (Chicago, 1992).

\textsuperscript{26} See G. Brauer and F. Rosing, Human Biological History of Southern Africa (Munich, 1986); A. Morris, "Biological relationships between Upper Pleistocene and Holocene populations in Southern Africa", in G. Brauer and F. Smith (eds.), Continuity or Replacement (Rotterdam, 1992), 131-43.
community for long. However, an entire set of new words is also a useful innovation when used carefully; much of Ehret’s work involving subclassification is based on such sets of lexical innovations or shared loanwords.\footnote{See, for example, C. Ehret, \textit{Southern Nilotic History} (Evanston, 1971), \textit{Ethiopians and East Africans} (Nairobi, 1974), \textit{The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary} (Berlin, 1986) and ‘Proto-Cushitic reconstruction’, \textit{SUGHA}, \text{vii} (1987), 7–184.}

An innovation starts at some point and spreads (or doesn’t spread) across part of a language community. Changes rarely spread across entire communities all at once because at some point they meet some kind of resistance, based on geographical, social, age or other factors. Thus a language community A develops two variants, a new variant B and the unchanging variant C. At this point B and C are ‘dialects’ of A. This is already a misrepresentation because C is just the remnant of A. Later, other changes start. As long as social, geographical and age factors have not changed, there is a possibility that these changes might spread to cover exactly the same area as the first, but this is not often the case as changes are not usually simultaneous or coextensive. Thus, as successive changes spread across different parts of B and C, they become crisscrossed by what we call isoglosses. At some point, B and C become so different from the original A and from each other that we call them each languages, a useful label but a fiction because linguists do not agree how different one dialect needs to be from another before it becomes a language. Hence the disagreements over the number of African languages and of how well family trees represent linguistic reality.

There are thus pitfalls here for anyone interpreting a family tree diagram. First is the danger of conflating linguistic and non-linguistic evidence. Linguists intend trees solely as graphic representations of linguistic phenomena. As such, they should be worked out independently first and only related to other historical events later. Rarely in Africa do we have enough information to link languages unambiguously to historical peoples, much less to historical events, but historians and archaeologists are frequently prone to interpret a linguistic tree as a literal historical development, as a movement of peoples from one point in space or time to another, or as one component of a homogeneous cultural ‘tradition’.

Second, a standard tree diagram makes it appear that languages diverge all at once. A node on a family tree suggests that all the changes that have accumulated in B and C over time took place simultaneously at a single point in time, but such was rarely the case. In fact, since languages are continually changing, a node only represents changes occurring just prior to that point, during that time, or slightly after it. Thus, some of the innovations that have come to characterize a language or sub-group today may have occurred subsequent to its node on the family tree.

Third, there is the temptation to interpret nodes on a family tree as indicating physical movements of peoples away from one another, causing the language they spoke to diverge over time. But does a node signify that language A split into B and C because communities speaking B and/or C drifted apart from one another, or did communities B and/or C separate after the linguistic changes occurred? In fact, it might mean either. Most
linguistic innovations take place in part or all of a language *in situ*. An innovation starts at some point within a language community and rolls across the community and into others. It might stop at a geographical boundary, a social one (e.g. a class dialect) or a linguistic one with another language. Thus, peoples do not need to move for their language to change or be influenced by others.

Alternatively, if speakers of community B do move away permanently, it is likely that language B will undergo further changes that will then come to define it. Geographical separation inevitably leads to linguistic separation when communities no longer need to communicate with one another. But in either of these migratory scenarios, the relative chronology of the linguistic change and the physical move is not clear. The change(s) simply define B as a language and say little about the events that produced it. If B later splits into D and E, they will then inherit the changes defining B, thus forming the basis for grouping D and E together as sub-groups of B.

Fourth, a family tree diagram implies linguistic change in a vacuum, unaffected by adjacent communities. It also implies that all languages are the natural descendants of the line of languages above them, handed on organically from one generation to the next. Neither of these is necessarily true. Few languages evolve in a vacuum, and if communities are in contact long enough, material will pass from one to the other, sometimes in sufficiently large quantities that a language’s original genetic affinity becomes obscured and the language appears to be the descendant of an adoptive ‘parent’ other than its natural one.

Family tree diagrams are thus static models that greatly oversimplify complex patterns of linguistic change. Languages are the result of long and complicated dynamic processes not easily reducible to a single, two dimensional, tree diagram. Tree diagrams can certainly be useful, but they are limited and potentially deceptive.

Trees often contain an additional pitfall, mentioned earlier. Branching nodes should be based on innovation, but since we have often lacked comprehensive linguistic data in Africa, we have constructed trees on the basis of lexicostatistics. This suffers from its own set of drawbacks, including the tendency of lexicostatistically derived trees to reflect retention, not innovation. Readers should thus beware of trees based exclusively on the use of methods such as lexicostatistics. If classifications based on lexicostatistics (or lexical dialectometry) correspond to those based on other criteria, such as phonological or morphological innovation, they may be taken as relatively reliable. But if lexical and non-lexical indices differ, then lexical similarity most often indicates later contact whereas phonological and morphological similarity suggests shared inheritance.

Classifications tend to be incomplete at intermediate levels

One of the most perplexing aspects of African language classification is the...
fact that the top and bottom of family trees are reasonably secure, while the intermediate levels are less sure and are often disputed (see, for example, African language families above). Thus, most comparative linguists working in Africa today agree on the four indigenous language families (see Table 1) as well as on the immediate sub-groups to which most of today’s languages and dialects belong, but there is little agreement on the intermediate levels between these two. This is not just an African problem. Indo-European has been identified as a language family for nearly two centuries, yet for most of that period scholars have not agreed on how its major internal sub-families, such as Germanic, Romance or Slavic, relate to one another.

One reason for the lack of clarity about these intermediate relationships lies in the complicated and interwoven distribution of African language communities. Most of the major movements that led to the general outlines of the linguistic map of Africa occurred millennia ago. In the meanwhile, language communities have splintered and interacted, risen and fallen. As successor languages developed, they became defined by certain innovations, but with the further passage of time, these were overlaid by changes stemming from elsewhere, and some language communities became absorbed linguistically by communities speaking languages of different families. After the passage of so much time, it is hard to untangle the web, to sort out the old from the new and the inherited from the transferred. The kind of detailed detective work necessary to unravel this web has only been carried out in a few places due to a general lack of interest, funding and scholars relative to the magnitude of the task. Hundreds of languages are still not adequately described, and smaller languages carrying a wealth of historical detail are rapidly disappearing.

The lack of clarity about the intermediate levels of family trees has important implications for historians who seek to extrapolate the movement of peoples and ideas from family trees. If the intermediate branching is not clear, then any historical inferences made from it are unlikely to be reliable.

Former populations and substrata

If the intermediate levels of language development everywhere in Africa are unclear, other linguistic populations have disappeared, leaving barely a discernable trace of their former prominence. There is no doubt that languages continually die out, but as long as they came into contact with succeeding languages intensely enough to influence them through transfers, they can often be identified and described through these remnants. Some such languages are well known and have even been reconstructed in some detail. Southern Cushitic-speaking peoples once lived throughout Eastern

---

32 See, for example, Ebert, Southern Nilotic History, Ethiopians and East Africans and Eastern Africa in the Early Iron Age (forthcoming).
Africa, for example, but incoming Bantu-speakers started to absorb them some 2,000 years ago. In the process, however, Southern Cushites deeply influenced Bantu culture, leaving numerous transfers in Bantu languages as testimony to their former existence and influence.

Other former languages were not as influential, however. Pygmies or Pygmoïd populations today stretch nearly 1,000 miles from Cameroon across Zaire to Rwanda. They total some 200,000 people today and are physically, biologically, economically and culturally distinct from their neighbors. Pygmies are well adapted to the forest, popularly thought of as its original inhabitants, and genetic studies suggest they have developed separately there for up to 20,000 years. In spite of such an enduring presence, however, Pygmies have no language of their own. All speak the languages of their more numerous neighbors, usually Bantu or Ubangian, and it seems that they have been in the process of giving up their own languages for over 5,000 years. Moreover, the disparate Pygmy languages share little linguistically that would define them as a separate linguistic community. In fact, tenous linguistic and other evidence indicates that there were three separate linguistic populations in the past. Thus, while it is tempting to think that the Pygmies once had their own language(s) and only gave them up fairly recently in favor of those of their influential neighbors, linguistic evidence hardly supports this or gives us many details of what those original languages might have been.

The main determinants of whether a language lives or dies do not lie in its linguistic structure but rather in local political, economic and sociological circumstances. We shall never know the exact circumstances surrounding past extinctions, but a glance at the contemporary situation in Africa gives us broad clues. Two kinds of languages have political utility and are doing well. One consists of the few languages with national or international use, such as Swahili. The other consists of a larger number of middle-size languages with a regional distribution and use, less than national but larger than very local. The losers are a third group of languages who are probably condemned to extinction because no one outside their own small local communities wants to learn them or because there is a break in linguistic continuity within the communities as younger people are not able or willing to pass on their language to the next generation.

From linguistic classification to history

We have already seen some of the potential pitfalls of casually inferring historical events from linguistic evidence, but there are a number of well-developed techniques for doing so more rigorously. It is easy to identify the members of a language family or sub-family on a map and thus to define its geographical limits today. Within this area, it is usually possible to

---

23 Cavalli-Sforza et al., History and Geography of Human Genes.
26 Earlier languages, now extinct, can also often be identified via loan words, archaeology or oral traditions.
distinguish smaller areas of concentrated linguistic diversity from larger, relatively homogeneous ones. These smaller areas often indicate the older homelands of the family, where long settlement has allowed the most linguistic divergence, while the larger areas have been settled more recently. The histories of Bantu, Nilo-Saharan and Afroasiatic all illustrate this.

Bantu-speaking communities stretch across Africa from west to east and down to South Africa. Within that large area, the most diversity is in the extreme north-western corner of the distribution, while Bantu’s nearest relatives are spoken just to the west of this north-western zone, strongly suggesting that the Benue basin was the original proto-Bantu homeland where Bantu first split from its relatives and began to develop as a separate language. Similarly, although the best known Nilo-Saharan peoples are East African Nilotes like the Maasai, similar geolinguistic considerations suggest these languages originally came from around Lake Chad.35

The evidence for Afroasiatic (formerly known as Hamito-Semitic) is more complex. Previously, its homeland was thought to be in the Middle East because its two best known members (Arabic and Hebrew) are spoken there. Arabic is its largest member and a pervasive Western bias considered the eastern Mediterranean as the cradle of civilization. But the nearest relatives of Semitic, the other sub-groups of Afroasiatic, are dispersed across north, west and north-eastern Africa as far as Tanzania. In fact, of the 250 languages in the family, all except twenty are native to Africa, and the greatest diversity in Afroasiatic languages occurs in north-eastern Africa, thus making it likely that this was their original homeland. Three (Semitic, Cushitic, Omotic) of the six Afroasiatic sub-groups are exclusively or partly represented in this one area, while the other three are all in different places. This conclusion has been controversial, however, because of the aforementioned bias to the Middle East and the fact that archaeology has constantly produced a string of early dates for the Middle East.36

Khoisan illustrates a further pitfall in trying to link linguistic distribution with history when some links are missing. Most of the present Khoisan languages are spoken in Namibia and Botswana, with a smaller number in Angola, but 400 years ago Khoisan communities existed everywhere south of a line from southern Angola to a point near the South African-Mozambique border. Further north, in western Tanzania, we find Sandawe, a Khoisan language, together with Hadza, a language which, if not Khoisan in origin, has certainly been influenced by a Khoisan language. Finally, there are the Dahalo, whose language was similarly influenced by Khoisan, who live in north-east Kenya at the foot of the Tana River. Given that most Khoisan languages are now spoken in southern Africa, with a few outliers far to the north-east, one is tempted to conclude that southern Africa was the homeland.

36 The ‘three’ sub-groups of Afroasiatic in north-east Africa can be expanded to four if Coptic/Egyptian is included and the north-east is expanded somewhat. This paragraph has benefited from comments by Newman. He notes that the distribution of Berber, from cases in Egypt to the Atlantic, is consistent with a migration from east to west and that the present location of Chadic, despite the proliferation of Chadic languages, requires only one move.
from which Khoisan speakers slowly moved north before the community was severed by incoming Bantu communities during the first millennium A.D.

Such a conclusion would be wrong, as illustrated by the comparable Celtic situation. While the Celtic-speaking populations today are small and restricted to the western fringe of Europe, no one points to the British Isles or Normandy as the original homeland of the Celts. Rather, 3,000 years ago Celts covered much of what is now Iberia, France, Germany and central Europe, but were subsequently absorbed or pushed to the fringes by immigrant Slavic, Germanic or Romance speaking populations. Similarly, it is probable that Khoisan-speakers once inhabited the whole area from Kenya to South Africa - perhaps more - and were slowly pushed to the fringes or absorbed into immigrant Bantu-speaking communities over the past 2,000 years. Given this, it is hard to say where a proto-Khoisan homeland might have been.

While we can often locate the origins of language families in space, locating them in time is more difficult. Since languages develop in orderly sequence, linguists usually have no trouble dating linguistic phenomena relative to each other, but they have no accurate way of dating them absolutely or independently of non-linguistic phenomena. Thus nodes on a family tree or linguistic events tend to be dated in one of two ways. One is through glottochronology, as noted earlier. While this is controversial, many linguists tend to use it informally, nonetheless, and then juxtapose their intuited results with dates derived from archaeology.

Family trees do not only culminate in a single node at the top, however, but they also consist of a number of intermediate nodes. The top node represents the original proto-community, which linguists try to locate in space and time, but they can also do the same thing for communities represented by intermediate nodes. Most linguists assume that an original homeland was linked to a series of intermediate homelands from which progressively smaller communities fanned out. Five different models have been developed for exploring the nature of such linkages.

The traditional ‘migration’ model takes lines linking linguistic phenomena on family tree diagrams and interprets them historically in terms of lengthy migrations. That such migrations actually took place is clear. The largest concentration of English speakers today is in North America, and English speaking communities are spread throughout the world. But the greatest diversity within English is in England, and English’s nearest Germanic relatives are all located in north-western Europe. Thus north-western Europe is the most likely candidate for the original homeland, while England is the best candidate for an ‘English’ homeland and intermediate jumping-

---

off point for subsequent communities. Thus two major movements of English-speakers must have ensued, one from the Continent to England, and a second from England across the seas. 

Similarly, Polynesian settlers moved quickly over long distances across the South Pacific in one direction and across the Indian Ocean in the other. Arabic speakers rapidly expanded across North Africa in the second half of the first millennium. And maritime Swahili were able to move rapidly up and down the East African coast. In all these cases, an original large movement was followed by smaller local movements as the initial community slowly dispersed. But while sea-borne peoples or pastoralists can travel long distances quickly, subsistence farmers typically moved much more slowly.

A second model, the so-called 'wave-of-advance' model, appears to suit subsistence farmers and shifting cultivators better. As a family develops, it slowly clears more land and shifts its fields from season to season until no more fertile land is available within the immediate area and some members pioneer new land elsewhere. Over generations, the process is endlessly repeated and hundreds or thousands of kilometers are slowly covered in a process of expansionary drift. If farmers displace local hunter-gatherers in the process, then they often absorb them, leading to more rapid population increase and an increased rate of drift. While this model is appealing, there are two objections. One has to do with the period of time involved. Computer simulations predict that the rate of advance averages a kilometer a year. While this is compatible with the spread of early agriculture from Greece to Scotland, it would predict that Bantu farmers took over two millennia to move from eastern to southern Africa rather than the millennium that they actually took. The other objection is linguistic. A slow forward movement of people would involve the emergence of one linguistic subgroup after another in a constant chain-like movement. In Europe that chain would start with Greek, but no linguistic evidence has been adduced for this kind of link among different Indo-European linguistic groups. Similarly for Bantu, where linguists have had notorious trouble solving the jigsaw puzzle of linguistic pieces, linguistic chaining has not been proved.

A third model, also originally computer-generated, has been termed the 'discontinuous-spread' model. This is a somewhat inaccurate label, however, as it combines continuous waves of advance and local dispersals with some longer distance movements. The 'discontinuous-spread' model is central to Vansina's reinterpretation of much of the Bantu diaspora:

an original East Bantu spread first as a single language to cover much of the large area where its daughter languages are now spoken. This suggests that the spreading took several centuries at least during which the original language remained unified.

The implication was that an original language group first covered a huge area.

---

42. Note the small numbers of original speakers involved in some of these movements.
45. Collett, 'Spread of the Early Iron Age'.
46. Vansina, 'New linguistic evidence'.

relatively quickly and thinly. In time, however, local dialects slowly developed among people geographically separated from their neighbors, producing the present situation. The model is attractive as many areas in sub-Saharan Africa are unsuitable for subsistence farming, so individual groups of farmers have had to hop across ecological barriers from one cultivable area to another, and all recent analyses of Bantu expansion use it implicitly or explicitly. But while it fits the linguistic sub-groupings well enough, correlates with local archaeology are sometimes asserted rather than demonstrated, and the model relies on establishing a prior relationship between ceramic types. Nevertheless, it has the major advantage of accounting much better for the archaeological chronology.

A fourth model, the 'wave' model, involves the spread of linguistic features rather than movements of people. As a cultural, economic or political system became successful, its prestige and its linguistic norms spread to adjacent peoples speaking other languages. Such a pattern, whereby linguistic features radiated from a central point across adjacent areas, is commonly accepted by linguists working in Africa today.

The final model is a more radical version of the preceding one. Here, the influence of one group was so great that an adjacent group gave up its own language entirely. In the standard 'wave' model, linguistic features radiated out across adjacent languages or dialects, whereas in recent treatments whole new languages or even peoples were included. Again, no linguist would deny this possibility, although total replacement of one language by another is a local phenomenon and not as common as some historians imply. And, while larger languages are indeed now displacing smaller ones in Africa, this is a product of modern political circumstances and can not be assumed to have been as frequent in the past.

These dispersal models need not be mutually exclusive. Most dispersals certainly involved the assimilation of earlier populations by later ones, usually of hunter-gatherers by farmers. And some models—especially the 'wave-of-advance' and 'discontinuous-spread' models—were not intended as linguistic models when first proposed and so need to be verified linguistically. It is linguistically plausible to see ethno-linguistic groups splintering, subsistence farmers spreading slowly or quickly over the land, and prestigious languages being gradually adopted by neighbors, who might even give up their original language in the process, but language contact leaves certain telltale linguistic signs, as discussed further below, and the dispersal models sketched above need to be demonstrated linguistically before we can accept them as valid.

Some successes and lessons of linguistic classification

Initial classification and subclassification of related languages, and statements of probability about movements and migrations, are not direct statements about history, but are only a framework and need fleshing out. Reconstructed languages do not come with dates, and reconstructed vocabularies are only

---

47 Renfrew, *Archaeology and Language*, Vansina, 'New linguistic evidence', Fig. 3.
48 See Vansina's fine picture of how this might have worked, Vansina, 'New linguistic evidence'.
probabilities. Current geographical distributions do not explain why people moved. We thus need to bring other sources of data, such as archaeology, to bear if we are to make meaningful historical statements from linguistic evidence. It would be nice if variables such as linguistic affiliation, group identity, ceramic styles, archaeological stratum, technology, and economy overlapped, but experience suggests that the extent of overlap varies considerably and thus needs to be carefully established in each case.10

The problem of Bantu expansion illustrates some of the successes and limitations of juxtaposing linguistic classifications and archaeological data and trying to interpret them historically. Once Bantu languages were recognized as a group and the limits of their geographical distribution plotted, it was only a matter of time before attempts were made to correlate them with archaeological data. A Bantu archaeological profile seemed easy enough to establish because Bantu communities were culturally and economically homogeneous, their farming lifestyle contrasted clearly with that of their non-agricultural predecessors, and ‘Bantu’ pottery seemed to cohere stylistically, but the task has been fraught with difficulties.

Early historical interpretations derived from Guthrie’s work on Bantu languages in the 1950s and 1960s.11 The main problem was linguistic. Guthrie had reconstructed over 2,000 roots for proto-Bantu words, and he calculated that the languages which retained the highest percentage of these words were located in the southern headwaters of the Zaire. He then took this to mean that these languages were the most conservative and thus represented the point from which other Bantu languages had dispersed. Guthrie’s assumption was wrong, however, and Greenberg subsequently demonstrated that the initial Bantu homeland was more likely to lie in south-eastern Nigeria, as we have seen.5 The two linguistic positions were later reconciled by assuming southeastern Nigeria as the initial dispersal area, followed by a series of intermediate dispersal areas to the east and south.52

Subsequent linguistic studies were largely based on lexicostatistics, and historians once more sought to link these with Greenberg’s earlier scenario.53 Again, the resulting historical interpretations proved faulty, largely because parts of the underlying linguistic picture were flawed.54 In retrospect, these

54 Greenburg, *Languages of Africa*, drawing on material published in the 1940s and 1950s in the *Southwestern Journal of Anthropology*.
57 Accessing these interpretations is made more difficult because some of the linguistic data was never published. Likewise, Vanu, ‘New linguistic evidence’, rests on linguistic data not yet published, and Ehret, *Eastern Africa in the Early Iron Age*, makes some claims about the distribution of words without always presenting the raw data. It is exciting to present new hypotheses, but a scientific hypothesis has to be based on data.
interpretations all foundered for three main reasons. First, the arboreal frameworks provided by the linguists, while fine constructs, were not direct models of history and were often incomplete. In particular, these frameworks were not complete or correct at the intermediate levels, and linguists still need to develop better intermediate classifications of Bantu if they are to provide reliable historical models. Second, historians were too eager to interpret the linguistic findings literally without carefully examining what they were based on. All of us, historians, archaeologists and linguists, need to understand more about each other’s discipline, enough to make independent judgements of findings in other disciplines or at least to know what questions to ask so that we do not accept claims uncritically. Third, there were flaws in the archaeology: geographical coverage was spotty; archaeologists did not agree on the relations among certain ceramic types; and correlations between ceramic types and language populations were unclear.

Historians have thus reconsidered these early models and realized that it was not realistic to portray large homogeneous groups of iron- and pottery-using farmers speaking Bantu languages moving boldly across the landscape displacing autochthonous hunters in their wake. The most recent interpretations have replaced a few rapid large migrations by a subtle combination of subsistence drift (“wave-of-advance”), frequent shorter movements over several centuries and local diffusion. While Vansina bases his case on largely unpublished lexicostatistical data, his proposals are entirely linguistically based; he uses the data critically and he hardly refers to archaeological data. By contrast, Ehret’s analysis of Bantu expansion in eastern and southern Africa presents his own data, makes intermediate groupings based largely on lexical innovation and interweaves linguistic and archaeological data.

The Swahili of Africa’s east coast provide another case study of a possibly deceptively neat correlation between linguistic and archaeological data. Swahili settlements developed a distinctive archaeological profile soon after their establishment, and the details of the Swahili dialect/language continuum are also clear. Hence, it has been possible to reconstruct early Swahili history in some detail, based on the fit among linguistic, archaeological and oral data. Since three of the six languages related to Swahili are spoken in north-eastern Kenya, the rule of least moves suggests that the Swahili was the initial dispersal area for all six communities. Further, the largest cluster of large early archaeological sites, all dated to c. 800 A.D., also occurs in north-eastern Kenya. Finally, oral traditions among some of the related communities point to a migration from the same general northern area.

Recent archaeological work has now cast some doubt on this earlier scenario, however. Much of the newer work has been conducted along the

---

that can be seen and tested. Linguists know, however, that editors are loath to publish voluminous linguistic data. Perhaps we can develop an electronic method of making this kind of data available to those who want to examine it.

55 Vansina, ‘New linguistic evidence’.
57 The actual scenario is more complex than sketched here, and includes correlative documentary and ethnographic data as well. D. Nurse and T. Spear, *The Swahili* (Philadelphia, 1985); Nurse and Hienbusch, *Swahili and Sababili*. 
coast and hinterland of northern and central Tanzania, the region where Swahili's more distant linguistic relatives are now located, and it provides a series of dates covering the first eight centuries A.D., thus suggesting that this area may have been settled earlier and served as a dispersal area. Whether it replaces the dispersal area in northeastern Kenya or simply precedes it, however, remains to be seen.

There has been much less success in other cases, however, because the linguistic or archaeological pictures are less clear. In parts of West Africa, most obviously Nigeria, the current linguistic picture is incredibly tangled; new languages are still being 'discovered'; there is little data for many of them; and thus statements about their status can only be considered provisional.

LINGUISTIC RECONSTRUCTION

Lexical reconstruction

The Comparative Method not only groups related languages but also provides a means of reconstructing details of past stages of a language, back to the family's proto-form. Vocabulary, sound systems, morphology, syntax and semantic shifts can all be reconstructed. For historians, the most tangible aspect of reconstruction is vocabulary, because this allows direct access to cultural terms and concepts used by historic populations in the past. The size of reconstructed vocabularies varies according to distance from the present and the quality and quantity of the data underlying the reconstructions. Where our database consists of existing and well described languages, reconstructions of proto-languages in the not-too-distant past can be sizeable and sure. By triangulating backwards from today's language data, we can often put together several thousand lexical items and their referents for proto-languages. The further back we go in the past and the less firm our data base, however, the smaller and more tentative our reconstructions can be.

The size and quality of reconstructions thus varies considerably. Despite recent comparative work by Ehret, the detailed lexical study of Afroasiatic as a whole is not far advanced. Of the six branches, most reconstruction has focused on Semitic and Egyptian and, with the occasional exception such as Ehret's reconstructions of proto-Cushitic, the others have been relatively neglected. Not surprisingly, given the time depths and problems of classification involved, little has been done on the reconstruction of Khoisan. While some suggestions have been offered on shared lexical items, most have been controversial.

While Nilo-Saharan as a grouping was first proposed by Greenberg thirty

54 Syntax: the structure and arrangement of sentences.
55 See, for example, J. Vansina, Paths in the Rainforest (Madison, 1990) or Schoenmakers, 'Early history in Eastern Africa's Great Lakes region' for excellent examples of the use of reconstructed vocabulary in history.
57 Ehret, Historical Reconstruction of Southern Cushitic and 'Proto-Cushitic reconstruction'.
years ago, much of the ensuing work has been concerned with verifying
details of its internal configuration, reconstructing intermediate proto-s, and
describing some of the lesser known languages. Few scholars have been
engaged in this work, and many of the communities live in areas wracked by
civil war. Consequently, we are only now beginning to see substantial lexical
reconstructions and detailed tree structures.  

Niger-Congo is so vast and complex that most efforts to date have been
directed at description and internal classification of individual sub-parts of
the family. Niger-Congo languages are spoken in some of the most linguisti-
cally splintered and complicated areas of the world, especially in southeastern Nigeria. While considerable reconstruction has been carried out
within narrow Bantu, elsewhere reconstruction within Niger-Congo has
been limited in scope and area, and earlier lexical reconstructions for what
would now be called Niger-Congo were not underpinned by rigorous
statements of phonetic correspondence.  

Our most pressing need at this point, in my opinion, is not for further work
at the highest levels, but for reconstruction--lexical and otherwise--at
intermediate levels. Work on this level will provide us with a more secure
foundation for extrapolating backwards to family proto-s, and it will also
provide us with more detail for the events of the last 2,000 to 4,000 years.
Again, few linguists are working on this because it is painstaking and time
consuming and not as prestigious as producing a reconstruction of a whole
family or phylum. Even at the level of family or phylum, progress is painfully
slow. Few scholars are involved (even fewer since the collapse of the USSR,
where there were several dedicated and talented workers), the number of
languages is large and good comparative data is often lacking.

Words-and-things

While reconstructed vocabularies inevitably include general vocabulary or
grammatical items of little interest to historians, terms that refer to social
institutions and activities, the spiritual world, the economy, domestic crops
and animals, food production and preparation, tools and weapons, flora and
fauna are clearly of great significance for historical reconstructions. Not only
do these provide direct evidence for activities and institutions in the past, but
linguists can also determine which words and meanings are inherited, which
inherited but with their meaning changed, and which borrowed from
neighboring cultures, thus allowing historians to also establish their prove-
nance. Historical reconstructions are further enhanced if one has a series of
linguistic stages, each with a reconstructed vocabulary, that allows one to see
changes in these activities and institutions along with the sources of the
innovations.

---

63 See Bender, Nilo-Saharan Languages; Ehret, Comparative Historical Reconstruction
of Nilo-Saharan.
64 Guthrie, Comparative Bantu; A. Meeussen, Bantu Lexical Reconstructions
(Tervuren, 1986).
65 Westermann, Die Sudansprachen und Die Westlichen Sudansprachen.
66 See, for example, Nurse and Himbeusch, Sesehi and Saboki.
67 The overall technique is often referred to as ‘words-and-things’ (‘Woerter und
Sachen’).
There is a noticeable difference between simply reconstructing vocabularies and teasing history out of them, however. Over the past few decades many linguists working in Africa have produced reconstructed vocabularies, but few of these linguists have been able to produce a really coherent picture of the communities using these vocabularies because they lack the training to exploit fully their work historically. It has thus largely been left to historians with linguistic training to draw the necessary conclusions. Most linguists are open to suggestions from historians about which semantic fields would be useful to investigate.

*From reconstruction to history*

While I have treated classification and reconstruction separately for expository purposes, they are often combined in practice. Distinguishing language families, their sub-groups and their geographical distribution provides historians with a peg on which they can hang details. Related languages imply that their communities have a shared history, inheritance, culture, ancestor and point of origin. Reconstruction, particularly lexical, of proto-languages establishes some of these common components and provides details. A reconstructed vocabulary of the appropriate semantic and cultural areas can thus lead to a detailed picture of an earlier community.

The seminal historical works which have relied heavily on reconstruction of vocabulary share certain common features. They start at some definable point in the history of the community. To establish this point in time and space, they often use glottochronology or archaeologial dates together with the techniques outlined above to determine proto and intermediate homelands. They then rely on the 'words-and-things' technique to establish cultural, social, political and economic features of the community, based on the presumption that a word in a language has a referent in real life. As the community develops historically, it comes into contact with others and separates into successor communities, processes that can be derived from linguistic events represented in a tree diagram. These developments can be further fleshed out by looking at which vocabulary was kept and which was discarded, at changes in the meanings of inherited items, at internal innovations, at new regional words and meanings, at how words dispersed across boundaries, and at how new words entered the language. Most of these accounts finish at some point in the present millennium as soon as evidence from written records or oral traditions becomes available.


69 See, for example, works cited above. It is noteworthy that most broad history (i.e. not local) based on the use of linguistics has been done for east, central, and southern Africa, not for north or west Africa. Vansina (pers. comm.) suggests that this may be because historians in north and west Africa had access to more written records and did not therefore feel a need to diversify their sources.
There are obviously differences among these works, brought about by differences in local conditions and the writers' particular interests. Rainforests preserve few archaeological remains, so Vansina relied little on archaeology in his study of Western Bantu speakers. Conversely, earlier inhabitants of the rainforest gave up their original languages for those of later arrivals or the desiccation of the Sahara forced its former inhabitants out, making it harder to rely on loan word evidence in such areas than in large areas of west, east, and south Africa where source languages for these loans often still exist. In the absence of data from archaeology or loan words, linguists rely more on reconstructions and recent patterns in current languages.

At the risk of stating the obvious, a final caveat should be issued about relying too heavily on linguistic reconstructions. Ancestral English and German-speaking communities were adjacent on the north German plains 2,000 years ago. For a short period after the Anglo-Saxons moved to Britain, their culture presumably remained similar to that they had left behind. Examination of common roots and institutions thus provide a good index for that period, but are of more limited use as a tool as the English moved farther from their Germanic origins. Similarly, one would not want to base a history of France primarily or exclusively on the language spoken in what is now France. When the Roman legions first crossed the Alps, they entered a country where most people spoke Celtic and some spoke Basque. Within a few centuries, though, almost the entire population had changed their language (and much of their culture) to late Latin, now French. It would thus be as risky to analyze contemporary French history and culture in terms of Celtic or Basque as it would to reconstruct pre-Roman history in terms of Latin or French.

**New Contact Models**

Significant new models of language contact have arisen during the last few decades that have dramatically changed what linguists and historians can learn from transferred material. A parody of traditional historical linguistic statements might sound like this: 'Language X was first spoken in X-land but then moved east, where it came in contact with Language Y and borrowed some vocabulary for pastoralism.' The obvious problem with this statement is that it omits any reference to people or societies and treats languages as though they acted independently of their communities. Statements like these are also ambiguous: did the community speaking Language X move east with their language, or was their language adopted by people in the east? A less obvious problem lies in the phrase 'borrowed some vocabulary'. Traditionally, historical linguistics recognized only one form of contact, 'borrowing'. What was seen as 'borrowed' was largely vocabulary - loanwords - which was then used as the basis for statements about contact.

This is not meant to belittle the method. Separating inherited from borrowed vocabulary has played an important role in deciphering prehistory.

---

70 Vansina, *Paths in the Rainforests*.

71 'Borrowing' is defined as 'the incorporation of foreign features into a group's native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features'. S. Thomason and T. Kaufman, *Language Contact, Creolization, and Genetic Linguistics* (Berkeley, 1988), 37. Loan words or loan material are what are borrowed.
and will continue to do so, but our methods have become more sophisticated in their ability to indicate different kinds of ‘borrowing’ and related historical events. Since the term ‘borrowing’ is closely linked to a specific contact model (see below), I will replace it as a general term to refer to any form of transfer of linguistic material under any circumstances by the neutral term ‘transfer’, of which ‘borrowing’ is but one sub-type.72

The last thirty years or so have seen a marked shift in general attitudes towards linguistic change. Formerly, it was only thought possible to see change by looking at a long period of time, usually centuries or more. By comparing the start and the end of the period, change could be established. It was thus impossible to see change actually taking place, or, if it were possible, certainly impossible to discern its direction. But starting with Labov, sociolinguists in several countries have now conducted surveys of ongoing change in time and space, and they have started to measure change, mainly phonetic, in progress and identify directions in it. They have linked these changes to details of socio-economic status, social change, and attitude to language and other factors affecting language. They have established certain principles of phonetic and phonological change, and of the social motivation for change. Finally, Labov and others have applied their findings from synchronic dialect geography, sociolinguistics, and phonetics to historical and comparative linguistics.73 While such methods cannot be applied directly to historical situations 500 or 5,000 years ago, it should be possible to deduce unknown social situations from known linguistic facts by assuming that principles of linguistic change have not changed significantly over the centuries.

At the same time, sociolinguists and historical linguists have been finding an increasing number of cases where features other than vocabulary have been transferred as well as cases in which transfer did not derive from simple borrowing or adjacency.74 We now have evidence for communities giving up their language entirely for another, for situations that produce pidgins and creoles, and for situations that sit uneasily and undefined somewhere between or outside these models.75 Thus ‘borrowing’ has now been

---

72 Several people have urged me not to substitute ‘transfer’ for ‘borrowing’ because the latter is so well established and familiar. That attitude would lead to our never altering models.
74 E.g. P. Bakker and M. Mous (eds.), Mixed Languages (Amsterdam, 1994).
augmented with 'language shift' and 'pidginization' as contact models. Each of these situations is typically characterized by distinctive linguistic features, but since the models are not agreed on, we are still some distance from achieving reliable results.\textsuperscript{76}

\textit{Borrowing}

'Borrowing' occurs most often between languages spoken by adjacent communities. How much borrowing occurs is the result of factors such as the length of the adjacency, the relative size of the communities, their relative power and prestige, the typological fit of the languages, and the extent of bilingualism. Vocabulary is the first item to be borrowed, followed by sound units, phonetic habits, phonological processes, derivational morphology, inflectional morphology and details of syntax, in roughly that order. Within vocabulary, cultural items are borrowed before core vocabulary, non-grammatical items before grammatical ones, nouns before verbs, and so on. Thus, by examining the range and scale of items borrowed, it is possible to access the nature and intensity of the contact situation in addition to the specific words/items borrowed.\textsuperscript{77}

\textit{Language shift}

In 'language shift', a community or part of a community gives up its language en masse, usually for a language in which they have long been bilingual.\textsuperscript{78} People making the shift thus already have access to most of the vocabulary of the new language, so they do not have to borrow vocabulary.\textsuperscript{79} Rather, people adopting a new language retain ingrained phonological, morphological and syntactic habits and carry them over into the new language just as second language speakers characteristically import distinctive accents, word use and grammatical structures into their second language. Thus the order of linguistic events in cases of language shift is opposite to that in borrowing, with changes in phonology, morphology and syntax preceding those in vocabulary. If the new speakers then manage to persuade their host to adopt some of these changes, we talk of shift-induced

\textsuperscript{76} For more details, see Thomason and Kaufman, \textit{Language Contact}, of which this is largely a summary. For a different interpretation, see T. Vennemann, 'Etymologische Beziehungen im Alten Europa', \textit{Der Gießener Bauern}, 231 (1998), 42-44.


\textsuperscript{78} Most cases of language shift are not made by communities giving up their language en masse as the result of conscious decision. Rather they are made by individuals, families, or small groups of people, and typically they leave little linguistic trace. There are, however, documented cases of language shift in this century, and given current political trends in Africa, it is likely to happen quite often over the coming decades. We should assume it occurred in the past, too.

\textsuperscript{79} They may, in fact, already have borrowed considerable vocabulary into their original language during the period of bilingualism that normally precedes shift.
change. The circumstances under which people are encouraged to abandon their own language for another limit the likelihood that it will be found very widely, and a shift can usually only be detected if a whole community shifts together, stays together as a cohesive entity for a period of time after the shift and manages to persuade the host community to adopt their norms.

**Pidginization**

In contrast to 'borrowing' and 'shift', pidgins are new, somewhat simplified languages which arise when numbers of people from different language backgrounds are thrown together and negotiate a new language. Pidgins and creoles result from particular social circumstances found in the west coast of Africa, the Caribbean and Papua New Guinea, and while they are not limited to these areas, they do not appear to be widespread in world history. The end result depends on a number of variables, including linguistic universals, the features of the participating languages, the population of each community and access to the dominant language. The process is a complicated one, but 'pidginization' has its own distinctive linguistic profile that enables us to distinguish it from the other cases of contact.

**From contact models to historical models**

While in theory one can distinguish these different forms of transfer and so make refined deductions on how they came about historically, it is not always easy to do so in practice. Transfers may be discernible shortly after the event, but as time passes intrusions are smoothed out and become less obvious. It is also difficult because linguistic features characterizing these processes overlap to some extent. Finally, it is difficult because of the wide range and complexity of linguistic and non-linguistic variables that influence transfer. Linguistic variables include the extent of bilingualism, the typological fit between the languages and the reliability of our transfer models, while non-linguistic variables include the relative size, power and prestige of the superstrate and substrate communities, the likely length of contact, and such slippery issues as popular attitudes toward the languages involved, the relative status of their speakers and the nature of their coexistence.

---


82 For more detailed discussion, see Thomason and Kaufman, *Language Contact*, ch. 7. Certain African languages can be easily demonstrated to be simplified in some sense (e.g. Songhai, Swahili), but unless this can be clearly linked to use in a specific set of historical circumstances, they can not be assumed to be pidgins.
While our current linguistic models are still inadequate to the task, this is a growth area, and it promises to greatly expand what we can deduce about historical situations from the nature and scale of linguistic transfers that took place. Further progress depends on careful study of what linguistic features are transferred and under what circumstances. Since a description of the circumstances is usually not available for historical situations, models will have to be constructed on the basis of living languages or languages whose circumstances are known and then applied to historical situations in Africa. Our degree of success is likely to be directly proportional to the time from the present. For current languages we have much more real data available for inspection, while data for older periods is often sparse and reconstructed. This is particularly vital for components of language other than vocabulary. A sound system or a tense system for an older stage in Africa is bound to be a reconstruction, and the further back we go in time, the less reliable it is likely to be.

Nevertheless, impressive results have already been achieved, even for the more distant past. Starting fifty years ago, Leslau described Cushitic features that have penetrated the Semitic languages of Ethiopia (e.g. Amharic), presumably during the first millennium of our era. Contact induced such significant changes in the vocabulary, sound system, noun, verb and syntax of the verb that someone who only knows modern Hebrew or Arabic can only smile at what they find in Amharic. Over the following decades other scholars have expanded our knowledge and refined the analysis. Recently, Thomason and Kaufman have claimed that these induced features in Ethiopian Semitic resulted both from borrowing through contact with neighboring Cushitic communities and from language shift. During this shift, indigenous Cushitic communities gave up their language for Semitic, carrying some of their own linguistic features with them, but their own language died in the process. More recently, biological evidence has pointed to such massive gene flow between these communities that it is now difficult to distinguish between them.

Most cases of historical contact in Africa that lie far in the past are less well documented. The historical linguistic literature is full of case studies where the only linguistic evidence is a list of loanwords, and the non-linguistic evidence, if any, consists only of a vague folk memory of a former population. Blood tests have not been done; there has been little archaeological work; and the former linguistic population has vanished. In these circumstances, it is hard to produce a meaningful interpretation.

Even well substantiated cases from the more recent past are often hard to decipher. In south-eastern Africa, Nguni-speaking peoples settled among indigenous Khoisan communities early in the first millennium A.D. Several centuries of close contact ensued, during which the Nguni languages absorbed click sounds, vocabulary and possibly other phonological features

84 Cavalli-Sforza et al., History and Geography of Human Genes. Another well documented case of the correlation between political and linguistic events is Songhai, where massive linguistic changes took place during the period of the old Songhai Empire. See Nicolai, Parentes Linguistiques.
from Khoisan, while somewhat later many Khoisan-speakers shifted to Nguni languages. We still do not know the exact circumstances under which these linguistic transfers took place. An early hypothesis was that Nguni men took Khoisan wives, who then influenced their children to adopt Khoisan linguistic features. More recent analyses reject this in favor of an explanation based on *hlonipa*, a form of taboo practiced by the Nguni.\(^6\) Linguistically, this took the form of avoiding speaking the name of fathers-in-law and other male relatives by marriage. Thus people began to substitute other words, sounds and syllables, some from Khoisan, for the taboo names. Whatever the real situation, it is obvious that extensive bilingualism and close contact must have been involved, the intensity of contact indicated by gene flow from Khoisan to Nguni.\(^7\)

Recent examples from the later half of this millennium are more frequent, and they often serve as valuable case studies for building more sophisticated models for contact, as discussed above.\(^8\) It should be clear from this section that the old emphasis on the primacy of internal change has altered and that interest in contact is increasing. All languages consist of inherited linguistic material, subsequent internal changes and changes induced by contact with other languages or dialects, and fuller analysis of all these processes is likely to enhance the usefulness of linguistic analysis for historians considerably.

**LINGUISTIC CHANGE**

Linguists and non-linguists differ in their views of linguistic change. In my own experience, most non-linguists think that most linguistic change is induced by outside contact. When they think of change, they naturally focus on vocabulary as reflective of culture, and much lexical change does indeed come from outside. Most linguists view this quite differently, however, because they think of lexical change as trivial and uninteresting. At the 1995 meeting of the International Conference on Historical Linguistics, half of the 120 presentations had to do with change in some form, but very few dealt with lexical change or change induced by outside contact. Rather, linguists are currently interested in change resulting from internal factors, such as imbalances in the system or universal linguistic factors. They concentrate on phonetic and phonological change, morphological change, syntactic change and interactions among these. While these types of changes can be induced by contact with other languages, linguists tend instinctively to look for internal reasons. It is not obvious to most of them, for example, that change in the tense system of the verb might be related to the social or economic development of a community, nor is it obvious that such a change might be brought about by contact with another language. Even if it were obvious,


\(^8\) See studies of borrowing, shift and mixed cases cited above. An entire forthcoming issue of *SUGIA* is devoted to the topic of language contact and describes a number of other cases.
what strikes them as more interesting is not the source of the innovation, but how the existing system was restructured as a result of it. Even for sociolinguists like Labov, social factors are linked to the spread of change across communities, not the linguistic nature of the change itself.

Parts of this debate on linguistic change will no doubt produce interesting results for historians in the long run, but these contemporary interests have not yet washed over older interests in classification, reconstruction and contact. If and when it does it will result in more reliable classification because we will be able to rely more on morphological and syntactic indices of relatedness and not just vocabulary and sound. It will also produce better reconstructions because we will be able to reconstruct more components of proto-languages. And we will understand more about contact because we will see a greater range of phenomena that can be transferred from one language to another.

**OVERVIEW**

Linguistics is more relevant to African historians for certain time periods than for others. Linguistic development can be represented arborescantly; conventional linguistic trees, like baobabs, appear upside down with their roots in the air. The lowest nodes and branches on our family trees are quite recent, because they represent the emergence of today’s languages and dialects. Comparative linguistics spans the whole period from these back to the roots of the four African families (Niger-Congo, Nilo-Saharan, Afrasian, Khoisan), while historical linguistics covers development from the roots down to yesterday. Historical and comparative linguistics can thus span many thousands of years. At the far end of the historical and linguistic continuum before the clear emergence of today’s language families, the predictive power of historical and comparative linguistics is very low, although no linguist would deny the earlier presence of language.89

The linguistic picture starts to come into better focus at a second stage, with the clear emergence of today’s four African language families. The proto-forms of Nilo-Saharan and Niger-Congo can safely be assigned to a period before 10,000 B.P., or even 15,000 to 20,000 B.P. Khoisan-speaking populations were certainly present in southern Africa at least 10,000 years ago, maybe earlier. The age conventionally assigned to proto-Afrasian is 8,000 B.P., although Ehret feels a considerably earlier date is appropriate.90 Classification of these proto is more advanced now. The quantity of reconstruction varies from family to family. Relatively little has been done for Khoisan (if it is a family) or Niger-Congo, although the work of Guthrie and Meeussen for Bantu, now thirty years old, is invaluable. The last two decades have seen increased reconstruction of Afrasian and Nilo-Saharan. The quality and scope of work being produced today is better than that in earlier decades, and what is being done will provide a sounder platform both for more detailed investigation of events during this second stage and for pushing more reliably into the further past.

89 As mentioned earlier, some linguists want to go beyond today’s language families to a few reconstructed super families. Little has been reconstructed for the proto-language(s) of these super families, and it is difficult to link them to real places and events, so they are of little use to historians at present.

90 Ehret (pers. comm.).
An even better focus is provided at the third stage, from 5,000 B.P. to the early second millennium A.D., the period covering the development of today's sub-families and individual languages. At this stage, it has been easier to link classificatory trees and reconstructions to results from other disciplines and to produce detailed and realistic scenarios. Recent synthetic works are impressive and would have been impossible a few decades ago.

At the fourth and final stage, the nearest end of the continuum covering recent centuries, less spectacular results have been achieved. Linguistics can certainly be of use here and our data is comparatively rich, but most historical linguists apparently feel that earlier periods are more challenging, and the recent period is better covered by oral traditions, written accounts and archaeology.

Since the 1960s new interests and methods have arisen. Just as historians now have more diversity of evidence available to them, so historical linguists have had to incorporate initiatives from other branches of linguistics. While recent work in phonological and syntactic change has started to affect thinking in historical linguistics, the most obvious impact is the range of new contact models deriving from the work of sociolinguists. Thus Labov's work is that of a sociolinguist with an interest in historical linguistics, while Thomason and Kaufman are historical linguists endeavoring to assimilate sociolinguistic insights.

An initial reading of much recent sociolinguistic work could easily cause a non-linguist to wonder about its relevance, because it is avowedly contemporary and based on case studies of urban situations, while most African history occurred in rural settings a long time ago. It is also based on analyses of the relationship between linguistic and social factors in the present and near past. This does not really matter, however, because linguistic processes are universal, as are the general relationships between linguistic events and social situations. Once we have assimilated the lessons about the relationship between social situations and linguistic results, we should be able to project these back into past situations where only the linguistic results are visible, and thus make better deductions about the possible historical scenarios that led to them. They are therefore likely to be most useful in situations and periods from 5,000 B.P. to the recent past for which we have ready access to direct unreconstructed linguistic data.

SUMMARY

This essay discusses the methods and achievements of historical and comparative linguistics and introduces some new possibilities of potential interest to historians. Its main components are language classification, linguistic reconstruction of past stages and new models of language contact and change. While many historians are familiar with lexicostatistics, most comparative linguists working in Africa today use the longer and better established Comparative Method. This aims to establish relatedness between languages and between them and an ancestor language. It provides historians with a taxonomic framework to test against frameworks from other disciplines. There is no necessary one-to-one relationship between the history of a language and the history of the people who speak it today, nor between a family tree diagram and migrations.

The Comparative Method also aims at reconstructing ancestral languages from which later languages derive. While linguists are able to reconstruct considerable
parts of ancestral languages, reconstructed vocabulary for key cultural, economic, political and societal concepts is the most useful component for historians. Reconstruction and classification have been of most use so far for events from 5000 BCE to the recent past.

Finally, linguists have recently been inspired by new case studies and by theoretical work outside historical linguistics. The traditional model, with largely static language communities affected by 'borrowing' deriving from lengthy geographical proximity, is being supplemented by a range of different contact types and of ways in which linguistic material is transferred. And the view that most linguistic change is internal, with other superficial components such as vocabulary being affected from outside, is being replaced by a model that views internal and external change as both important and views linguistic history as an account of the interplay between stasis and change.