

BEFORE THE DAWN

Recovering
the Lost History of
Our Ancestors

NICHOLAS WADE



PENGUIN BOOKS

LANGUAGE

It may be worth while to illustrate this view of classification, by taking the case of languages. If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, had to be included, such an arrangement would, I think, be the only possible one. Yet it might be that some very ancient language had altered little, and had given rise to few new languages, whilst others (owing to the spreading and subsequent isolation and states of civilisation of the several races, descended from a common race) had altered much, and had given rise to many new languages and dialects. The various degrees of difference in the languages from the same stock, would have to be expressed by groups subordinate to groups; but the proper or even only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and modern, by the closest affinities, and would give the filiation and origin of each tongue.

CHARLES DARWIN, ON THE ORIGIN OF SPECIES

ACROSS THE WORLD, linguists estimate, there are some 6,000 different languages. All are descendants of older languages that are no longer spoken. In a few cases these parent languages have survived in written form, like Latin, or can be reconstructed from their descendants, like proto-Indo-European, the inferred ancestor of a vast family of languages spoken from Europe to India.

The 6,000 languages, in other words, are not an unrelated miscellany but all belong to various branches of a single family tree of human languages. Those branches must presumably have converged at their trunk to a single language, the first ever spoken, which was perhaps the mother tongue of the ancestral human population.

If so, it should be possible to draw up a genealogy of the world's languages, showing their tree of descent from the mother tongue. As Darwin perceived, such a tree should be recognizably similar to a parallel tree showing the emergence of human races from the ancestral population. And if a tree of language could be interwoven with a genetic tree of human populations, and the two trees linked to the various cultures discovered by archaeologists, a new and unified framework would be created for understanding all of human prehistory.

One immediate obstacle to this grand synthesis is that most historical linguists believe language trees cannot be constructed farther back than a mere 5,000 years from the present, or perhaps 10,000 years at most. Geneticists, however, are not so pessimistic. They have developed sophisticated statistical techniques for constructing genetic trees and believe the same approach should work for languages.

The geneticists' methods, if they work, may help resolve several long-running disputes in historical linguistics. Foremost among these is the question of the unusual distribution of the world's languages.

Language Spread Zones and Mosaic Zones

Across the United States a single dominant language is spoken. New Guinea, by contrast, has some 1,200 languages, a fifth of the world's total, jammed into an area a quarter the size of the continental United States. Why should the linguistic situations be so different?

Linguists call a large area dominated by a single language a spread zone. An area parceled into many small regions, each of which has its own language, is a mosaic zone. Most of the world's language zones fall into one or other of these two patterns, and throughout history there seem to have been occasional alternations between them. The forces that generate mosaic zones and spread zones are significant shapers of history and culture.

Mosaic zones arise in part because language mutates so rapidly, even from one generation to another, that in only a few centuries it passes beyond easy recognition. Just six hundred years later, the English of Chaucer seems half way toward a foreign language. Within a language there are dialects, that often change from village to village and were probably even more

distinctive in days when people seldom traveled far from home. Even in England, up until the late 1970s, speakers could be located by their accent to an area as small as 35 miles in diameter.

This variability is extremely puzzling given that a universal, unchanging language would seem to be the most useful form of communication. That language has evolved to be parochial, not universal, is surely no accident. Security would have been far more important to early human societies than ease of communication with outsiders. Given the incessant warfare between early human groups, a highly variable language would have served to exclude outsiders and to identify strangers the moment they opened their mouths. Dialects, writes the evolutionary psychologist Robin Dunbar, are "particularly well designed to act as badges of group membership that allow everyone to identify members of their exchange group; dialects are difficult to learn well, generally have to be learned young, and change sufficiently rapidly that it is possible to identify an individual not just within a locality but also within a generation within that locality."²⁶⁵

In warfare, dialect may serve to distinguish friend from foe. When Jephthah and the men of Gilead defeated the Ephraimites, guards were posted to prevent the survivors escaping back across the Jordan. "And it was so," the bible recounts in chilling detail, "that when these Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said, Nay; Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand."²⁶⁶

On Easter Monday in 1282, the people of Sicily rose up against the occupying French troops of Charles of Anjou. "Every stranger whose accent betrayed him was slaughtered, and several thousand Frenchmen were said to have been killed in a few hours," the historian Denis Mack Smith writes of the massacre known as the Sicilian Vespers.²⁶⁷ The linguistic challenge was to say "ceci" (pronounced "chaychee"), the Italian word for chickpeas.

The mutability of language reflects the dark truth that humans evolved in a savage and dangerous world, in which the deadliest threat came from other human groups. Mosaic zones presumably come into being when small tribal groups coexist for a long time in the same place, with none being able to overrun the others. Even if the original settlers all speak the same lan-

guage, dialects quickly evolve in each group's territory, as a badge of identity and a defense against outsiders. The longer this situation lasts, the greater the diversity of languages that are spoken.

New Guinea, a premier example of a mosaic zone, appears to have so many languages because it has been stable for a very long time. There seem to be two principal language families, Trans Guinea in the central mountains, and Austronesian languages spoken around the coastal plains. Trans Guinea is the language of earlier settlers, possibly even the original ones who arrived 40,000 years ago, while Austronesian is thought to have arrived with rice-growing seafarers who expanded from Taiwan throughout the islands of the Pacific.

Each of New Guinea's languages is spoken, on average, by some 3,000 people living in 10 to 20 villages. Tribal competition, as well as the deeply forested mountains and valleys, is one reason for the extreme balkanization. "Political fragmentation is a fact of life in New Guinea communities," writes William Foley, an expert on the island's languages. "Unlike most of Eurasia and much of Africa, the region does not have a history of state formation, either of empire or nation type. The basic unit of social structure is the clan, and competition between clans is the basic arena in which political life is played out."²⁶⁸ Thus three factors that have shaped the island's rich mosaic of languages are competition, the inability of any one language group to dominate the others, and a long period of time for diversification to occur.

The same process may have occurred on a worldwide basis after modern humans first left the ancestral homeland. Linguistically, a single worldwide spread zone would have been created, because the small group that left Africa presumably spoke a single language. But that spread zone would have been occupied by mutually hostile tribes who deterred travel across their territory by any who didn't speak their tongue. Over the generations this worldwide spread zone would have crystallized into a mosaic zone of increasingly divergent languages. New Guinea and parts of Australia may represent the remnants of that ancient mosaic zone. Given the territoriality of early people, reinforced by language barriers, it is little wonder that the world's population has been so immobile, at least as reflected in its genetic composition, until recent times.

Discovery and exploitation of a new, uninhabited territory would open up a new language spread zone, though that too, once occupied, would

gradually fragment into the mosaic pattern. South America, with its many Amerind-derived languages, is a recently created mosaic zone. But two areas of the world have been inhabited so recently that they still look like spread zones. One is Polynesia, the other is that of the arctic regions, first occupied when the Inuit peoples developed the technology for living there.

Once a spread zone has crystallized into a mosaic zone, what forces can make it revert to a spread zone? Three possibilities are climatic disaster, a transition to agriculture, and warfare.

If a large land area is wiped clean of people, those who recolonize the empty lands will create a spread zone of their own language. The Last Glacial Maximum depopulated the northern part of the Eurasian continent between 20,000 and 15,000 years ago. Those who returned could have been the speakers of the ancient language that preceded proto-Indo-European and other large language families. This postulated ancient superfamily is called Nostratic by some scholars, and proto-Eurasiatic by the linguist Joseph Greenberg. Or possibly it was the Younger Dryas cold snap, beginning around 13,000 years ago, that paved the way for Eurasiatic and its daughter languages.²⁶⁹

Another major perturber of mosaicism may have been agriculture. Colin Renfrew of the University of Cambridge and other archaeologists, such as Peter Bellwood of the Australian National University in Canberra, believe that from each center where agriculture was first developed, populations may have expanded outward, spreading their languages with them.

Bellwood and the geographer Jared Diamond argue that no fewer than 15 major language families are the result of farmers expanding from the first centers of agriculture.²⁷⁰ In some cases a single center spawned several different language families, they suggest. Presumably this could have happened if an agricultural center covered several highly diversified languages in a mosaic zone, all of whose populations were amplified by the new farming technology.

Diamond and Bellwood propose that the center of agriculture in the Near East was the source of at least two major language families. One was the Indo-European family of languages. Another was Afroasiatic, which they say spread southwest into Africa. A third could have been Dravidian which, even before Indo-European, had expanded in a southeasterly direction into India. (Dravidian is distantly related to Elamite, an ancient language spoken in southwestern Iran; the eastern branch of Indo-European presumably arrived in India later, pushing the Dravidian-speakers southward.)

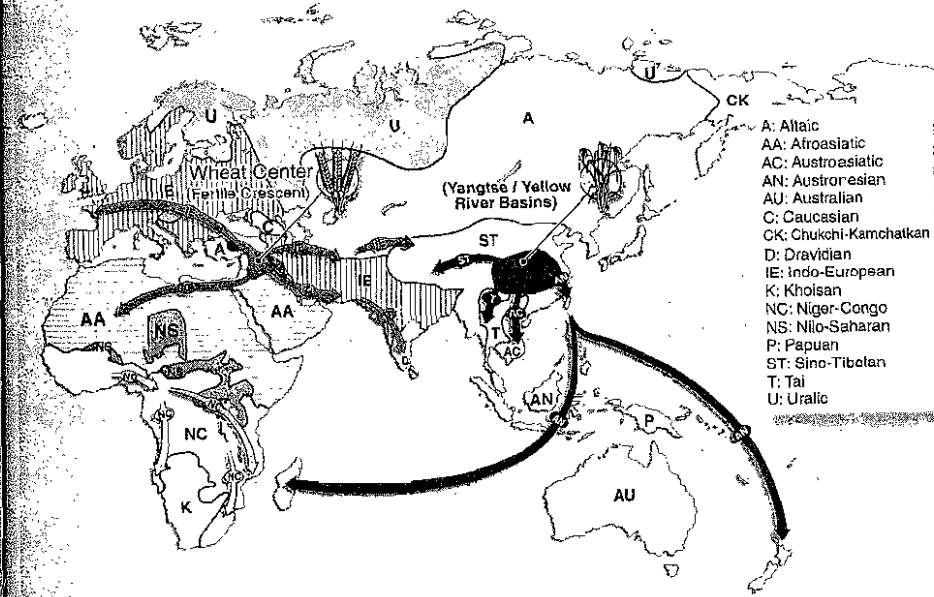


FIGURE 10.1. LARGE LANGUAGE FAMILIES MAY HAVE ARISEN THROUGH FARMING.

The language/farming hypothesis holds that populations expanded from the regions where agriculture was invented, spreading their languages with them. If several languages were spoken within such a region, all could be exported from it. The Indo-European and Afroasiatic languages may have originated in the wheat center, according to the hypothesis, and perhaps Dravidian too. The Sino-Tibetan, Tai and Austroasiatic language families are proposed to have spread from the rice center, along with Austronesian, whose speakers reached Taiwan and from there expanded across the southern oceans.

These language expansions would have taken place up to 9,000 years ago (see arrows). The map of the world, however, shows the distribution of present day language families. People speaking an Indo-European language known as Tokharian expanded into northwest China but their language is now extinct.

Also shown is the Bantu expansion in Africa, labeled for Bantu's Niger-Congo language family, which occurred some 4,000 years ago.

The proposal of the Fertile Crescent as a spawner of language families is ingenious, but the origin of each of the language families involved is a matter of dispute. In the case of Afroasiatic, linguists such as Christopher Ehret, of the University of California, Los Angeles, vigorously dispute Bellwood and Diamond's proposal that the language family originated in the Near East.²⁷¹

A second major homeland of language families, according to the Diamond-Bellwood thesis, was the region of the Yangtze and Yellow river basins where

rice was first cultivated some 9,000 years ago. The rice region, in their view, was the origin of no fewer than four different language families. Speakers of Austroasiatic, a group of 150 languages that includes Vietnamese and Cambodian, spread out to southeast Asia. They were followed by a second wave of rice farmers, speaking the Tai family of languages, which includes Thai and Laotian. Third were the Sino-Tibetan speakers. Fourth were the Austronesians, who reached Taiwan before 5,000 years ago and then set sail across the Pacific, becoming the first inhabitants of Polynesia, and finally reaching New Zealand in around AD 1200.

The Maori colonization of New Zealand was, in a sense, the final step in a 50,000 year journey.

In Africa, the Bantu language family was spread by farmers who developed an agricultural system based at first on yams and later including millet and sorghum. Starting around 4,000 years ago, in their homeland in eastern Nigeria–western Cameroon, the Bantu speakers migrated southward in two migrations. One headed down the west coast, the other crossed to east Africa and then moved south down the east coast. The latter group of migrants mingled with Nilo-Saharan speakers around the Great Lakes region of east Africa, and displaced the Khoisan speakers. Bantu languages, though just one branch of the Niger-Congo superfamily, are now spoken across a broad zone of subequatorial Africa.

Diamond and Bellwood list the Bantu expansion as being the least controversial of their 15 asserted cases of language/farming spread. But a major factor in the Bantu speakers' success, besides their farming practices, was their mastery of ironworking. Iron weapons were part of the package that made their advance through the length and breadth of Africa so irresistible, raising the possibility that warfare was also an agent of the Bantu expansion.

Warfare is a third major perturber of mosaic zones, whether by itself or combined with new agricultural techniques. During the first millennium BC, Nilotic-speaking peoples expanded southward from Ethiopia to the Great Lakes region of eastern Africa, overcoming Cushitic-speaking farmers in the Kenyan highlands. They were able to displace agricultural societies, Christopher Ehret believes, because of a superior military tradition based on assigning young men at adolescence to age sets, which served as military companies on a permanent war footing. "Over the long term of their history, most Nilotes had an institution and apparently an attitude toward war that

recurrently gave them the advantage over all their neighbors, except for other Nilotic peoples, whenever conflict arose," Ehret writes.²⁷² (These southern Nilotes included the Kalenjin of Kenya, now renowned for their more peaceful achievement of dominating world middle-distance running records.)

The Coming of the Indo-Europeans

The Indo-European languages provide a leading test case for whether warfare or agriculture has been the dominant generator of new spread zones. The spread zone of Indo-European stretches from western Europe to the Indian subcontinent. The family includes extinct languages such as Latin, ancient Greek, Hittite and Tokharian, once spoken in northwestern China. The living descendants of proto-Indo-European include, besides English, the other Germanic languages (German, Dutch, Icelandic, Norwegian), the Slavic languages (Russian, Serbo-Croat, Czechoslovak, Polish), the Baltic languages (Latvian, Lithuanian), the Italic languages (Italian, French, Spanish, Portuguese) and the Celtic languages (Breton, Welsh, Irish).

Where was the homeland of the speakers of proto-Indo-European? When did they live? How did they and their language spread? On these questions there exist two main schools of thought, one of which asserts that Indo-European spread by the sword, the other by the plough.

In a series of papers written between 1956 and 1979, the archaeologist Marija Gimbutas identified the Indo-Europeans with the people who built the characteristic burial mounds, called *kurgan* in Russian, in the steppe area to the north of the Black Sea and the Caspian. The Kurgan people, benefiting from the domestication of the horse, started expanding from their homeland sometime after 4000 BC. By 2500 BC, in Gimbutas's estimation, these warrior-pastoralists had reached the extremities of Britain and Scandinavia, and their language developed into its many descendant tongues that are spoken from Europe to India today.

This view is supported on linguistic grounds by Ehret, who argues that if the Indo-Europeans had been peaceful farmers, many words to do with cereals should trace back to them. But Indo-European literatures are full of allusions to fighting. "We find preserved in early myths and legends almost

everywhere among Indo-Europeans a glorification of battle, and particularly of death in battle, not entirely unknown elsewhere in the world, but of an intensity not often matched. We also find widely in these stories a division of society that singles out warriors as an elite group," Ehret says.

A rival hypothesis was proposed in 1987 by the archaeologist Colin Renfrew.²⁷³ He argued that the Indo-Europeans must have been the first farmers, and that they spread out from their homeland because the new agricultural techniques allowed the population to grow and therefore expand. Looking to the archaeological evidence bearing on the spread of agriculture, Renfrew placed the homeland of the first Indo-European speakers in Anatolia, now Turkey, the region where some of the earliest Neolithic settlements have been found. Because the Neolithic revolution started expanding through Europe around 9,500 years ago, Renfrew's hypothesis required the Indo-European languages to have arrived several thousand years earlier than implied by Gimbutas's Kurgan warrior theory and indeed than the date favored by most historical linguists.

It seemed for a time that genetics might decide the issue. The first genetic insight into the peopling of Europe came from Luca Cavalli-Sforza of Stanford University. Working just with the protein products of genes, since DNA sequencing was not then available, he showed there was a genetic gradient, based on 95 genetic markers, that spread across Europe in a southeast to northwest direction. He and the archaeologist Albert Ammerman suggested the gradient was caused by Neolithic farmers moving across Europe in a slow wave of advance. Although the farmers were assumed to intermarry with the existing foragers, giving rise to the observed genetic gradient, the basic engine behind the wave of advance was assumed to be the population growth of the more numerous farmers.²⁷⁴

This idea lent serious but not conclusive weight to Renfrew's theory. Cavalli-Sforza noted that several other genetic gradients emerged from his data besides the one possibly associated with farmers from the Near East. Another gradient suggested a flow of genes westward from the steppe area above the Black Sea. This gradient "supports Gimbutas' hypothesis," he and his coauthors said, just as the first gradient supported Renfrew's.²⁷⁵

New assessments of population numbers have undercut Renfrew's original idea that population growth was the engine of Indo-European expansion. The archaeologist Marek Zvelebil, of the University of Sheffield in

England, writes that "Demographically, there is no evidence for population pressure sufficient to encourage first farmers to migrate, nor is there evidence for rapid population growth. Archaeological evidence does not record rapid saturation of areas colonized by Neolithic farmers, or demographic expansion [with one possible exception]."²⁷⁶

But Renfrew's theory could still be correct even if Indo-European-speaking farmers did not overwhelm the indigenous population of Europe. The farmers' language could have been adopted by the European hunter-gatherers along with the new agricultural technology. In terms of population numbers, relatively few farmers entering Europe from the Near East could have had a catalytic effect in spreading both their language and their farming techniques. Perhaps they bought or captured extra wives from the Paleolithic inhabitants, and the next generation moved a few miles farther into Europe, also adding wives from the existing forager population. The farther this wave of farmers advanced into Europe, the more its Neolithic genes would get diluted with Paleolithic genes. But regardless of the shifting composition of the genetic pool, each generation of farmers would speak the language of its parents' community, presumably Indo-European.

In this way, the new farming techniques would have triggered a language change throughout the area to which they were applied, but with only a small number of Anatolian immigrants relative to the indigenous forager population. This could explain how it is that Europeans speak Indo-European languages yet carry only 20% or less of the genes of those assumed to have introduced the languages.

Can Languages Be Dated?

European genetics seems at present compatible with both theories of Indo-European spread. A more decisive test would be to put a date on when proto-Indo-European was spoken, since the two theories imply very different times of expansion. The Kurgan warrior expansion started some 6,000 years ago, the spread of farming from the Near East some 9,500 years ago.

The dating of languages is not yet a settled science. One approach is to estimate the rate of historical change in a group of languages by analyzing similarities in vocabulary. Glottochronology, one version of this method, de-

depends on estimating the percentage of cognates that two languages have in common. (Cognates are words derived from a common ancestor; *apple* is a cognate of German's *Apfel* but not of French *pomme*.)

The cognates that glottochronologists examine are not chosen randomly but belong to special vocabularies, drawn up by the method's inventor, Morris Swadesh, from items that are particularly resistant to linguistic change. These include words for numbers, pronouns and parts of the body. A Swadesh list of 100 words is the most commonly used.

In comparing two languages, a linguist will decide how many Swadesh-list words in each are true cognates with each other. The fewer cognates, the longer ago the languages diverged, and there are various methods of translating the percentage of matching cognates into a date of language split. In Ehret's view, a 5% match indicates a language split of about 10,000 years ago, a 22% agreement means a divergence around 5,000 years ago, and two languages that parted ways only 500 years ago will retain 86% of their Swadesh-list vocabulary in common.

Given the simplicity of the method, glottochronology can produce surprisingly plausible dates. But it has flaws. Linguists have put considerable effort into criticizing glottochronology, perhaps more than in trying to get it to work better. The result has been continuing disagreement among linguists as to whether it is a usable technique. At a conference held at Cambridge University in 1999, opinion ranged from one extreme to the other. Robert Blust, of the University of Hawaii, gave a paper explaining why the glottochronology kind of method "doesn't work" for Austronesian languages, and James Matisoff, of the University of California, Berkeley, talked about "the uselessness of glottochronology for the subgrouping of Tibeto-Burman." They were followed by Ehret, who explained how well glottochronology works for dating language splits in the Afroasiatic family.²⁷⁷

Historical linguists are much more enthusiastic about a quite different dating technique called linguistic paleontology. The idea is to reconstruct words for objects of material culture in a language family and date the language by noting the times at which such objects first appear in the archaeological record.

In many Indo-European languages, for example, there are words for wheel that are clear cognates of each other. Greek has *kuklos* (a word that is also the origin of circle), Sanskrit *cacras*, Tokharian *kukäl*, and Old English

hweowol (initial "k"s in proto-Indo-European turn to "h" sounds in the Germanic family branch). Since the daughter languages of proto-Indo-European have cognate words for wheel, they must be derived from a common source, and linguists assert that this was the proto-Indo-European word for wheel, which they reconstruct as **k^wek^wlos* (the asterisk indicates a reconstructed word).

Now, the earliest known wheels in the archaeological record date from 3400 BC (5,400 years ago). The proto-Indo-European language must have split into its daughter languages sometime after this date, the argument goes, since how else could the daughter languages, spoken over an enormous region, all have cognate words for wheel?

Similar arguments can be made for words like yoke, axle, and wool. Work on this issue by linguists like Bill Darden of the University of Chicago has encouraged many linguists in their belief that Indo-European was a single language as recently as 5,500 years ago and that its daughter languages could not have come into existence until after this date.²⁷⁸

Linguistic paleontology is an ingenious exercise of the linguist's craft. But it has two conceptual weaknesses. One is that a splendid new invention like the wheel is likely to spread like wildfire from one culture to the next, carrying its own name with it. Linguistic paleontologists claim they can spot such borrowed words. It's true that "Coca-Cola" is easy enough to recognize as a foreign borrowing in many languages, but the more ancient the borrowing, the more a word may take on the coloration of its host language. One of the criticisms linguists level at glottochronology is that it is confounded by unrecognized borrowed words.

Another weakness in linguistic paleontology is the danger of constructing highly plausible words that didn't, in fact, exist. Related words for bishop exist in Greek (*episkopos*), Latin (*episcopus*), Old English (*bisceop*), Spanish (*obispo*) and French (*evêque*), from which the proto-Indo-European word **apispek* for bishop could be reconstructed; but of course, in a language spoken at least 5,000 years ago, no such word existed. As for wheel, proto-Indo-European is thought to have had a word **k^wel*, meaning to turn or twist, of which **k^wek^wlos* is assumed to be a duplication. But it could be that proto-Indo-European had no word for wheel, and what happened was that its daughter languages each independently used their inherited **k^wel*/turn words to form their own words for wheel. In which case proto-Indo-European

could have been spoken thousands of years before the invention of the wheel.

A New Date for Proto-Indo-European

A better, more systematic way of dating languages has long been needed, and biologists hope they may have provided it by adapting one of their own methods for drawing phylogenetic trees. The favored approach is called a maximum likelihood method because it asks what is the most probable shape of tree to account for the observed data. In the case of language families, the data are each language's list of Swadesh words, along with a designation of which are cognates and which are not.

The idea of applying a maximum likelihood method to language history was laid out by Mark Pagel, an evolutionary biologist at the University of Reading in England. Pagel showed that with a list of just 18 words he could generate a maximum likelihood tree for 7 languages (Welsh, Romanian, Spanish, French, German, Dutch and English) that was the same as the tree constructed by linguists with purely linguistic techniques.²⁷⁹

The method has now been further developed by Russell D. Gray, an evolutionary biologist at the University of Auckland in New Zealand. Gray has carefully analyzed the problems of glottochronology and adapted the method so as to address them. One of the problems is unrecognized borrowing. Unrecognized loan words make languages appear younger than they are. But they also knit the side branches of a language together, making a netlike structure. Netlike structures can be tested for and the offending words eliminated.

Another problem that has vexed glottochronology is that languages may evolve at different rates. Both modern Icelandic and Norwegian are known to have evolved from Old Norse, which was spoken between AD 800 and 1050. Norwegian and Old Norse have 81% of their Swadesh list words as cognates, correctly implying a separation of 1,000 years ago. But modern Icelandic, which has been much more isolated, shares 99% of its words with Old Norse, wrongly implying the two languages separated only 200 years ago.²⁸⁰ Rate variation can be taken account of in the maximum likelihood approach, essentially by choosing trees with the minimum amount of variation necessary to fit known dates of language divergence.

The mathematical techniques for addressing both word borrowing and variation in evolution rate were available because biologists had encountered the same two problems in drawing up trees based on DNA data. As with languages, some genes evolve at faster rates than others. And just as words may be borrowed instead of inherited, an organism may acquire genes through borrowing as well as by inheritance; bacteria, for instance, transfer packets of genes to each other, which is why they so quickly acquire genes for resistance to antibiotics.

In one maximum likelihood approach currently favored by biologists, called the Bayesian Markov chain Monte Carlo method, the DNA sequences of various genes are fed into a computer that generates a large number of possible trees by which the genes might be related. The program samples the classes of tree that seem most promising (there are far too many for even the fastest computer to examine each one), and then repeats the whole process a large number of times. At each iteration there are fewer promising trees, and eventually the process will converge on a single, most probable tree to account for the data.

With this powerful tree-drawing technique, Gray and his colleague Quentin Atkinson have constructed a family tree of Indo-European. For data, he relied on a 200 word Swadesh list for 84 Indo-European languages drawn up by the linguist Isidore Dyen, to which he added data from three extinct languages (Hittite and the two versions of Tokharian, known as Tokharian A and B).

Gene trees can often be anchored in real time by matching a date from the fossil record to one of the tree's branch points. The same can be done with maximum likelihood trees constructed for languages. Having found the statistically most likely tree to account for the Indo-European data, Gray then constrained certain branch points in the tree to fit attested historical dates for divergence of certain languages. Hittite must have been a separate language by 1800 BC, the date of the oldest known inscription. Greek must have been separate by 1500 BC, the date of the Linear B inscriptions. Latin and Romanian started to diverge when Roman troops withdrew south of the Danube in AD 270.

Altogether Gray plugged in 14 known dates, constraining the tree to fit itself to the dates in the most statistically probable way. Because the branch lengths of the tree are proportional to elapsed time, anchoring the tree to

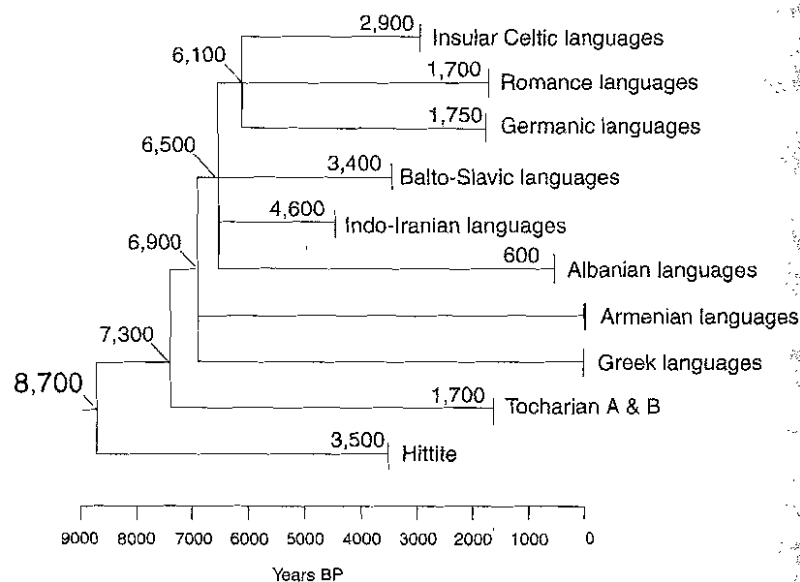


FIGURE 10.2. A GENETICIST'S TREE OF THE INDO-EUROPEAN LANGUAGE FAMILY. A tree of Indo-European was constructed by Russell Gray and Quentin Atkinson using an advanced statistical method. Because the tree is anchored to 14 known dates of recent language origin, the dates of its ancient branch points can be estimated. Figures show the years before the present at which languages split apart.

According to the Gray-Atkinson tree, the original language, called proto-Indo-European by linguists, split 8,700 years ago into the two branches, of which the first led to Hittite and the second to all the other Indo-European languages. The early date assigned to proto-Indo-European suggests that it was the language of the people who introduced farming into Europe from the Middle East.

English is a member of the Germanic group of languages, as are Dutch, Swedish and Icelandic. The Romance language family includes French, Italian and Spanish. Russian, Czech and Lithuanian are among the members of Balto-Slavic. Hittite, now extinct, was the language of the Hittite empire in what is now Turkey; Tokharian was spoken in western China.

historical events allows all the other branch points in the tree to be dated. Gray's tree was published in *Nature* in November 2003, with a terse description of the rather complex methodology behind its construction.²⁸¹ The first reaction of many historical linguists was that he had done nothing new because his tree of Indo-European was just like theirs. But that very fact, in Gray's view, was the best possible validation of his method.

The novel feature of his tree was not its shape but its dates. They were very different from anything the linguists had imagined. The tree showed that

proto-Indo-European was spoken before 8,700 years ago, the date at which it underwent its first split, when the branch leading to Hittite split off from all the rest. This date is nearly 3,000 years older than the 5,500 to 6,000 years ago date favored by many historical linguists for the breakup of Indo-European.

Gray's dates, if correct, are somewhat revolutionary because they show the roots of Indo-European are far older than expected and that language can be traced back far deeper in time than most linguists think likely. Moreover, a reliable dating method would at last allow language change to be correlated with the information emerging from archaeology and population genetics.

Many linguists say Gray's dates can't be right, essentially because they conflict with the dates given by linguistic paleontology. But linguistic paleontology is a fuzzy technique, dependent on judgment and vulnerable to undetected borrowing and fallacious reconstructions. Gray's technique applies a sophisticated statistical method, of proven value in phylogeny, to a reliable data set, the Dyen list, which represents the fruit of Indo-European linguistic scholarship. As a pioneering approach, it may well need refinement, or turn out to have some unexpected flaw. But as compared with linguistic paleontology, it doesn't seem so obviously less credible.

Gray says he has great respect for the scholarship and methods of historical linguistics and hopes linguists will come around to taking his tree seriously, once they understand that his technique avoids the much discussed errors of glottochronology.

Using a simpler phylogenetic technique, Peter Forster, an archaeologist at the University of Cambridge, has drawn up a family tree of several Celtic languages including Gaulish, the version spoken in ancient France before the Roman conquest, as well as Welsh, Breton and Gaelic. Celtic is a major branch of Indo-European. Forster's tree implies that Indo-European had diverged around 10,000 years ago, and that Celtic had split into Gaulish and its British branches by 5,200 years ago.²⁸² These dates have wide margins of error, but are in the same range as Gray's.

Gray's date of 8,700 years ago for the first split in the Indo-European language tree lends considerable weight to the Renfrew hypothesis that the invention of agriculture drove the spread of Indo-European.

The implications reach beyond the specific case of Indo-European. Success of the biologists' tree building methods would mean that languages can

be reconstructed back to 9,000 years ago, considerably farther back in time than many linguists have supposed. The prospects for reconstructing even older trees of human languages may not be entirely hopeless.

The Greenberg Synthesis

Gray's tree building bears on a dispute that has long divided historical linguists. The issue is how best to assess the relationships between today's languages, given that language changes so fast. The world's 6,000 living languages lie at the tips of a long-vanished tree. Can that tree be reconstructed for other families besides Indo-European? Can these families be grouped into superfamilies so as to reach time depths even deeper than that of Indo-European?

The classification of languages is a matter of considerable disagreement. Many linguists, being familiar with the extreme mutability of language, are skeptical of attempts to find ancient relationships between living tongues. Languages change so fast, they believe, that the number of words two diverging languages may share because of a true cognate relationship quickly dwindles to near zero. Indeed, the number of cognates may fall to the same level as the number of word resemblances that arise purely by chance. Unless that point is recognized, the incautious researcher may assert relationships where none exist. The only acceptable way of avoiding such traps, many linguists believe, is with an approach called the comparative method. The comparative method is highly reliable. Its drawback is that it is so rigorous that it does not reach very far back in time.

Other linguists believe that the comparative method is useful for confirming a postulated relationship between two languages, but is too strict to help detect such relationships in the first place. A leading figure in this school of thought has been the late Joseph H. Greenberg of Stanford University. During his lifetime Greenberg classified almost all of the world's languages, showing how they could be grouped into some 14 superfamilies.

The superfamily classifications achieved by Greenberg and his colleague Merritt Ruhlen have been greeted warmly by geneticists because these groupings of languages largely mesh with the population splits inferred from gene-based genealogies. But many linguists repudiate Greenberg's language

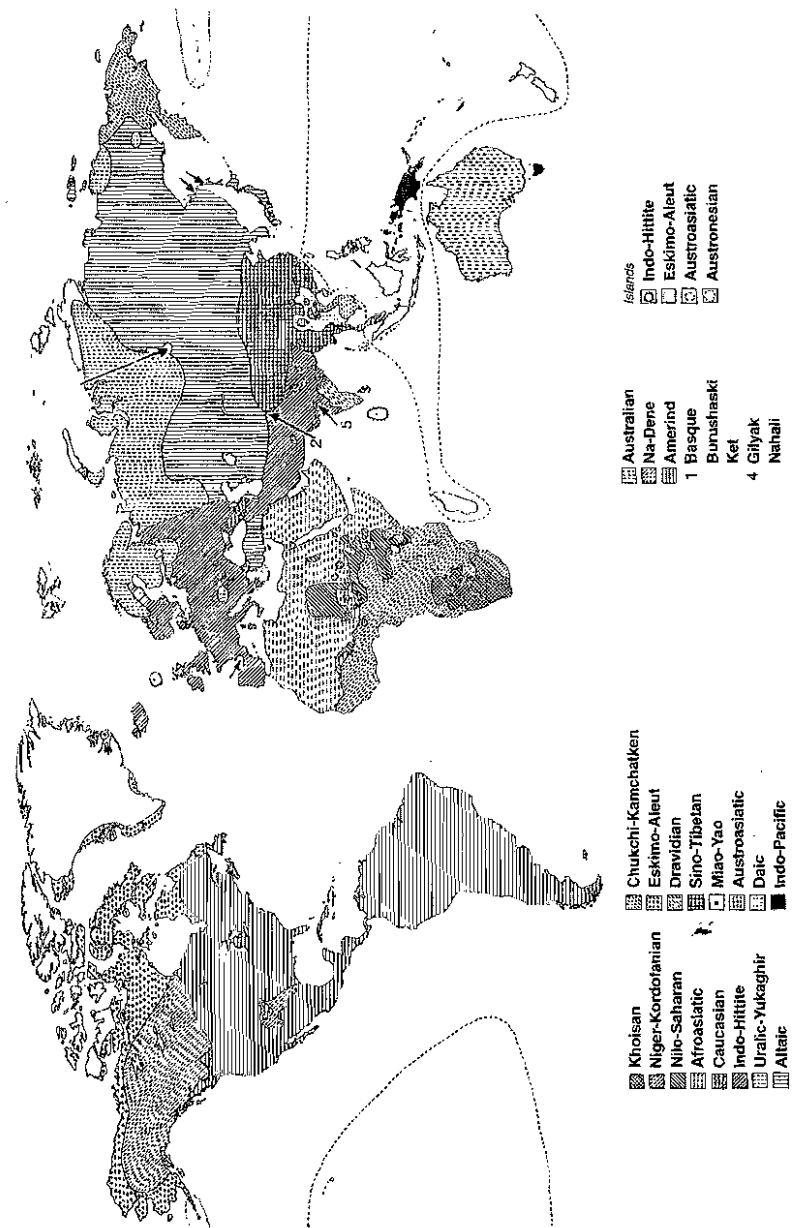


FIGURE 10.3. THE WORLD'S LANGUAGE SUPERFAMILIES.

The language superfamilies of the Old World, as defined by Joseph Greenberg and Merritt Ruhlen. The Basque and Burushaski languages, shown by arrows 1 and 2, are entirely unrelated to their neighbors and may be relicts of more ancient languages. Ket (arrow 3) may be the mother tongue of the Na-Dene languages of North America.

families, arguing that his method is unreliable and that his work contains errors.

Greenberg was not formally an outsider to the linguistic establishment. He served as president of the Linguistic Society of America and was one of the few linguists to have been elected to the National Academy of Sciences. Aside from his work on classification, he founded a subfield of linguistics known as typology, to do with universal patterns of order in the grammatical elements of language. His 1962 article on typology is said to be the most widely cited in the history of linguistics.

Greenberg's training, however, was not in linguistics but social anthropology. He did fieldwork studying the ethnography of pagan cults among the Hausa-speaking people of west Africa, spent the years from 1940 to 1945 in the Army Signal Intelligence Corps, mostly decrypting Italian code, and after the war turned his attention to the interrelationship of African languages.

These had been largely the purview of English and French linguists who had classified them with the help of various criteria, like the physical type of the speakers, that Greenberg deemed irrelevant to language origins. He developed his own, purely linguistic method, which he later called mass comparison. It was based on comparing grammar and some 300 items of vocabulary, such as pronouns and words for parts of the body, that as Swadesh had found are less prone to linguistic change. Greenberg would fill notebooks with lists of languages down the left column and word meanings along the top, and simply search in his mind's eye for relationships.

He started out with Hausa, trying to see what other languages it might be related to by comparing common words and deciding if the languages fell into groups. Over the space of 5 years, Greenberg kept arranging the 1,500 then known languages of Africa into larger and larger assemblies, until he had grouped them into just 16 superfamilies, and finally only four. He put the odd and ancient click languages of southern Africa into the group named Khoisan. The languages of central Africa, including the widely spoken Bantu languages, he assigned to a group he called Niger-Kordofanian. He decided that the Bantu languages must have originated in west Africa, because that is where their diversity is greatest. From that it followed that the present-day Bantu languages, which are distributed down the west and east coasts of Africa, must have arisen from a migration out of the homeland that had split into two streams, one going directly down the west coast, the

other crossing the breadth of Africa and then turning south down the east coast. This inference was later confirmed by archaeologists.

Greenberg's third group was Nilo-Saharan, a family of languages spoken by Nilotic peoples like the Nuer and the Dinka as well as by people of the Saharan region and by the Songhay of west Africa. The fourth group of languages, spoken in a swath across northern Africa, he named Afroasiatic. This family includes Berber of northwestern Africa, ancient Egyptian, and Semitic, a branch to which belong Arabic, Hebrew and Akkadian, the extinct language of the Assyrians and Babylonians.

Greenberg's sweeping classification of African languages has stood the test of time and is broadly accepted, although scholars continue to rearrange the furniture. The African languages are of particular interest because of their diversity and presumed antiquity. At the latest count some 2,035 are now known, of which 35 belong to the Khoisan family, 1,436 to Niger-Congo (a new name for Greenberg's Niger-Kordofanian), 196 to Nilo-Saharan and 371 to Afroasiatic.²⁸³

Ehret has attempted to date the period when the proto-languages of Greenberg's four groups were spoken. On archaeological evidence, he estimates that proto-Khoisan was first spoken about 20,000 years ago. The ancestral tongue of the Niger-Congo family may date back to 15,000 years ago, since a junior branch of the family had spread across the yam growing regions of west Africa from 8,000 years ago. Proto-Nilo-Saharan, on the basis of glottochronology, may be 12,000 years old.²⁸⁴

Afroasiatic is a language family of general interest since its West-Semitic branch includes Hebrew, Aramaic and Arabic, the founding languages of three popular religions. Many people have assumed the ancestral homeland of proto-Afroasiatic was in the Near East, some for a miscellany of unscientific reasons, others because the Near East is a known center of early agriculture from which growing populations might have expanded into Africa, carrying their language with them. But an African origin seems more likely, in Ehret's view. Of the six major branches of Afroasiatic, five lie in Africa—Berber in northwest Africa, Chadic around Lake Chad at the southern edge of the central Sahara, Cushitic in the Horn of Africa, Omotic in the Ethiopian highlands, and ancient Egyptian.

Following the rule that the region of greatest diversity is usually the homeland, this distribution points strongly to an ancestral homeland for

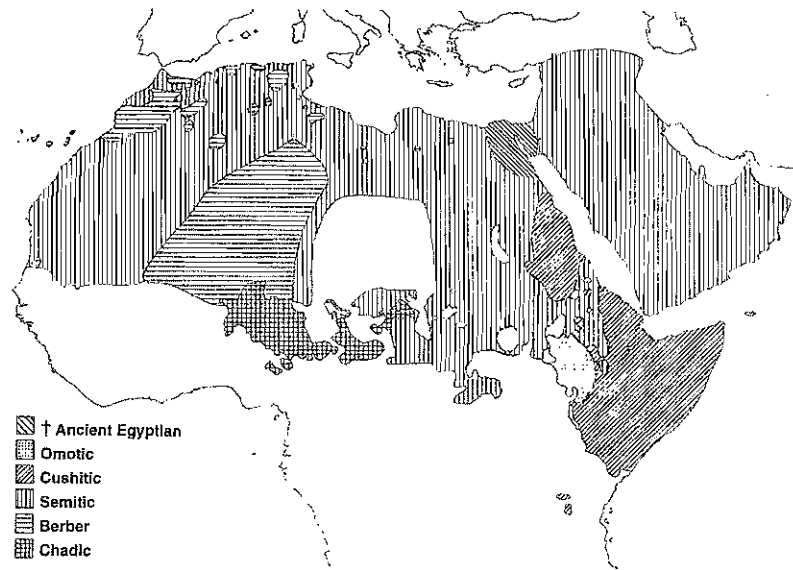


FIGURE 10.4. THE AFROASIATIC LANGUAGE FAMILY.

The major branches of the Afroasiatic language family. Arabic is now spoken in the area shown as belonging to Ancient Egyptian.

Afroasiatic somewhere in northern Africa, which the Semitic speakers left to invade the Near East, perhaps some 9,000 years ago.²⁸⁵ (Later, about 7,000 years ago, some crossed back from Yemen into Ethiopia, giving it the country's principal language of Amharic.) Also pointing to an African homeland, the earliest branching of proto-Afroasiatic was into Omotic and the rest, and the second branching was into Cushitic and the rest. Since Omotic and Cushitic are both restricted to Africa, that has "put it beyond doubt that the ancestral language, proto-Afroasiatic, was spoken in Africa," writes Ehret.²⁸⁶

Though Greenberg's classification of African languages is now broadly accepted, it was for many years bitterly resisted by British Africanists. In linguistics as in other academic fields, specialists tend to resent the generalist who shows how their little patch relates to a larger order. Paul Newman, a linguist at Indiana University, recalls visiting the London School of Oriental and African Studies around 1970, some 15 years after the first publication of Greenberg's African work. He was told that it was quite safe for him to go into the common room, as long as he did not mention Greenberg's name.²⁸⁷

After his African classification, Greenberg turned his attention to the question of American Indian languages. Taking note of the archaeological findings that the Americas had been settled only recently, Greenberg expected to find far fewer language families than in Africa. But American linguists, then undergoing a splittist phase, had agreed at a conference in 1976 that no fewer than 63 independent language families were spoken in the Americas. Greenberg, using the same mass comparison method he had developed for Africa, announced there were just three—Amerind, Na-Dene and Eskimo-Aleut.²⁸⁸

Greenberg's conclusions induced the same agitation among American linguists as his African classification had among the British. And even though American linguists had generally accepted his grouping of African languages, they now assailed him with a fury that startled the population geneticists who were beginning to take an interest in his work. Luca Cavalli-Sforza, an eminent geneticist at Stanford University, wrote of his dismay at the linguists' diatribes against Greenberg.²⁸⁹

Cavalli-Sforza's confidence in Greenberg's approach stemmed from the fact that, at least in general outline, he had confirmed it by an independent approach. Before methods of DNA analysis became available, Cavalli-Sforza and colleagues had worked out a genetic family tree of the world's populations in terms of protein differences. Comparing this tree to Greenberg's list of major language families, Cavalli-Sforza showed that peoples who were grouped together on his world population tree tended to fall into the same language family, as defined by Greenberg.²⁹⁰ Further analysis proved that the correspondence between the world's human population tree and Greenberg's language families was statistically significant.²⁹¹

The Comparative Method versus Mass Comparison

Despite Cavalli-Sforza's support for Greenberg's findings, linguists continued to assail Greenberg's work on grounds of factual errors and methodology. As even Greenberg's supporters concede, he was interested in the big picture, not the details. Numerous small errors, of the type scholars usually do their best to avoid, crept into his work. Some were errors of transcription,

some perhaps the result of working in haste as he reviewed the grammar and vocabulary of hundreds of languages, transcribing everything with his own hand and usually without a graduate student to check things. Were the errors fatal, as his Americanist critics contended, or trivial, as his supporters averred? The verdict of the Africanists, who came to agree with him, is that the errors were not significant. "There are . . . more errors in data-entering than one expects in such a work," writes Lionel Bender, an Africanist at Southern Illinois University, about Greenberg's book on African languages. "Nevertheless, he got it right for the most part and his African classification culminating in the 1963 book is a tremendous advance."²⁹²

The larger point of Greenberg's critics was that in establishing relationships among languages he had failed to use what is known as the comparative method, the orthodox approach to classifying languages. The method is based on identifying sets of related words that change in predictable ways between members of a language family. The French and Italian words for "goat" are not particularly similar, but when compared with other words it is clear that a "k" sound in Italian corresponds with a "ch" sound in French, and a "p" in Italian corresponds with an "f" or "v" in French.²⁹³ These sound correspondences exist because many French and Italian words are cognates, or descendants of the same parent word in their common ancestor tongue of Latin.

LATIN	ITALIAN	FRENCH	ENGLISH
<i>capra</i>	<i>capra</i>	<i>chèvre</i>	<i>goat</i>
<i>caput</i>	<i>capo</i>	<i>chef</i>	<i>head</i>
<i>canis</i>	<i>cane</i>	<i>chien</i>	<i>dog</i>

Once the rules of sound correspondence between contemporary languages have been established, the word in the parent language can be reconstructed. Scholars have reconstructed an extensive vocabulary in proto-Indo-European, the hypothesized ancestral tongue of many European and Indian languages. Any claim that a language is part of the Indo-European family can then be tested by seeing if its grammar and vocabulary can be derived, by the established rules, from proto-Indo-European. From the instances above, English might not seem so promising a candidate, but the initial "k" sound in Latin is known to correspond with an "h" in the Germanic group of languages, making *head* and the German word *haupt* (now a

figurative word for head) cognates with Latin's *caput*. By the same rule Latin's *canis* is cognate with German's *hund* and the English word *hound*, all being derived from proto-Indo-European **kwon*.

Rigorous application of the comparative method has freed linguistics from many false etymologies and crank theories. Many linguists insist that the comparative method is the only acceptable way of testing whether languages are related to each other. This position is based on the belief that, since words change so fast, two daughter languages will soon have only a small percentage of their vocabulary in common and at this point the number of true cognates may be exceeded by chance resemblances and words that sound alike because the two languages under comparison each borrowed them from a third.

Because the signal of the true cognates is soon overwhelmed by the noise of specious ones, the roots of a family of languages, linguists say, can be traced no farther back than about 6,000 years or so, the period when most linguists believe proto-Indo-European was spoken.

Greenberg, in his method of mass comparison, did not look for sound correspondences, nor did he try to reconstruct proto-languages to confirm his findings. Hence, in the view of many linguists, his method and findings cannot be trusted.

Whatever the theoretical objections to Greenberg's method, the bottom line is the empirical question of whether or not it works. Africanists have decided it did indeed work for African languages. But this apparently persuasive circumstance has not changed linguists' views about the validity of Greenberg's method. In a recent essay on Greenberg's Afroasiatic family, Richard Hayward, of the London School of Oriental and African Studies, writes that the "only admissible evidence" for establishing that languages have a common ancestry is by the comparative method and sound correspondences. "Now it was on the basis of 'mass comparison,' rather than the comparative method, that the canon of the Afroasiatic languages was established by Greenberg, and although this methodology . . . has, in the present writer's view, come up with the right conclusions, a methodology that does not invoke the rigour of the principle stated in the last paragraph [i.e., that of the comparative method] cannot make predictions, and so falls short of true theoretical status," Hayward writes.²⁹⁴ In other words, even if Greenberg got the right answer, it was by the wrong method.

If the faculty of human language were extremely ancient, and if human populations were highly mixed, the likelihood of languages on the same continent being related to each other might be small, and it would be appropriate to assume languages were unrelated unless proven otherwise. But since fully modern language probably evolved only 50,000 years ago, and since today's populations still strongly reflect the original patterns of human migration, the reverse is the case: all languages are probably offshoots of a single mother tongue and related to each other at one level or another. In circumstances where history and archaeology make language relationships very likely, such as in the Americas, a lesser standard of proof would perhaps be appropriate. It is surely in Africa, where languages have had longest to diversify, that Greenberg's mass comparison method stood least chance of success, yet it is there that linguists judge it to be most successful.

Linguists' insistence on comparative method as the only acceptable classification tool is a matter of some frustration to researchers who would like to integrate the findings of population genetics and archaeology with a linguistic tree. Without a guide from conventional linguists to deep language relationships, population geneticists tend to rely on the work of Greenberg and Merritt Ruhlen, his Stanford University colleague, as the best available guide to the overall structure of the world's languages.

The Eurasiatic Superfamily

As Greenberg worked on classifying the languages of the Americas, he realized that they must be related in some way to languages on the Eurasian continent, if indeed the Americas had been inhabited by people migrating from Siberia. So to help with the American classification, he started making lists of words in languages of the Eurasian land mass, particularly personal pronouns and interrogative pronouns.

"I began to see when I lined these up that there is a whole group of languages through northern Asia," he said in an interview in 1999. "I must have noticed this 20 years ago. But I realized what scorn the idea would provoke and put off detailed study of it until I had finished the American languages book."²⁹⁵

This was the beginning of Greenberg's next major classification, a link-

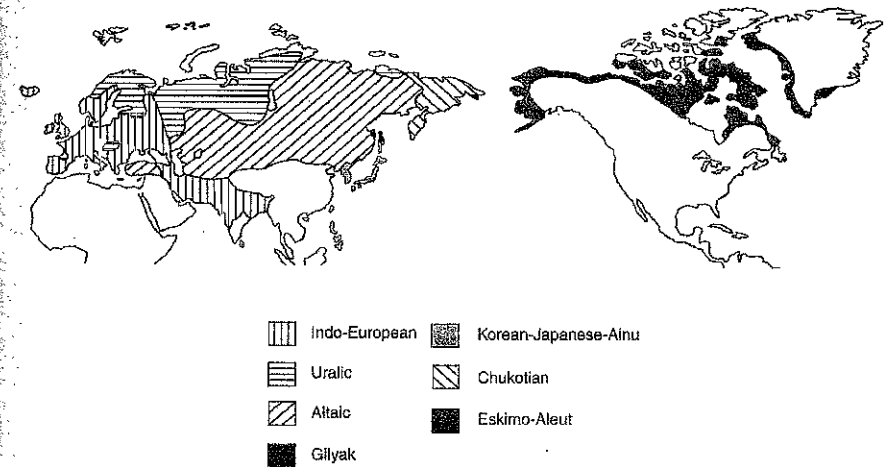


FIGURE 10.5. THE DISTRIBUTION OF EURASIATIC.

The family of Indo-European languages, according to the linguist Joseph Greenberg, belongs to a more ancient superfamily called Eurasiatic. Other members include the Uralic and Altaic families, the Korean-Japanese-Ainu group and the Eskimo-Aleut languages of North America.

ing of many of the major language families of Europe and northern Asia into a single superfamily that he called Eurasiatic. This ancestral tongue, in his view, gave rise to eight families of languages, now spoken in a great swath across northern Eurasia, from Portugal to Japan, and, since Eskimo languages too are included, from Alaska to Greenland.

The best-known member of the Eurasiatic superfamily is the language family known as *Indo-European*, which itself has 11 branches:

1. The Anatolian group, not well known because all its member languages are now extinct. Its principal member is Hittite, the language of the Hittite empire that was centered in Anatolia (now Turkey), and reached its height between 1680 and 1200 BC.
2. Armenian
3. Tokharian, a pair of languages known as Tokharian A and Tokharian B and spoken in northwest China in the second half of the first millennium AD. Though at the east of the Indo-

- European range, Tokharian seems more closely related to languages of the west; the origin and history of its speakers is unclear.
4. Indo-Iranian, which includes the ancient Sanskrit as well as many modern Indian languages such as Urdu and Hindi, along with the ancient and modern languages of the Iranian region.
 5. Albanian
 6. Greek
 7. Italic, which includes Latin and its modern descendants, such as Italian, French, Spanish, Portuguese and Romanian.
 8. Celtic, which includes Irish and Scottish Gaelic.
 9. Germanic, including Danish, Swedish, Norwegian and Icelandic; German, Dutch and Yiddish; and English.
 10. Baltic, including Latvian and Lithuanian.
 11. Slavic, the branch comprised of Russian, Polish, Czech and Serbo-Croatian.

The second major family of Eurasiatic is *Uralic-Yukaghir*, a far flung family that includes Hungarian, Finnish and Estonian in the west and many Siberian languages in the east. This family, in Greenberg's view, includes Ket, a hard to classify Siberian language that may be the source of Na-Dene, the second of the three language families of the Americas along with Amerind and Eskimo-Aleut.

Third is *Altaic*, which includes the Turkish and Mongolian language groups.

Fourth is *Korean-Japanese-Ainu*, a grouping that has no generic name; Ainu is the language spoken by the original inhabitants of northern Japan.

Fifth is *Gilyak*, the language of a dwindling number of people who live in northern Sakhalin, the large island north of Japan, and in a small region opposite Sakhalin on the Siberian mainland.

Sixth is *Chukotian*, a language family of eastern Siberia that includes Chukchi and Koryak.

Seventh is *Eskimo-Aleut*, a family spoken from Siberia to Greenland.

Eighth is *Etruscan*, an extinct language of the Romans' adversaries in ancient Italy.

Greenberg's book on the grammar of his proposed Eurasiatic family was published in 2000; the second volume, on shared vocabulary, appeared

posthumously in 2002. His grouping was developed independently of Nostratic, the superfamily advocated by a Russian school of linguists, but overlaps with it to a great extent. Nostratic differs from Euroasiatic in that it includes Afroasiatic, at least in early versions, and some Nostraticists exclude Japanese and Ainu. An important difference of methodology is that Nostraticists insist proto-languages be reconstructed as the basis for comparison, a procedure that Greenberg skips.

To English speakers, it may not be instantly obvious that their language has anything whatsoever in common with Finnish, Turkish, or Inuit, let alone Japanese, as the Eurasiatic hypothesis asserts. Given the speed of language change, and the 10,000 years or more that separate all these daughter tongues from the assumed proto-Eurasiatic, only a few echoes would be expected. As Greenberg's critics rightly point out, it is hard to be sure that the signal of these faint echoes rises above the noise of chance resemblance.

But consider the comparison of English with, say, Japanese. Given that *wakaru* means understand in Japanese, guess the meaning of *wakaranai*. Apart from the oddity of putting a negative at the end of the verb, it seems natural that *wakaranai* should mean don't understand, and so it does.

In many Indo-European languages, questions are expressed with words starting with "k" or "kw" sounds, though the "kw" has become a "w" in English. French has *quoi* (what?), Italian *come* (how?) and Latin *quando*, *quis*, and *quid pro quo*. So *wakaranaika*? Don't you understand?

It could be just by chance that the Indo-European and Japanese families use "k" sounds for question words. But an interrogative in *k* is found in every branch of Eurasiatic, Greenberg says.²⁹⁶ In the Uralic family, Finnish has *ken*, meaning who? In Altaic, Turkic has *kim*, with the same meaning. In all dialects of Eskimo who is *kina*.

There are many interrogative words, so if one rummages around in all the languages of a proposed family, it's perhaps not so hard to find a few k-words. The same may be true of n-words for negatives. Greenberg's case for Eurasiatic rests not on any specific case but on the combination of a large number of such similarities that he has turned up. These include 72 types of grammatical similarity, though most are shared by only some of the eight postulated families of Eurasiatic. Nonetheless, "This grammatical evidence is quite sufficient in itself to establish the validity of the Eurasiatic family," Greenberg says.

Turning to words, as distinct from grammar, it's probably reasonable to assume that a given sound will ricochet around a related set of meanings over time. The assumption raises the chances of spotting a relationship between language families, but also of picking up accidental similarities. No single group of cognate words is conclusive, but large numbers can begin to make a case. Greenberg has found 437 groups of cognates for Eurasiatic, though very few have examples from every family.²⁹⁷ One of the most interesting concerns a set of meanings based on the putative Eurasiatic word for finger, which Greenberg thinks was *tik*. Raise your first or index *tik* and you make a universally understood sign for the number one. Point it horizontally and you are drawing attention to something. On that basis, Greenberg cites the following echoes of this ancient word.

In the Indo-European family, linguists have reconstructed a proto-Indo-European root **deik*, meaning to show, from which comes the Latin word *digitus* for finger, and the English words *digit* and *digital*. In the Altaic family, the Turkish word for sole or only is *tek*. In the Korean-Japanese-Ainu group, there is Ainu's *tek* and Japanese's *te*, both meaning hand. As for Eskimo-Aleut, Greenlandic has *tikiq* for index finger, Sirenik and Central Alaskan Yupik have *tekeq*.

Greenberg put particular emphasis on another group of cognates, which he saw as providing a link between Eurasiatic and Amerind. It is a set of meanings centered on the word hand and including both give (to give is to hand something over) and measure (the width of the hand is often used as a measure, and in English is the name of the unit for measuring the shoulder height of horses and ponies). Many American Indian languages use a *ma* or *mi* sound as their word for hand (Algonquian **mi*, Uto-Aztecan **ma*, Tequistlatec *mane*, Guato *mara*). In the Eurasiatic family, Indo-European has a root **me-* meaning measure, whence *metric*, as well as Latin *manus*, hand; Gilyak has *man*, to measure by hand spans and *-ma*, a word added to numbers to indicate units of hand spans; Korean has *mān*, an amount or measure.

In Greenberg's view, Eurasiatic and Amerind were sister superfamilies, younger than the original languages of the Old World, of which the strange isolate languages like Basque and Burushaski (spoken in a small region of the northwestern Indian subcontinent) are relics. "The Eurasiatic-Amerind family represents a relatively recent expansion (circa 15,000 [years before the present]) into territory opened up by the melting of the Arctic ice cap. Eurasiatic-

Amerind stands apart from the other families of the Old World, among which the differences are much greater and represent deeper chronological groupings," he wrote in his last work.²⁹⁸ It was, perhaps, a final gibe at his critics, who insisted that languages could be traced back no farther than 5,000 years or so; Greenberg was insisting he could see three times farther than they.^{*299}

Echoes of the First Language

Nothing makes linguists heave wearier sighs than talk of the ancestral human language. The subject, in their general view, is not worth even talking about because, as every serious specialist knows, the roots of language cannot be traced back farther than 5,000 years, 10,000 at the very most. "Given present knowledge of language change and probability," writes Johanna Nichols, "... descent and reconstruction will never be traceable beyond approximately 10,000 years. Methods now being developed reach back much earlier but do not trace descent. Among other things, this means that linguistics will never be able to apply phylogenetic analysis to the question of when language arose and whether all the world's languages are descended from a single ancestor."³⁰⁰

Though Nichols's prediction may prove correct, biologists are not quite so pessimistic. With DNA, their phylogenetic trees reach back hundreds of millions of years, and 50,000 years ago is like yesterday. If Indo-European started to split up 8,700 years ago, as Gray's statistics say, languages may be reconstructible far further back in time than linguists have supposed.

* Greenberg died at the age of 85, the year after publishing the first volume of his work on Eurasiatic. Paul Newman, a linguist at Indiana University, recalls visiting him in his final illness. Greenberg mentioned his regret that he had not gotten around to classifying the languages of southeast Asia. "He looked at me," Newman said, "almost with tears in his eyes, and said that without classifying them, he hadn't finished his work with the world's languages."

Another friend, Harold Fleming of Boston University, a specialist in African languages, paid a farewell visit together with Greenberg's longtime colleague Merritt Ruhlen. The two were the last scholars to talk with him. The conversation turned to a recent work on Gilyak, an obscure language of northern Sakhalin. Fleming mentioned that there was a Gilyak word *irf* meaning fox or jackal. Greenberg was familiar with the word and noted that its "r" was trilled like a French "r."

"What a memory!" Fleming wrote in a tribute. "What a scholar! And what a shame that his vast and unique knowledge of human languages had to leave us, could not be electronically stored, and that such a great scientist had to die under a cloud of misguided criticism!" Source is note 299.

The very existence of Swadesh lists is proof that some words are retained longer than others. Might some be retained for long enough to reconstruct the tree of language 5 times farther back than Gray has done, close to the source of the ancestral tongue? Some words—*new, tongue, where, thou, one, what, name, how*—have half-lives greater than 13,000 years, and another seven words—*I, we, who, two, three, four, five*—are even more resistant to change, according to calculations by Mark Pagel. Such words, in his view, “can potentially resolve very old time depths,” beyond the 5,000 to 10,000 years so often proposed for linguistic data.³⁰¹ The word for one, he notes, has a half-life of 21,000 years. This means it has a 22% chance of not changing in 50,000 years.

Could these long-lived Swadesh words support a genealogy that coalesced on a single proto-language? Greenberg played with the idea that he had found a word that might be a remnant of the mother tongue. It is the group of cognates, mentioned above, that are based on the set of ideas one/finger/point and derived from the root **tik*. Greenberg spotted what he assumed were cognates of this word in at least one member language of many of his language superfamilies. He mentioned the group in a lecture in 1977 but never published it, whether because of his own reservations or from fear of incurring more than the usual deluge of ridicule from his fellow linguists.

In the Eurasiatic family, as noted above, **tik* words range from the English *digital* and Greek *daktulos* to Eskimo *tiqik* for “index finger.” According to Ruhlen, Greenberg first noticed when defining the Nilo-Saharan family that several of its languages had words of the general form *t-k* for the word one.³⁰² The word for one in proto-Afroasiatic has been reconstructed **tak*. In the Austroasiatic family, Cambodian or Khmer has *tai* as the word for hand, and Vietnamese has *tay*. In Amerind languages there are several *tik*-like words meaning finger or alone.

Even linguists who support Greenberg have little patience with the suggestion that the **tik* word may be an echo of the mother tongue. Yet given Pagel’s calculations, it is not impossible that some words still spoken today have very ancient pedigrees, and even that Greenberg’s **tik* is indeed a faint but indelible whisper from the distant days when the world was one.

11

HISTORY

The remarkable success of the English as colonists, compared to other European nations, has been ascribed to their “daring and persistent energy”: a result which is well illustrated by comparing the progress of the Canadians of English and French extraction; but who can say how the English gained their energy? There is apparently much truth in the belief that the wonderful progress of the United States, as well as the character of the people, are the results of natural selection; for the more energetic, restless, and courageous men from all parts of Europe have emigrated during the last ten or twelve generations to that great country, and have there succeeded best. . . . Obscure as is the problem of the advance of civilisation, we can at least see that a nation which produced during a lengthened period the greatest number of highly intellectual, energetic, brave, patriotic, and benevolent men, would generally prevail over less favoured nations. Natural selection follows from the struggle for existence; and this from a rapid rate of increase. It is impossible not to regret bitterly, but whether wisely is another question, the rate at which man tends to increase, for this leads in barbarous tribes to infanticide and many other evils, and in civilised nations to abject poverty, celibacy, and to the late marriages of the prudent. But as man suffers from the same physical evils as the lower animals, he has no right to expect an immunity from the evils consequent on the struggle for existence. Had he not been subjected during primeval times to natural selection, assuredly he would never have attained to his present rank.

CHARLES DARWIN, THE DESCENT OF MAN

WITH SETTLEMENT and the invention of agriculture, human societies embarked on a trajectory quite different from the foraging life that had hitherto been their only choice. The new behaviors that had now been developed allowed people to construct complex societies and urban civilizations.