

## **WARNING CONCERNING COPYRIGHT RESTRICTIONS**

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproduction of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If electronic transmission of reserve material is used for purposes in excess of what constitutes "fair use", that user may be liable for copyright infringement.

commonly accompany some other word get eroded and then glued onto it. Tense markers may come from auxiliaries; for example, as I've mentioned, the English *-ed* suffix may have evolved from *did*: *hammer-did* → *hammered*. Case markers may come from slurred prepositions or from sequences of verbs (for example, in a language that allows the construction *take nail hit it*, *take* might erode into an accusative case marker like *ta-*). Agreement markers can arise from pronouns: in *John, he kissed her*, *he* and *her* can eventually glom onto the verb as agreement affixes.

• Syntactic constructions can arise when a word order that is merely preferred becomes reanalyzed as obligatory. For example, when English had case markers, both *give him a book* and *give a book him* were possible, but the former was more common. When the case markers eroded in casual speech, many sentences would have become ambiguous if order were still allowed to vary. The more common order was thus enshrined as a rule of syntax. Other constructions can arise from multiple reanalyses. The English perfect *I had written a book* originally came from *I had a book written* (meaning "I owned a book that was written"). The reanalysis was inviting because the SOV pattern was alive in English; the participle *written* could be reanalyzed as the main verb of the sentence, and *had* could be reanalyzed as its auxiliary, begetting a new analysis with a related meaning.



The third ingredient for language splitting is separation among groups of speakers, so that successful innovations do not take over everywhere but accumulate separately in the different groups. Though people modify their language every generation, the extent of these changes is slight: vastly more sounds are preserved than mutated, more constructions analyzed properly than reanalyzed. Because of this overall conservatism, some patterns of vocabulary, sound, and grammar survive for millennia. They serve as the fossilized tracks of mass migrations in the remote past, clues to how human beings spread out over the earth to end up where we find them today.

How far back can we trace the language of this book, modern American English? Surprisingly far, perhaps five or even nine thousand years. Our knowledge of where our language has come from is considerably more precise than the recollection of Dave Barry's Mr. Language Person: "The English language is a rich verbal tapestry

woven together from the tongues of the Greeks, the Latins, the Angles, the Klaxtons, the Celts, and many more other ancient peoples, all of whom had severe drinking problems." Let's work our way back.

America and England first came to be divided by a common language, in Wilde's memorable words, when colonists and immigrants isolated themselves from British speech by crossing the Atlantic Ocean. England was already a Babel of regional and class dialects when the first colonists left. What was to become the standard American dialect was seeded by the ambitious or dissatisfied members of lower and middle classes from southeastern England. By the eighteenth century an American accent was noted, and pronunciation in the American South was particularly influenced by the immigration of the Ulster Scots. Westward expansions preserved the layers of dialects of the eastern seaboard, though the farther west the pioneers went, the more their dialects mixed, especially in California, which required leapfrogging of the vast interior desert. Because of immigration, mobility, literacy, and now the mass media, the English of the United States, even with its rich regional differences, is homogeneous compared with the languages in territories of similar size in the rest of the world; the process has been called "Babel in reverse." It is often said that the dialects of the Ozarks and Appalachia are a relict of Elizabethan English, but this is just a quaint myth, coming from the misconception of language as a cultural artifact. We think of the folk ballads, the hand-stitched quilts, and the whiskey aging slowly in oak casks and easily swallow the rumor that in this land that time forgot, the people still speak the traditional tongue lovingly handed down through the generations. But language does not work that way—at all times, in all communities, language changes, though the various parts of a language may change in different ways in different communities. Thus it is true that these dialects preserve some English forms that are rare elsewhere, such as *afeared*, *yourn*, *bism*, and *et*, *holp*, and *clome* as the past of *eat*, *help*, and *climb*. But so does every variety of American English, including the standard one. Many so-called Americanisms were in fact carried over from England, where they were subsequently lost. For example, the participle *gotten*, the pronunciation of *a* in *path* and *bath* with a front-of-the-mouth "a" rather than the back-of-the-mouth "ah," and the use of *mad* to mean "angry," *fall* to mean "autumn," and *sick* to mean "ill," strike the

British ear as all-American, but they are actually holdovers from the English that was spoken in the British Isles at the time of the American colonization.

English has changed on both sides of the Atlantic, and had been changing well before the voyage of the *Mayflower*. What grew into standard contemporary English was simply the dialect spoken around London, the political and economic center of England, in the seventeenth century. In the centuries preceding, it had undergone a number of major changes, as you can see in these versions of the Lord's Prayer:

CONTEMPORARY ENGLISH: Our Father, who is in heaven, may your name be kept holy. May your kingdom come into being. May your will be followed on earth, just as it is in heaven. Give us this day our food for the day. And forgive us our offenses, just as we forgive those who have offended us. And do not bring us to the test. But free us from evil. For the kingdom, the power, and the glory are yours forever. Amen.

EARLY MODERN ENGLISH (C. 1600): Our father which are in heaven, hallowed be thy Name. Thy kingdom come. Thy will be done, on earth, as it is in heaven. Give us this day our daily bread. And forgive us our trespasses, as we forgive those who trespass against us. And lead us not into temptation, but deliver us from evil. For thine is the kingdom, and the power, and the glory, for ever, amen.

MIDDLE ENGLISH (C. 1400): Oure fadir that art in heuenes halowid be thi name, thi kyngdom come to, be thi wille don in erthe es in heuene, yeue to us this day oure bread ouir other substance, & foyeue to us oure dettis, as we forgeuen to oure dettouris, & lede us not in to temptacion: but delyuer us from yuel, amen.

OLD ENGLISH (C. 1000): Faeder ure thu the eart on heofonum, si thin nama gehalgod. Tobecume thin rice. Gewurthe in willa on eorthan swa swa on heofonum. Urne gedaeghwamlican hlaf syle us to daeg. And forgyf us ure gyltas, swa swa we forgyfath urum gyltedum, And ne gelaed thu us on contnungen ac alys us of yfele. Sothlice.

The roots of English are in northern Germany near Denmark, which was inhabited early in the first millennium by pagan tribes called the Angles, the Saxons, and the Jutes. After the armies of the

collapsing Roman Empire left Britain in the fifth century, these tribes invaded what was to become England (Angle-land) and displaced the indigenous Celts there into Scotland, Ireland, Wales, and Cornwall. Linguistically, the defeat was total; English has virtually no traces of Celtic. Vikings invaded in the ninth to eleventh centuries, but their language, Old Norse, was similar enough to Anglo-Saxon that aside from many borrowings, the language, Old English, did not change much.

In 1066 William the Conqueror invaded Britain, bringing with him the Norman dialect of French, which became the language of the ruling classes. When King John of the Anglo-Norman kingdom lost Normandy shortly after 1200, English reestablished itself as the exclusive language of England, though with a marked influence of French that lasts to this day in the form of thousands of words and a variety of grammatical quirks that go with them. This "Latinated" vocabulary—including such words as *donate*, *vibrate*, and *desist*—has a more restricted syntax; for example, you can say *give the museum a painting* but not *donate the museum a painting*, *shake it up* but not *vibrate it up*. The vocabulary also has its own sound pattern: Latinated words are largely polysyllabic with stress on the second syllable, such as *desist*, *construct*, and *transmit*, whereas their Anglo-Saxon synonyms *stop*, *build*, and *send* are single syllables. The Latinated words also trigger many of the sound changes that make English morphology and spelling so idiosyncratic, like *electric*–*electricity* and *nation*–*national*. Because Latinated words are longer, and are more formal because of their ancestry in the government, church, and schools of the Norman conquerors, overusing them produces the stuffy prose universally deplored by style manuals, such as *The adolescents who had effectuated forcible entry into the domicile were apprehended* versus *We caught the kids who broke into the house*. Orwell captured the flabbiness of Latinated English in his translation of a passage from Ecclesiastes into modern institutionalese:

I returned and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise; nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all.

Objective consideration of contemporary phenomena compels the conclusion that success or failure in competitive activities exhibits

no tendency to be commensurate with innate capacity, but that a considerable element of the unpredictable must invariably be taken into account.

English changed noticeably in the Middle English period (1100–1450) in which Chaucer lived. Originally all syllables were enunciated, including those now represented in spelling by “silent” letters. For example, *make* would have been pronounced with two syllables. But the final syllables became reduced to the generic schwa like the *a* in *allow* and in many cases they were eliminated entirely. Since the final syllables contained the case markers, overt case began to vanish, and the word order became fixed to eliminate the resulting ambiguity. For the same reason, prepositions and auxiliaries like *of* and *do* and *will* and *have* were bled of their original meanings and given important grammatical duties. Thus many of the signatures of modern English syntax were the result of a chain of effects beginning with a simple shift in pronunciation.

The period of Early Modern English, the language of Shakespeare and the King James Bible, lasted from 1450 to 1700. It began with the Great Vowel Shift, a revolution in the pronunciation of long vowels whose causes remain mysterious. (Perhaps it was to compensate for the fact that long vowels sounded too similar to short vowels in the monosyllables that were now prevalent; or perhaps it was a way for the upper classes to differentiate themselves from the lower classes once Norman French became obsolete.) Before the vowel shift, *mouse* had been pronounced “mooce”; the old “oo” turned into a diphthong. The gap left by the departed “oo” was filled by raising what used to be an “oh” sound; what we pronounce as *goose* had, before the Great Vowel Shift, been pronounced “goce.” That vacuum, in turn, was filled by the “o” vowel (as in *hot*, only drawn out), giving us *broken* from what had previously been pronounced more like “brocken.” In a similar rotation, the “ee” vowel turned into a diphthong; *like* had been pronounced “leek.” This dragged in the vowel “eh” to replace it; our *geese* was originally pronounced “gace.” And that gap was filled when the long version of *ah* was raised, resulting in *name* from what used to be pronounced “nahma.” The spelling never bothered to track these shifts, which is why the letter *a* is pronounced one way in *can* and another way in *came*, where it had formerly been just a longer version of the *a* in *can*. This

is also why vowels are rendered differently in English spelling than in all the other European alphabets and in “phonetic” spelling.

Incidentally, fifteenth-century Englishmen did not wake up one day and suddenly pronounce their vowels differently, like a switch to Daylight Savings Time. To the people living through it, the Great Vowel Shift probably felt like the current trend in the Chicago area to pronounce *hoi* like *bat*, or the growing popularity of that strange surfer dialect in which *dude* is pronounced something like “diihh-hoooood.”



What happens if we try to go back farther in time? The languages of the Angles and the Saxons did not come out of thin air; they evolved from Proto-Germanic, the language of a tribe that occupied much of northern Europe in the first millennium B.C. The western branch of the tribe split into groups that gave us not only Anglo-Saxon, but German and its offshoot Yiddish, and Dutch and its offshoot Afrikaans. The northern branch settled Scandinavia and came to speak Swedish, Danish, Norwegian, and Icelandic. The similarities in vocabulary among these languages are visible in an instant, and there are many similarities in grammar as well, such as forms of the past-tense ending *-ed*.

The ancestors of the Germanic tribes left no clear mark in written history or the archeological record. But they did leave a special mark on the territory they occupied. That mark was discerned in 1786 by Sir William Jones, a British judge stationed in India, in one of the most extraordinary discoveries in all scholarship. Jones had taken up the study of Sanskrit, a long-dead language, and noted:

The Sanskrit language, whatever may be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed that no philologist could examine them all three, without believing them to have sprung from some common source, which, perhaps no longer exists; there is a similar reason, though not quite so forcible, for supposing that both the Gothic [Germanic] and the Celtic; though blended with a very different

idiom, had the same origin as the Sanskrit; and the old Persian might be added to the same family . . .

Here are the kinds of affinities that impressed Jones:

ENGLISH:	brother	mead	is	thou bearest	he bears
GREEK:	phrater	methu	esti	phereis	pherei
LATIN:	frater		est	fers	fert
OLD SLAVIC:	bratre	mid	yeste	berasi	beretu
OLD IRISH:	brathir	mith	is		beri
SANSKRIT:	bhrater	medhu	asti	bharasi	bharati

Such similarities in vocabulary and grammar are seen in an immense number of modern languages. Among others, they embrace Germanic, Greek, Romance (French, Spanish, Italian, Portuguese, Romanian), Slavic (Russian, Czech, Polish, Bulgarian, Serbo-Croatian), Celtic (Gaelic, Irish, Welsh, Breton), and Indo-Iranian (Persian, Afghan, Kurdish, Sanskrit, Hindi, Bengali, and the Romany language of the Gypsies). Subsequent scholars were able to add Anatolian (extinct languages spoken in Turkey, including Hittite), Armenian, Baltic (Lithuanian and Latvian), and Tocharian (two extinct languages spoken in China). The similarities are so pervasive that linguists have reconstructed a grammar and a large dictionary for a hypothetical common ancestor language, Proto-Indo-European, and a set of systematic rules by which the daughter languages changed. For example, Jacob Grimm (one of the two Grimm brothers, famous as collectors of fairy tales) discovered the rule by which *p* and *t* in Proto-Indo-European became *f* and *th* in Germanic, as one can see in comparing Latin *pater* and Sanskrit *piter* with English *father*.

The implications are mind-boggling. Some ancient tribe must have taken over most of Europe, Turkey, Iran, Afghanistan, Pakistan, northern India, western Russia, and parts of China. The idea has excited the imagination of a century of linguists and archeologists, though even today no one really knows who the Indo-Europeans were. Ingenious scholars have made guesses from the reconstructed vocabulary. Words for metals, wheeled vehicles, farm implements, and domesticated animals and plants suggest that the Indo-Europeans were a late Neolithic people. The ecological distributions of the natural objects for which there are Proto-Indo-European words—elm and

willow, for example, but not olive or palm—have been used to trace the speakers somewhere in the territory from inland northern Europe to southern Russia. Combined with words for patriarch, fort, horse, and weapons, the reconstructions led to an image of a powerful conquering tribe spilling out of an ancestral homeland on horseback to overrun most of Europe and Asia. The word "Aryan" became associated with the Indo-Europeans, and the Nazis claimed them as ancestors. More sanely, archeologists have linked them to artifacts of the Kurgan culture in the southern Russian steppes from around 3500 B.C., a band of tribes that first harnessed the horse for military purposes.

Recently the archeologist Colin Renfrew has argued that the Indo-European takeover was a victory not of the chariot but of the cradle. His controversial theory is that the Indo-Europeans lived in Anatolia (part of modern Turkey) on the flanks of the Fertile Crescent region around 7000 B.C., where they were among the world's first farmers. Farming is a method for mass-producing human beings by turning land into bodies. Farmers' daughters and sons need more land, and even if they moved just a mile or two from their parents, they would quickly engulf the less fecund hunter-gatherers standing in their way. Archeologists agree that farming spread in a wave that began in Turkey around 8500 B.C. and reached Ireland and Scandinavia by 2500 B.C. Geneticists recently discovered that a certain set of genes is most concentrated among modern people in Turkey and becomes progressively diluted as one moves through the Balkans to northern Europe. This supports the theory originally proposed by the human geneticist Luca Cavalli-Sforza that farming spread by the movement of farmers, as their offspring interbred with indigenous hunter-gatherers, rather than by the movement of farming techniques, as a fad adopted by the hunter-gatherers. Whether these people were the Indo-Europeans, and whether they spread into Iran, India, and China by a similar process, is still not known. It is an awesome possibility. Every time we use a word like *brother*, or form the past tense of an irregular verb like *break-broke* or *drink-drank*, we would be using the preserved speech patterns of the instigators of the most important event in human history, the spread of agriculture.

Most of the other human languages on earth can also be grouped into phyla descending from ancient tribes of astoundingly successful farmers, conquerors, explorers, or nomads. Not all of Europe is Indo-

European. Finnish, Hungarian, and Estonian are Uralic languages, which together with Lappish, Samoyed, and other languages are the remnants of a vast nation based in central Russia about 7,000 years ago. Altaic is generally thought to include the main languages of Turkey, Mongolia, the Islamic republics of the former USSR, and much of central Asia and Siberia. The earliest ancestors are uncertain, but later ones include a sixth-century empire as well as the Mongolian empire of Genghis Khan and the Manchu dynasty. Basque is an orphan, presumably from an island of aboriginal Europeans that resisted the Indo-European tidal wave.

Afro-Asiatic (or Hamito-Semitic), including Arabic, Hebrew, Maltese, Berber, and many Ethiopian and Egyptian languages, dominates Saharan Africa and much of the Middle East. The rest of Africa is divided among three groups. Khoisan includes the !Kung and other groups (formerly called "Hottentots" and "Bushmen"), whose ancestors once occupied most of sub-Saharan Africa. The Niger-Congo phylum includes the Bantu family, spoken by farmers from western Africa who pushed the Khoisan into their current small enclaves in southern, and southeastern Africa. The third phylum, Nilo-Saharan, occupies three large patches in the southern Saharan region.

In Asia, Dravidian languages such as Tamil dominate southern India and are found in pockets to the north. Dravidian speakers must therefore be the descendants of a people who occupied most of the Indian subcontinent before the incursion of the Indo-Europeans. Some 40 languages between the Black Sea and the Caspian Sea belong to the family called Caucasian (not to be confused with the informal racial term for the typically light-skinned people of Europe and Asia). Sino-Tibetan includes Chinese, Burmese, and Tibetan. Austronesian, having nothing to do with Australia (*Austr-* means "south"), includes the languages of Madagascar off the coast of Africa, Indonesia, Malaysia, the Philippines, New Zealand (Maori), Micronesia, Melanesia, and Polynesia, all the way to Hawaii—the record of people with extraordinary wanderlust and seafaring skill. Vietnamese and Khmer (the language of Cambodia) fall into Austro-Asiatic. The 200 aboriginal languages of Australia belong to a family of their own, and the 800 of New Guinea belong to a family as well, or perhaps to a small number of families. Japanese and Korean look like linguistic orphans, though a few linguists lump one or both with Altaic.

What about the Americas? Joseph Greenberg, whom we met earlier

as the founder of the study of language universals, also classifies languages into phyla. He played a large role in unifying the 1,500 African languages into their four groups. Recently he has claimed that the 200 language stocks of native Americans can be grouped into only three phyla, each descending from a group of migrants who came over the Bering land bridge from Asia beginning 12,000 years ago or earlier. The Eskimos and Aleuts were the most recent immigrants. They were preceded by the Na-Dene, who occupied most of Alaska and northwestern Canada and embrace some of the languages of the American Southwest such as Navajo and Apache. This much is widely accepted. But Greenberg has also proposed that all the other languages, from Hudson Bay to Tierra del Fuego, belong to a single phylum, Amerind. The sweeping idea that America was settled by only three migrations has received some support from recent studies by Cavalli-Sforza and others of modern natives' genes and tooth patterns, which fall into groups corresponding roughly to the three language phyla.



At this point we enter a territory of fierce controversy but potentially large rewards. Greenberg's hypothesis has been furiously attacked by other scholars of American languages. Comparative linguistics is an impeccably precise domain of scholarship, where radical divergences between related languages over centuries or a few millennia can with great confidence be traced back step by step to a common ancestor. Linguists raised in this tradition are appalled by Greenberg's unorthodox method of lumping together dozens of languages based on rough similarities in vocabulary, rather than carefully tracing sound-changes and reconstructing proto-languages. As an experimental psycholinguist who deals with the noisy data of reaction times and speech errors, I have no problem with Greenberg's use of many loose correspondences, or even with the fact that some of his data contain random errors. What bothers me more is his reliance on gut feelings of similarity rather than on actual statistics that control for the number of correspondences that might be expected by chance. A charitable observer can always spot similarities in large vocabulary lists, but that does not imply that they descended from a common lexical ancestor. It could be a coincidence, like the fact that the word for "blow" is *pneu* in Greek and *pniv* in Klamath (an American

Indian language spoken in Oregon), or the fact that the word for "dog" in the Australian aboriginal language Mbabaram happens to be *dog*. (Another serious problem, which Greenberg's critics do point out, is that languages can resemble each other because of lateral borrowing rather than vertical inheritance, as in the recent exchanges that led to *her negligées* and *le weekend*.)

The odd absence of statistics also leaves in limbo a set of even more ambitious, exciting, and controversial hypotheses about language families and the prehistoric peoplings of continents that they would represent. Greenberg and his associate Merritt Ruhlen are joined by a school of Russian linguists (Sergei Starostin, Aharon Dogopolsky, Vitaly Shevoroshkin, and Vladislav Illich-Svitych) who lump languages aggressively and seek to reconstruct the very ancient language that would have been the progenitor of each lump. They discern similarities among the proto-languages of Indo-European, Afro-Asiatic, Dravidian, Altaic, Uralic, and Eskimo-Aleut, as well as theorphans Japanese and Korean and a few miscellaneous language groups, reflecting a common ancestor proto-proto-language they call Nostratic. For example, the reconstructed Proto-Indo-European word for mulberry, *mor*, is similar to Proto-Altaic *mür* "berry," Proto-Uralic *marja* "berry," and Proto-Kartvelian (Georgian) *mar-cau* "strawberry." The Nostraticists would have them all evolve from the hypothetical Nostratic root *marja*. Similarly, Proto-Indo-European *melg* "to milk" resembles Proto-Uralic *malge* "breast" and Arabic *mlg* "to suckle." Nostratic would have been spoken by a hunter-gatherer population, for there are no names of domesticated species among the 1,600 words the linguists claim to have reconstructed. The Nostratic hunter-gatherers would have occupied all of Europe, northern Africa, and northern, northeastern, western, and southern Asia, perhaps 15,000 years ago, from an origin in the Middle East.

And various lumpers from this school have suggested other audacious superphyla and super-superphyla. One comprises Amerind and Nostratic. Another, Sino-Caucasian, comprises Sino-Tibetan, Caucasian, and maybe Basque and Na-Dene. Lumping the lumps, Starostin has suggested that Sino-Caucasian can be connected to Amerind-Nostratic, forming a proto-proto-proto language that has been called SCAN, covering continental Eurasia and the Americas. Austric would embrace Austronesian, Austro-Asiatic, and various minor languages in China and Thailand. In Africa, some see similarities between Niger-

Congo and Nilo-Saharan that warrant a Congo-Saharan group. If we were to accept all of these mergers—and some are barely distinguishable from wishful thinking—all human languages would fall into only six groups: SCAN in Eurasia, the Americas, and northern Africa; Khoisan and Congo-Saharan in sub-Saharan Africa; Austric in Southeast Asia and the Indian and Pacific Oceans; Australian; and New Guinean.

Ancestral stocks of this geographic magnitude would have to correspond to the major expansions of the human species, and Cavalli-Sforza and Ruhlen have argued that they do. Cavalli-Sforza examined minor variations in the genes of hundreds of people representing a full spectrum of racial and ethnic groups. He claims that by lumping together sets of people who have similar genes, and then lumping the lumps, a genetic family tree of humankind can be constructed. The first bifurcation splits the sub-Saharan Africans off from everyone else. The adjoining branch in turn splits into two, one embracing Europeans, northeast Asians (including Japanese and Koreans), and American Indians, the other containing southeast Asians and Pacific Islanders on one sub-branch, and aboriginal Australians and New Guineans on another. The correspondences with the hypothetical language superphyla are reasonably clear, though not perfect. One interesting parallel is that what most people think of as the Mongoloid or Oriental race on the basis of superficial facial features and skin coloring may have no biological reality. In Cavalli-Sforza's genetic family tree, northeast Asians such as Siberians, Japanese, and Koreans are more similar to Europeans than to southeast Asians such as Chinese and Thai. Strikingly, this non-obvious racial grouping corresponds to the non-obvious linguistic grouping of Japanese, Korean, and Altaic with Indo-European in Nostratic, separate from the Sino-Tibetan family in which Chinese is found.

The branches of the hypothetical genetic/linguistic family tree can be taken to depict the history of *Homo sapiens sapiens*, from the African population in which mitochondrial Eve was thought to evolve 200,000 years ago, to the migrations out of Africa 100,000 years ago through the Middle East to Europe and Asia, and from there, in the past 50,000 years, to Australia, the islands of the Indian and Pacific Oceans, and the Americas. Unfortunately, the genetic and migrational family trees are almost as controversial as the linguistic one, and any part of this interesting story could unravel in the next few years.

A correlation between language families and human genetic groupings does *not*, by the way, mean that there are genes that make it easier for some kinds of people to learn some kinds of languages. This folk myth is pervasive, like the claim of some French speakers that only those with Gallic blood can truly master the gender system, or the insistence of my Hebrew teacher that the assimilated Jewish students in his college classes innately outperformed their Gentile classmates. As far as the language instinct is concerned, the correlation between genes and languages is a coincidence. People store genes in their gonads and pass them to their children through their genitals; they store grammars in their brains and pass them to their children through their mouths. Gonads and brains are attached to each other in bodies, so when bodies move, genes and grammars move together. That is the only reason that geneticists find any correlation between the two. We know that the connection is easily severed, thanks to the genetic experiments called immigration and conquest, in which children get their grammars from the brains of people other than their parents. Needless to say, the children of immigrants learn a language, even one separated from their parents' language by the deepest historical roots, without any disadvantage compared to age-mates who come from long lineages of the language's speakers. Correlations between genes and languages are thus so crude that they are measurable only at the level of superphyla and aboriginal races. In the past few centuries, colonization and immigration have completely scrambled the original correlations between the superphyla and the inhabitants of the different continents; native English speakers, to take the most obvious example, include virtually every racial subgroup on earth. Well before that, Europeans interbred with their neighbors and conquered each other often enough that there is almost no correlation between genes and language families within Europe (though the ancestors of the non-Indo-European Lapps, Maltese, and Basques left a few genetic mementos). For similar reasons, well-accepted language phyla can contain strange genetic bedfellows, like the black Ethiopians and white Arabs in the Afro-Asiatic phylum, and the white Lapps and Oriental Samoyeds in Uralic.

Moving from the highly speculative to the borderline flaky, Shvovoshkin, Ruhlen, and others have been trying to reconstruct words ancestral to the six superphyla—the vocabulary of the language of

African Eve, "Proto-World." Ruhlen has posited 31 roots, such as *tik* "one" which would have evolved into Proto-Indo-European *deik* "to point" and then Latin *digit* "finger," Nilo-Saharan *dik* "one," Eskimo *tik* "index finger," Kede *tong* "arm," Proto-Afro-Asiatic *tak* "one," and Proto-Austro-Asiatic *ktig* "arm or hand." Though I am willing to be patient with Nostratic and similar hypotheses pending the work of a good statistician with a free afternoon, I find the Proto-World hypothesis especially suspect. (Comparative linguists are speechless.) It is not that I doubt that language evolved only once, one of the assumptions behind the search for the ultimate mother tongue. It's just that one can trace words back only so far. It is like the man who claimed to be selling Abraham Lincoln's ax—he explained that over the years the head had to be replaced twice and the handle three times. Most linguists believe that after 10,000 years no traces of a language remain in its descendants. This makes it extremely doubtful that anyone will find extant traces of the most recent ancestor of all contemporary languages, or that that ancestor would in turn retain traces of the language of the first modern humans, who lived some 200,000 years ago.



This chapter must end on a sad and urgent note. Languages are perpetuated by the children who learn them. When linguists see a language spoken only by adults, they know it is doomed. By this reasoning, they warn of an impending tragedy in the history of humankind. The linguist Michael Krauss estimates that 150 North American Indian languages, about 80% of the existing ones, are moribund. Elsewhere, his counts are equally grim: 40 moribund languages (90% of the existing ones) in Alaska and northern Siberia, 160 (23%) in Central and South America, 45 (70%) in Russia, 225 (90%) in Australia, perhaps 3,000 (50%) worldwide. Only about 600 languages are reasonably safe by dint of the sheer number of their speakers, say, a minimum of 100,000 (though this does not *guarantee* even short-term survival), and this optimistic assumption still suggests that between 3,600 and 5,400 languages, as many as 90% of the world's total, are threatened with extinction in the next century.

The wide-scale extinction of languages is reminiscent of the current (though less severe) wide-scale extinction of plant and animal species.



The causes overlap. Languages disappear by the destruction of the habitats of their speakers, as well as by genocide, forced assimilation and assimilatory education, demographic submersion, and bombardment by electronic media, which Krauss calls "cultural nerve gas." Aside from halting the more repressive social and political causes of cultural annihilation, we can forestall some linguistic extinctions by developing pedagogical materials, literature, and television in the indigenous language. Other extinctions can be mitigated by preserving grammars, lexicons, texts, and recorded speech samples with the help of archives and faculty positions for native speakers. In some cases, like Hebrew in the twentieth century, the continued ceremonial use of a language together with preserved documents can be sufficient to revive it, given the will.

Just as we cannot reasonably hope to preserve every species on earth, we cannot preserve every language, and perhaps should not. The moral and practical issues are complex. Linguistic differences can be a source of lethal divisiveness, and if a generation chooses to switch to a language of the mainstream that promises them economic and social advancement, does some outside group have the right to coerce them not to on the grounds that it finds the idea of them keeping the old language pleasing? But such complexities aside, when 3,000-odd languages are moribund, we can be sure that many of the deaths are unwanted and preventable.

Why should people care about endangered languages? For linguistics and the sciences of mind and brain that encompass it, linguistic diversity shows us the scope and limits of the language instinct. Just think of the distorted picture we would have if only English were available for study! For anthropology and human evolutionary biology, languages trace the history and geography of the species, and the extinction of a language (say, Ainu, formerly spoken in Japan by a mysterious Caucasoid people) can be like the burning of a library of historical documents or the extinction of the last species in a phylum. But the reasons are not just scientific. As Krauss writes, "Any language is a supreme achievement of a uniquely human collective genius, as divine and endless a mystery as a living organism."<sup>3</sup> A language is a medium from which a culture's verse, literature, and song can never be extricated. We are in danger of losing treasures ranging from Yiddish, with far more words for "simpleton" than the Eskimos were reputed to have for "snow," to Damin, a ceremonial

variant of the Australian language Lardil, which has a unique 200-word vocabulary that is learnable in a day but that can express the full range of concepts in everyday speech. As the linguist Ken Hale has put it, "The loss of a language is part of the more general loss being suffered by the world, the loss of diversity in all things."