INTRODUCTION

First Words

Whatever the traits that separate humans from our ape ancestors, complex language is clearly among them. The advantages conferred by the ability to string meaningful words together in unlimited combinations are obvious to anyone who has ever struggled to communicate across a language barrier. Words offer a bridge to another mind, allowing us to cooperate in complex ways and knitting our communities together.

But how did this powerful ability evolve? And how has language changed through time, from what was presumably one mother tongue to the babel of thousands of languages spoken today? This interdisciplinary special issue explores these twin problems of language evolution, and also peers ahead into our ever-evolving linguistic future. Five News stories explore the history and prehistory of language evolution, from the origin of speech to recent language changes, and three Viewpoints speculate on the future. Elsewhere in this issue, three Book Reviews explore the latest in a growing crop of books on this topic.

In several cases, old theories associated with leading scholars are breaking down. For example, as Holden reports (p. 1316), linguists and neuroscientists armed with new types of data are moving beyond the nonevolutionary paradigm once suggested by Noam Chomsky, and tackling the origins of speech head-on. As for the first language, tantalizing linguistic and genetic clues suggest that it was rich with clicking and sucking noises now heard only in a few corners of Africa (see Pennisi's story on p. 1319).

As humans colonized the world and separated into distinct groups, one language diverged into many. Over the past several decades, influential linguist Joseph Greenberg categorized many of these languages into appealingly simple schemes. But today some historical linguists, as they seek the ghosts of dead languages in living ones, find that the evidence does not support simple divisions (see story by Pennisi on p. 1321). Most linguists do agree on grouping together the 144 members of the Indo-European language family, which encompasses tongues as distant as English and Sanskrit. But researchers debate rival theories concerning which culture first spread this language family throughout Europe (see story by Balter on p. 1323 and the Book Review by Mithen on p. 1298).

An individual language can evolve over the time scale of centuries, as seen in the dramatic differences between Modern English and Old English. Bhattacharjee reports (p. 1326) that researchers now use computer models as well as historical texts to probe how such changes arise and spread.

Language evolution has not stopped, of course; in fact, it may be progressing more rapidly than ever before. Population growth, which is concentrated in non–English-speaking countries, and the rapid growth of technology and global communication are causing the demise of many languages while feeding the emergence and growth of new ones. Three Viewpoints explore these dynamics and speculate on the future of written, spoken (and now encoded) language. Graddol (p. 1329) provides a glimpse of how these and other forces may affect the evolution of languages worldwide over the next several decades, raising the notion that English may be waning in dominance and pondering a mixed multilingual future. Aho (p. 1331) describes the rapid evolution of computer languages, which are and increasingly will be an enormous hidden cost in societies' infrastructure. And Montgomery (p. 1333) explores the evolution and future of scientific discourse, where language is affecting and being affected by specialization and renewed attempts to bridge disciplines and cultures. These dynamics are further illustrated in two articles in *Science*'s STKE (see p. 1251). The new generation of studies in language evolution has not yet produced simple answers, but it is reframing the questions—a sure sign of progress.

–ELIZABETH CULOTTA AND BROOKS HANSON

CONTENTS

Related Book Reviews on p. 1298 and Science's STKE material at http://stke.sciencemag.org/cgi/content/abstract/stke.2222004eg4.

Science

CREDIT: NIGEL J. DENNIS/PHOTO RESEARCHERS INC.

SPECIAL SECTION

NEWS

The Origin of Speech

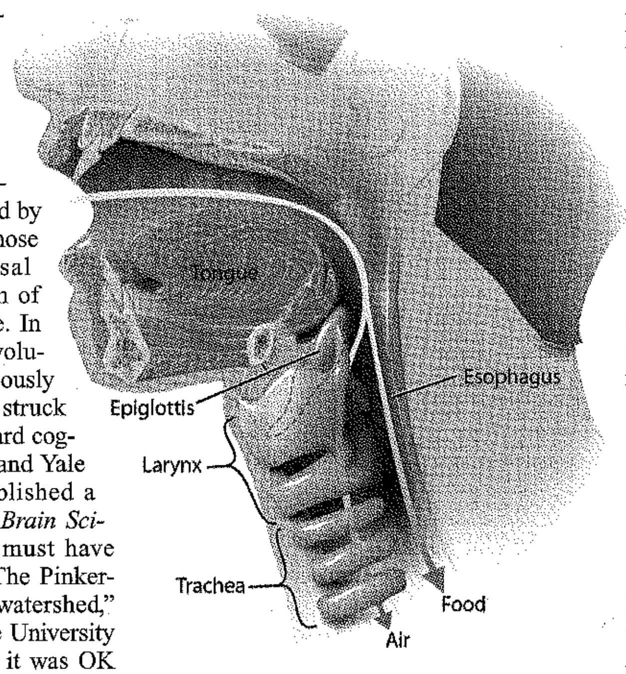
How did the remarkable ability to communicate in words first evolve? Researchers probing the neurological basis of language are focusing on seemingly unrelated abilities such as mimicry and movement

In the 1860s, both the British Academy and the Société de Linguistique de Paris warned their members not to discuss the origins of language, because the topic was so seductive—and so speculative—that it spawned endless, futile theorizing. More than a century later, Noam Chomsky, the most influential linguist of the last 50 years, wrote that language evolution and the brain mechanisms underlying it "appear to be beyond serious inquiry at the moment."

But the time now appears ripe for this endeavor. In the past decade, an unprecedented number of researchers from many disciplines have begun to tackle the origin of speech, spurred by new techniques as well as new ways of thinking. Among linguists, the question of language origins was long obscured by the dominance of Chomsky, whose theory of an innate "universal grammar" ignored the problem of how this language ability arose. In 1990, however, the wave of evolutionary thinking that had previously swept through biology finally struck linguistics too: That year, Harvard cognitive scientist Steven Pinker and Yale psychologist Paul Bloom published a long article in *Behavioral and Brain Sciences* arguing that language must have evolved by natural selection. The Pinker-Bloom paper was "a kind of watershed," says linguist James Hurford the University of Edinburgh, U.K. "Suddenly it was OK to talk about evolution of language in Chomskyian circles."

Meanwhile, advances in brain imaging, neuroscience, and genetics have enabled a new contingent of researchers to go ever deeper into our brains and our biological past. For a long time, researchers treated language ability as some sort of "miracle," says neuroscientist Michael Arbib of the University of Southern California (USC) in Los Angeles. Now, he says, researchers are breaking that miracle down into a series of smaller, more manageable "miracles," involving disparate capacities such as the ability to imitate facial expressions or to string movements together. They're not fantasizing that the human brain at some point suddenly found that it could speak with the tongues of angels, he says; rather, it achieved a more modest state some researchers call "language readiness," which opened the door to further advances in linguistic ability.

Language origins are "certainly worth talking about now," says Hurford, who in 1996 launched the first of a series of biennial conferences on language evolution* that have grown steadily. Hurford's Edinburgh colleague Simon Kirby has documented the leap in interest with a citation search: The number of papers dealing with both "language" and "evolution" more than doubled from the 1980s to 1990s. (See also Book Review, p. 1299.)



Dangerous talk. Side view of human vocal tract shows that because of our lowered larynx, food and drink must pass over the trachea, risking a fall into the lungs if the epiglottis is open.

ILLUSTRATION: C. SLAYDEN/SCIENCE

Yet despite all the activity, the new lines of evidence remain indirect, leaving plenty of room for interpretation—and conflict. "If you want a consensus, you won't get it," says cognitive scientist Philip Lieberman of Brown University. With no fossils of speech, the origin of language remains "a mystery with all the fingerprints wiped off," says brain scientist Terrence Deacon of the University of California, Berkeley.

The long view

Archaeologists have identified various milestones in human behavior in the 5-million-year evolutionary void between animal communication and human speech, but there is no consensus on which achievements imply the capacity for language. For example, the first stone tools date to 2.4 million years ago; some researchers think this may indicate linguistic facility, but others argue that toolmaking is far removed from speech. Another possible starting point is 2 million years ago, when the hominid brain began a period of rapid expansion, including in the primary brain areas associated with producing or processing language—namely Broca's area in the left frontal cortex and Wernicke's in the left temporal lobe (see brain model, p. 1318).

As for actually producing the sounds of words, or phonemes, skeletal studies reveal that by about 300,000 years ago, our ancestors had become more or less "modern" anatomically, and they possessed a larynx located at the top of the trachea, lower than in other primates (see diagram). This position increases the range of sounds humans can make, although it also makes it easier for food going down the esophagus to be misdirected into the windpipe, leaving us more vulnerable than other mammals to choking. Such anatomy could have developed for no other purpose than speech, says Deacon.

Other possible milestones come from genetic studies. For example, researchers at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, reported last year that the *FOXP2* "speech gene," which affects both language and the ability to articulate (*Science*, 16 August 2002, p. 1105), was apparently a target of natural selection. This gene may have undergone its final mutation fewer than 100,000 years ago—and no more than 200,000 years ago—perhaps laying the groundwork for a new level of linguistic fluency.

Most researchers are inclined to the view that language gradually emerged over perhaps a couple of hundred thousand years (*Science*, 20 November 1998, p. 1455). But all we know for certain, says Pinker, is that fully developed language was in place by at least 50,000 years ago, when humans in Europe were creating art and burying their dead, symbolic behaviors that point unequivocally to fluent language.

* www.ling.ed.ac.uk/evolang

The motor route

Understanding when language emerged will probably have to await better understanding of how it emerged. In recent years many researchers have become increasingly attracted to the notion that changes in the brain's motor areas were crucial for language capability.

Although we tend to associate language first with sound rather than movement, speech may be better understood as a motor activity, says Deacon. Like other fine motor activities such as threading a needle or playing the violin, speech demands extraordinarily fine and rapid motor control. Elaborate movements of the larynx, mouth, face, tongue, and breath must be synchronized with cognitive activity.

Thus researchers are probing the links between language and areas in the brain that control gestures, either hand movements or the articulatory gestures of mouth and tongue. Linguist Robert Kluender of the University of California, San Diego, says explorations of gestures, including sign language, offer glimpses into what might have been the "intermediate behavioral manifestations" between animal communication and speech.

Many researchers think hand and face gestures offer behavior that is more analogous to speech than are animal vocalizations. In all other mammals, both breathing and articulation are directed by brain areas quite separate from those associated with human speech, notes Pinker. Lieberman argues that nonhuman primates engage in "a limited number of stereotyped calls" such as alarm calls and that they don't have the interactive or combinatorial quality of language. Apes' anatomy is such that they "could produce a [phonetically] reduced form of human speech," adds Lieberman. "But they don't." They're much better at signing, because apes' motor behaviors have more flexibility and are more involved in social interaction—through gaze, mouth and facial movements, and limb gestures—than their calls, Lieberman says.

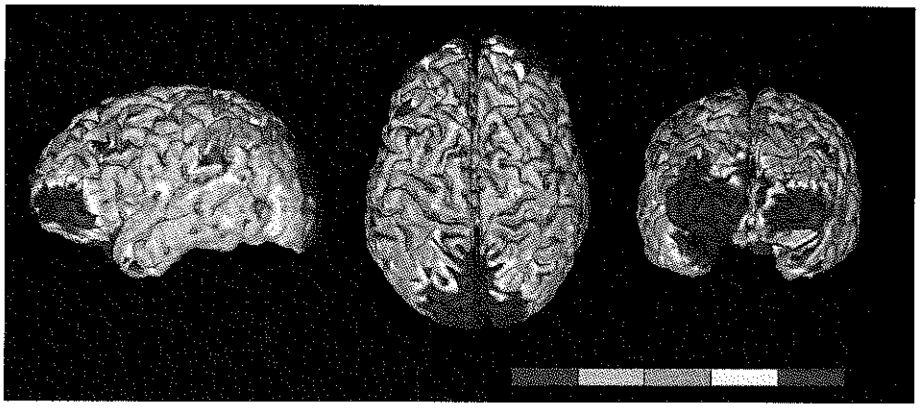
Lieberman argues that the crucial changes that laid the groundwork for language ability occurred in brain circuits connected with the basal ganglia, subcortical structures involved in movement. In his view the basal ganglia is the "sequencing engine" that makes combinations—both verbal and gestural—possible. As evidence he points to the fact that patients with Parkinson's disease, which disrupts the basal ganglia, suffer erosion of syntactical abilities, as well as problems with balance and movement.

Pinker's research, with cognitive scientist Michael Ullman of Georgetown University in Washington, D.C., lends weight to this view. They have shown that Parkinson's patients with basal ganglia damage have more trouble with regular verbs than with irregular ones. Conjugating a regular verb such as "walk," Pinker explains, is a combinatorial, sequential task that calls for adding the "ed" for past tense. But retrieving the past tense for an irregular verb such as "come" simply calls on long-term memory. Such tasks require other brain areas as well, but Lieberman argues that the basal ganglia are a common element in both movement and language disorders.

Indeed, although many other brain areas, including those responsible for articulation, hearing, planning, and memory, had to develop to support language, there is abundant behavioral evidence for an intimate connection between language and motor abilities, says Pinker. For example, psychologist David McNeill of the University of Chicago cites the case of a man who lost all sense of touch below the neck due to a strange virus. Although the man had to relearn the simplest movements, using cognitive and visual feedback to substitute for lost senses, he continued to gesture automatically when he spoke, even when researchers hid his hands from his own and listeners' view. "The hands are really precisely linked to speech articulation," says McNeill. "Gesture is not a behavioral fossil that was superseded by language but an indispensable part of language."

But not everyone is ready to dismiss the meaningfulness of animal calls, with differing views often dependent on a scientist's specialty. Primatologist Marc Hauser of Harvard, for example, believes that primate calls are better candidates for speech precursors than any gestures are. With chimp gestures, "nothing gives a suggestion of anything referential"—that is, having an explicit association with a concept or thing—he says. Primate alarm calls, in his view, "kind of look like words." For example, he cites work by psychologist Klaus Zuberbühler of the University of St. Andrews, U.K., who has reported that African Diana monkeys can modulate their alarm calls to indicate what type of animal (leopard or eagle) is threatening. Such sounds, says Hauser, "have a far greater ... connection to language than any discovery on nonvocal signals."

Many linguists, too, are unmoved by motor arguments, which they do not believe can explain how the brain developed syntax. "Motor organs are for muscular movements," says Derek Bickerton of the University of Hawaii, Manoa, and that puts them at the "end of the pipeline" of language production. "Whatever organizes motor movements is on a par with what organizes throwing movements," says Bickerton. "The purpose is to put things in a regular invariant sequence." That, he says, is very different from making sentences, which requires "putting



From ape to human. Magnetic resonance images of a bonobo brain are warped onto the shape of a human cortex, viewed from (left to right) the side, top, and front. Red and yellow areas in the temporal region (linked to language) and in the prefrontal and occipital regions had to be stretched the most to reach the human configuration, whereas blue areas are similar in apes and humans.

Hand and mouth. Chimps gesture with both face and hands to help express themselves.

CREDITS: (TOP TO BOTTOM) KARL ZILLES, PETER PIEPERHOFF, AND HARTMUT MOHLBERG/RESEARCH CENTER JÜLICH, INSTITUTE OF MEDICINE, GERMANY; NIGEL J. DENNIS/PHOTO RESEARCHERS INC.

things into an extremely plastic order determined by your conceptual structure."

Mirror, mirror

Despite such caveats, the motor-language connection continues to draw attention, in part because of a 1996 discovery that many see as the first hard data in years to bolster the theory. This is the so-called mirror neuron system found in monkeys' brains.

Mirror neurons' link to language depends on imitation, a skill largely unique to humans and considered vital to language. Although parrots and dolphins can do vocal mimicry, imitation is not as a rule a mammalian attribute: Even nonhuman primates do it poorly (contrary to the implication of the term "to ape"). But imitation is the way babies learn their first words. And it's the only way a common meaning can emerge for an abstract symbol, a phenomenon that linguists call "parity." "Imitation is the common thread for people writing about language origins," says neuroscientist Marco Iacoboni of the University of California, Los Angeles.

So researchers were excited when a team led by Giacomo Rizzolatti of the University of Parma, Italy, found what they considered a plausible antecedent for the human ability to imitate in the brains of monkeys. The team recorded electrical activity in macaques from 532 neurons in an area called F5, which is homologous to Broca's area in humans. Neurons in F5 are known to fire during monkeys' "goal-directed" hand and mouth movements—for example, when reaching for food.

What intrigued the researchers is that a subset of these neurons, which they dubbed mirror neurons, also became active when a monkey merely watched another monkey (or a human) perform the action. This finding "opened a whole new approach to the language evolution story," says Arbib of USC. "What would a mirror system for grasping be doing in the speech area of the brain?" The researchers concluded that these mirror cells form a system for matching the observation and execution of mouth and hand actions—the first steps toward imitation.

So far, mirror neurons have been found in only two brain areas in macaques, and the single-cell brain recording technique that revealed the macaque neurons isn't done on humans. But Iacoboni believes he has identified a similar circuit—"a core neural architecture for imitation"—in people. He combined the results of single-cell brain recordings in monkeys with functional magnetic resonance imaging in humans while they watched or imitated finger movements or facial expressions. Iacoboni says that in addition to Broca's, the circuit comprises an area in the superior temporal cortex (which overlaps with Wernicke's and has neurons that respond to face and body movements) and one in the parietal cortex, the homolog to the macaque area called PF, which combines visual and bodily information. "The neural architecture for imitation ... overlaps very well with well-known language areas in the human brain," says Iacoboni, who concludes that the dual-use nature of Broca's area in particular "suggests an evolutionary continuity between action recognition, imitation, and language."

Wired for imitation? Classic language areas—Broca's and Wernicke's (yellow)—overlap (orange) with areas critical for imitation (red).

Mirror neurons provide the "neural missing link" between movement and speech control, says Arbib. They also fit well with an old theory, the "motor theory of speech perception," developed in the 1950s by the late Alvin Liberman of Yale University's Haskins Laboratories. Psychologist Michael Studdert-Kennedy of Haskins Labs explains that when children imitate their first words, experiments have shown that they (unlike another imitator, the parrot) are guided by the "gestural" features of the sound—that is, by the actions of the mouth rather than by a sound's acoustic features. A well-known trick to demonstrate this is known as the McGurk effect: If you watch someone pronounce the syllable "ga" while listening to a recording of someone saying "ba," you will likely hear "da," a sound anatomically between the other two.

This means "you perceive speech by referring the sounds you hear to your own production mechanism," says Studdert-Kennedy. Humans, unlike other animals, are equipped with an intuitive sense of how their body parts correspond with those of others. Thus a small child knows how to raise its hand in response to a parental wave. "There's obviously a direct representation of your body in its body," says Studdert-Kennedy.

The theory developed new life when Studdert-Kennedy brought it to bear on questions of language evolution. Mirror neurons, he says, "for the first time provide an example of a direct physiological hookup between input and output": the observation of an action and its imitation. Indeed, Rizzolatti's group recently reported that the macaque has "audiovisual" mirror neurons: Some of the cells in F5 fire not only when a macaque watches a meaningful grasping action, but when it hears the sound of one, such as the sound of breaking peanuts (*Science*, 2 August 2002, p. 846). Arbib believes that mirror systems probably exist in other parts of the brain for many other behaviors.

He and others feel that mirror neurons offer the first concrete neurological evidence of abilities crucial to language, but it's a long way from a few firing neurons to speech. Some scientists think the potential significance of mirror neurons may be exaggerated. Macaques, after all, can't speak and they can't imitate either, notes Pinker. In his view, mirror neurons' "relevance to language is still pretty fuzzy."

The first syntax: words or waves?

Despite such drawbacks, mirror neuron theory is being invoked by both sides in the schism over whether the earliest language—that is, symbolic sounds or gestures connected by some sort of rules of syntax—used the voice or the hands.

Those who favor gestural origins, such as psychologist Michael Corballis of the University of Auckland, New Zealand, point out that mirror neurons are found in brain areas responsible for grasping. "I think it's extremely likely that language evolved in our early ancestors as a manual system, not as a vocal one," as far back as a million years ago, says Corballis. He notes that when robbed of speech, people quickly develop sign language, as has been shown by the case of a community of deaf Nicaraguans who created their own language.

Given the strong role of manual and facial gesture in speech and the relatively recent final mutation of the *FOXP2* gene, Corballis argues that "autonomous" speech may not have become fully developed until the cultural explosion beginning 50,000 years ago. The mirror system, he believes, reinforces his theory, because it apparently evolved first for manual control. It "probably picked up vocal and facial control quite late in hominid evolu-

CREDIT: MARCO IACOBONI, JOHN BACHELLER, AND ARTHUR TOGA/UCLA

tion," he says, as speech became the preferred modality for communication for various reasons, such as the need to free the hands for work or to talk in the dark.

But others believe equally strongly that even if movement and language are inseparable, language is primarily an oral, not manual, behavior. Psychologist Peter MacNeilage of the University of Texas, Austin, has developed a theory that monkey oral behaviors (not vocalizations) are precursors of human syllables, and he argues that the mirror neuron system—especially the recent discovery of neurons that respond to lip smacking and nut cracking—bolsters his ideas.

MacNeilage suggests that the brain's supplementary motor area (an area adjacent to the primary motor cortex that is important for motor memory and sequential movements) controls the physical constraints on vocal expression. The actions of chewing, sucking, and licking took on communicative content—a job for Broca's predecessor—in the form of lip smacks, tongue smacks, and teeth chatters. The next stage, says MacNeilage, was to give voice to these behaviors by bringing the larynx into play. This theory fits well with the fact that the unique sounds of click languages, which some speculate may have been the original mother tongue (see next story), do not use the larynx. Once the larynx was involved, a phonology—a set of sounds that could be combined in endless ways to form a large vocabulary—developed, and this in turn paved the way for the emergence of syntax.

"I don't believe manual gestural communication got to the point of the combinatorial phonology that I'm talking about, because if it did we'd still have it," says MacNeilage. In his view, if sign language had become that complex, there would have been no reason strong enough—the desire to talk in the dark notwithstanding—to cause a transition to vocal speech. "Nobody who argues that we went from sign to speech has given us an adequate translation theory," he says.

Others say the "which came first" debate is beside the point. "Evolution selected the ability to combine speech and gesture under a meaning," says McNeill. "The combination was the essential property"; neither gesture nor speech could have evolved without the other, he says. It doesn't matter which came first, agrees Zuberbühler: "Once an individual reaches a certain threshold in its cognitive sophistication, it will inevitably express itself in a sophisticated way," through any means at its command, he says.

The deepest questions—such as how humans became symbolic thinkers and developed "theory of mind," or awareness of others' thought processes—remain far from resolved. Researchers say one way to tackle them will be through ever-finer brain imaging technology so they can, as Bickerton puts it, "find out the flow chart for a sentence in the brain." Harvard's Hauser and colleagues believe that research in animals may identify behavioral analogs for "recursion": the ability to string words together in infinite hierarchical combinations. Arbib predicts that the discovery of other types of mirror systems, in both humans and animals, will help yield a better "taxonomy" of the language conundrum, especially if bolstered by computational modeling. But answers won't come all at once. "I see this as a process of gradual convergence. The problem space is shrinking" at long last, says Bickerton. "It will be solved when that space goes to zero, not when someone comes up with the killer solution."

—CONSTANCE HOLDEN

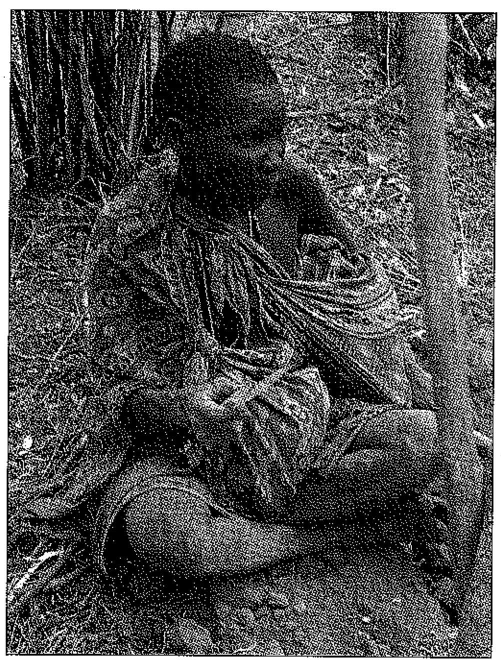
NEWS

The First Language?

Genetic and linguistic data indicate—but can't quite prove—that our ancient ancestors spoke with strange clicking noises

In the 1980 movie *The Gods Must Be Crazy*, a soda bottle falls out of the sky and lands among some strange-sounding Africans. Their excited chatter, punctuated by rapid-fire sucking and clicking noises, sounded intriguing but alien to audiences around the world. But a handful of studies of this seemingly esoteric language suggest that our early ancestors depended on such clicks to communicate. The latest linguistic work points to clicks as having deep roots, originating at the limits of linguistic analysis sometime earlier than 10,000 years ago, and genetic data suggest that click-speaking populations go back to a common ancestor perhaps 50,000 or more years ago.

Although the idea is far from proven, "it seems plausible that the population that was ancestral to all living humans lived in the savanna and used clicks," says vertebrate systematist Alec Knight of Stanford University. Knight estimates that today only about 120,000 people rely on these odd sounds. Even so, they are providing new insights into how humans evolved the gift of gab, particularly when researchers add up the results of different kinds of data. "There's a lot of mileage to be gained by cross-referencing linguistic, genetic, and archaeological data and theories," says Nigel Crawhall, a graduate student studying click languages at the University of Cape Town, South Africa.



All alone. Researchers ponder why the Hadzabe live so far from other click speakers.

CREDIT: NIGEL CRAWHALL

Clicks in context

Today clicks are part of typical conversation for about 30 groups of people, most from Botswana, Namibia, South Africa, and nearby. The only recognized non-African click language is Damin, an extinct Australian aboriginal language used only during manhood initiation ceremonies. Among African click speakers, daily conversations can be dominated by clicks, and sometimes verbal sounds drop out completely.

Adept tongue and inward air movements distinguish clicks from other nonverbal utterances. They are really just very strongly pronounced consonants, says Amanda Miller-Ockhuizen, a linguist at Cornell University in Ithaca, New York. Click speakers have click sounds in common, but they have different words and therefore very different languages.* Some researchers argue that click languages are far more different from each other than English is from Japanese.

But that diversity is only now being

* To hear click sounds, go to hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/index.html

appreciated. In the 1960s, the influential Stanford linguist Joseph Greenberg put all click languages under one umbrella, which he named the Khoisan language family after the two biggest groups included: herders known as Khoe and hunter-gatherers called San. Now, however, historical linguists are challenging Greenberg's classification, examining Khoisan with more stringent analytical methods and splitting it into several language groups. "It's been easy to say they are all in one family," says Bonny Sands, a linguist at Northern Arizona University in Flagstaff, "because nobody has gone and looked."

The latest work divides the Khoisan family into at least three geographically and linguistically distinct ones. And a few of these languages don't fit in any known families, Crawhall notes. For example, in 1995 Sands reexamined the grammar, meanings, and sounds of Hadzane, spoken by about 1000 Hadzabe people in north-central Tanzania, 2000 kilometers away from the majority of click speakers. She "proved that Hadzane cannot be shown to be related to any of the other families," says Crawhall. Rather, says Sands, linguistically Hadzane is unlike any other known language.

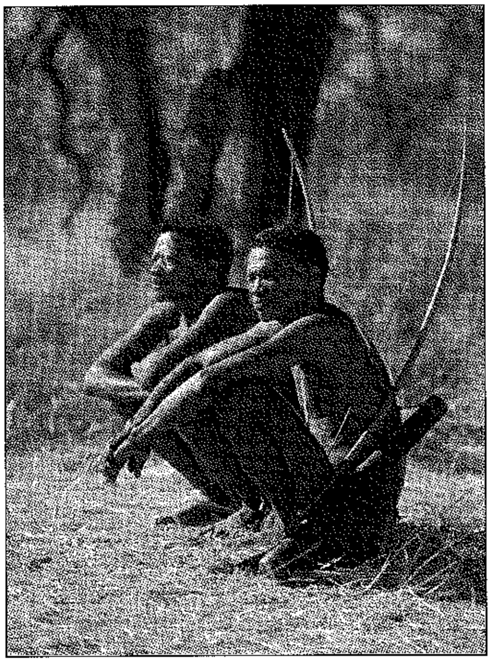
That suggests that either Hadzane had a separate origin from other click tongues or that it and other existing click languages derive from a very ancient protoclick language. Sands thinks that there have always been multiple click languages, but "if there was originally only one click family, it must be many tens of thousands of years old," she says. That's further back than linguistic studies can establish.

Tracking ancient populations

But genetic data on click speakers have also been streaming in, and these results can offer a glimpse into the more distant past. In 1991, one study hinted that Hadzabe were an ancient people based on the great diversity in their DNA; mutations accumulate over time, so diverse sequences imply an ancient population. Most recently, at a physical anthropology meeting last year, human geneticist Sarah Tishkoff of the University of Maryland, College Park, reported great diversity in the DNA of the Hadzabe and another click-speaking group in eastern Africa, the Sandawe.

The puzzling origins of these groups and their clicks intrigued Knight and Stanford anthropological geneticist Joanna Mountain. Last year, they decided to use genetic data to decipher the relationship between the isolated Hadzabe and the San in southern Africa. They thought that perhaps the Hadzabe had recently moved into Tanzania from the south, bringing clicks with them, or that the San had been part of a northern group that migrated south. "We expected a recent shared heritage, but the data indicated something opposite [from the recent origins] we expected," Knight recalls.

Knight, Mountain, and their colleagues examined mitochondrial DNA and parts of the Y chromosome from 49 Hadzabe and about 60 people from three other Tanzanian populations. They also gathered Y chromosome data from a San group, the Ju|'hoansi (also known as the !Kung) from Namibia and Botswana, and two non–click-speaking groups in central Africa.



Silent stalkers. !Kung hunters may use clicks while sneaking up on prey in the savanna.

Similar patterns in certain DNA segments indicate relatedness—and the Hadzabe and San turned out not to be closely related at all. The genetic sequences suggest that the two went separate ways very early on in their histories; neither group had migrated recently either northward or southward to bring clicks to the other. "The research suggests that the Hadzabe are the descendants of one of the first groups to split off" from an ancient pool of click speakers, says Crawhall.

Some researchers think the split between the Hadzabe and all other click speakers could have been as early as 100,000 years ago, but Knight puts it between 70,000 and 50,000 years ago. That's roughly the time frame proposed for the exodus of modern humans out of Africa, which some say might have been spurred by the development of language itself. But Knight warns that the dating is the most tentative part of their study.

Such an early origin for clicks appeals to linguist Michael Corballis of the University of Auckland, New Zealand, who has argued for years that 100,000 years ago, our only "words" were gestures: a flick of a finger, a twirl of the wrist, and so on (see p. 1316). "It may be that clicks themselves date back to a time when language was not autonomously vocal; they were a kind of [preverbal] way of adding sound," or a steppingstone to human speech, he says.

Knight thinks that only groups that retained ancestral hunting lifestyles continued to need clicks, and other click languages died out when early humans moved into new environments. That fits with evidence from living Hadzabe, who told Knight that when they hunt, they use clicks—and verbal talk disappears. Filmmaker John Marshall of Documentary Educational Resources in Watertown, Massachusetts, who has made dozens of films of click speakers, has noted this too. "I know from experience that using only clicks to communicate works well when stalking game," he explains. He and Knight suggest that whereas voices can spook animals, clicks mimic rustling grass, a typical sound on the dry savanna and one less likely to send game running.

Plausible as it all sounds, the theory of clicks as the first language is by no means proven. Even though Knight's work expands on Sands's ideas about the history of clicks, she's worried that Knight may be pushing his data too far. Genetic and language evolution don't necessarily go hand in hand. "The most he can say is that [the two] are correlative," she says. Thus there's no way to prove whether clicks made up the mother tongue, she argues.

Meanwhile, some researchers, such as linguistic historian Christopher Ehret of the University of California, Los Angeles, still stand by Greenberg's all-inclusive family for the click languages and downplay the genetic data. Furthermore, whereas most researchers insist that all clicks stem ultimately from the same ancestral tongue, Sands wonders whether clicks might have evolved several times, with Damin in Australia and Hadzane as examples. "Clicks are part of the normal language mechanism that people have," Sands notes, and children make clicks as they are learning to speak.

All agree that nothing can be settled without more work. Knight and Mountain are seeking DNA from more groups, and Sands and Crawhall are scrambling to bag more click languages for the linguists' portfolio. Sands worries that they can't work fast enough; one group has just 10 speakers left. But as the data stream in, Knight remains optimistic. "In the year 2000, we didn't know anything compared to what we know now," he says.

—ELIZABETH PENNISI

CREDIT: ANTHONY BANNISTER/GALLO IMAGE/CORBIS

NEWS

Speaking in Tongues

Researchers struggle to find a cohesive classification scheme to organize the world's thousands of languages

In the 1980s, after puzzling over the pronouns in a bewildering array of languages spoken in the Americas, linguist Joseph Greenberg of Stanford University thought he had made a breakthrough. By focusing on similarities in the meanings of pronouns and other words beginning with the same letters in Eskimo, Aleutian, Apache, and other native American tongues, he was able to organize mountains of data and classify all 2000 or so languages into just three groups. The achievement was particularly appealing because those groups roughly corresponded to genetic and archaeological evidence thought to indicate three large migrations into the Americas.

It was just the kind of work researchers long for. It helped make sense of the world's babel, classifying languages from the Arctic to Tierra del Fuego and from Greenland to Siberia. And it offered insight into human history with a simple scheme. Many researchers studying the peopling of the Americas eagerly embraced the work.

There was just one problem: Greenberg's work was wrong, says Lyle Campbell, a linguist at the University of Canterbury in Christchurch, New Zealand. "Antiscientific" and "crap," growls historical linguist Donald Ringe of the University of Pennsylvania in Philadelphia. The pair are among a cadre of historical linguists who argue that Greenberg's way of collecting and analyzing data led to spurious results and that the linguistic similarities he noticed could be due to coincidence rather than shared origin. Campbell and other native American linguists insist that there are upward of 150, rather than three, language families in the Americas.

Greenberg died in 2001, but deep schisms over how the world's languages are related live on. Greenberg's academic descendants continue to insist that they are able to trace languages further and further back in time, and in doing so piece together extended families and protolanguages ancestral to those families. But they are in a tug of war with others wary of the limits of linguistic clues, who content themselves with maps full of complexity and refuse to peer back further than 10,000 years. At that point, argue researchers such as Ringe and Campbell, languages had already diverged so much that establishing earlier kinships is difficult if not impossible. "We're seen as the wet blankets," says Robin Barr, linguist-in-residence at American University in Washington, D.C.

There are a few points of agreement, such as Greenberg's classification of Arabic, Egyptian, and Hebrew into the Afro-Asiatic language family (see map, p. 1322). But still, controversy prevails. Tracing the common ancestors of languages and grouping their descendants into ever larger collections is not for the faint-hearted. Most linguists rely solely on studies of words and grammar to date the split of one language into two, and the results are often anything but clear. And the addition of other types of data, such as genetic clues to migrations, tends to add more complexity than clarity. Genetic clues can reveal events older than those traced by linguistic studies, but it's not clear that genes accurately reflect what has happened to a language, warns Bonny Sands, a linguist at Northern Arizona University in Flagstaff.

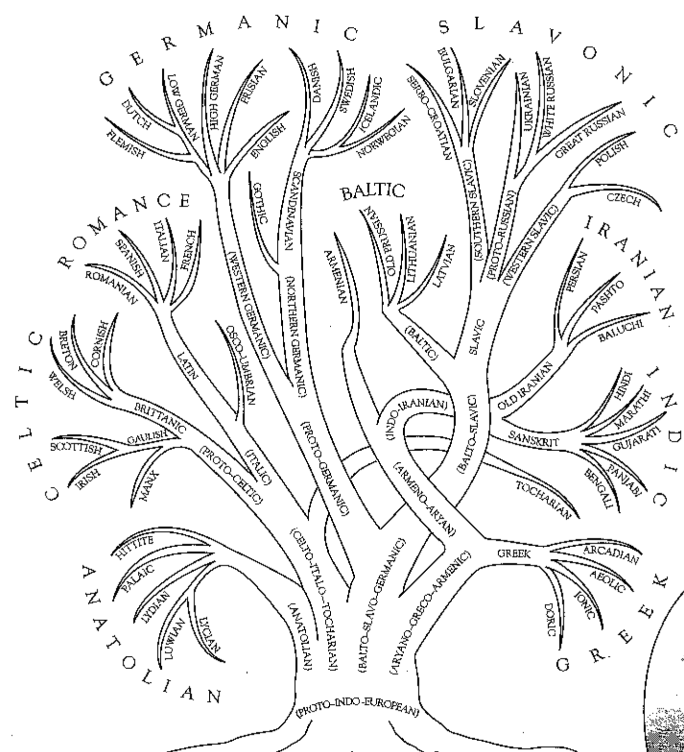
Pater to father

The notion that even seemingly distant languages are related dates back to the late 1700s, when British jurist Sir William Jones and others noticed similarities among Sanskrit, Latin, and Greek. Jones suggested that all these languages might stem from a common source. A century later, even before biologists were toying with the idea that organisms could be arranged in family trees, German linguist August Schleicher developed trees as a way to portray connections between languages. About the same time, others worked out a systematic way to establish kinship among languages—the so-called comparative method. These scholars and their successors laid the framework for modern historical linguistics.

To assess the degree of relatedness using the comparative method, investigators look for the "same" word across several languages, seeking systematic patterns of change. To take an example from one mother language, many Latin words that start with p begin with f in English. Thus "father" is parallel to *pater* and "fish" to *pisces*. This time-consuming approach requires an in-depth understanding of the languages involved, because researchers must identify and exclude words borrowed from another language and words that evolved in parallel; such parallel evolution often occurs in words that represent sounds, such as cuckoo, and words that come from baby talk, such as papa.

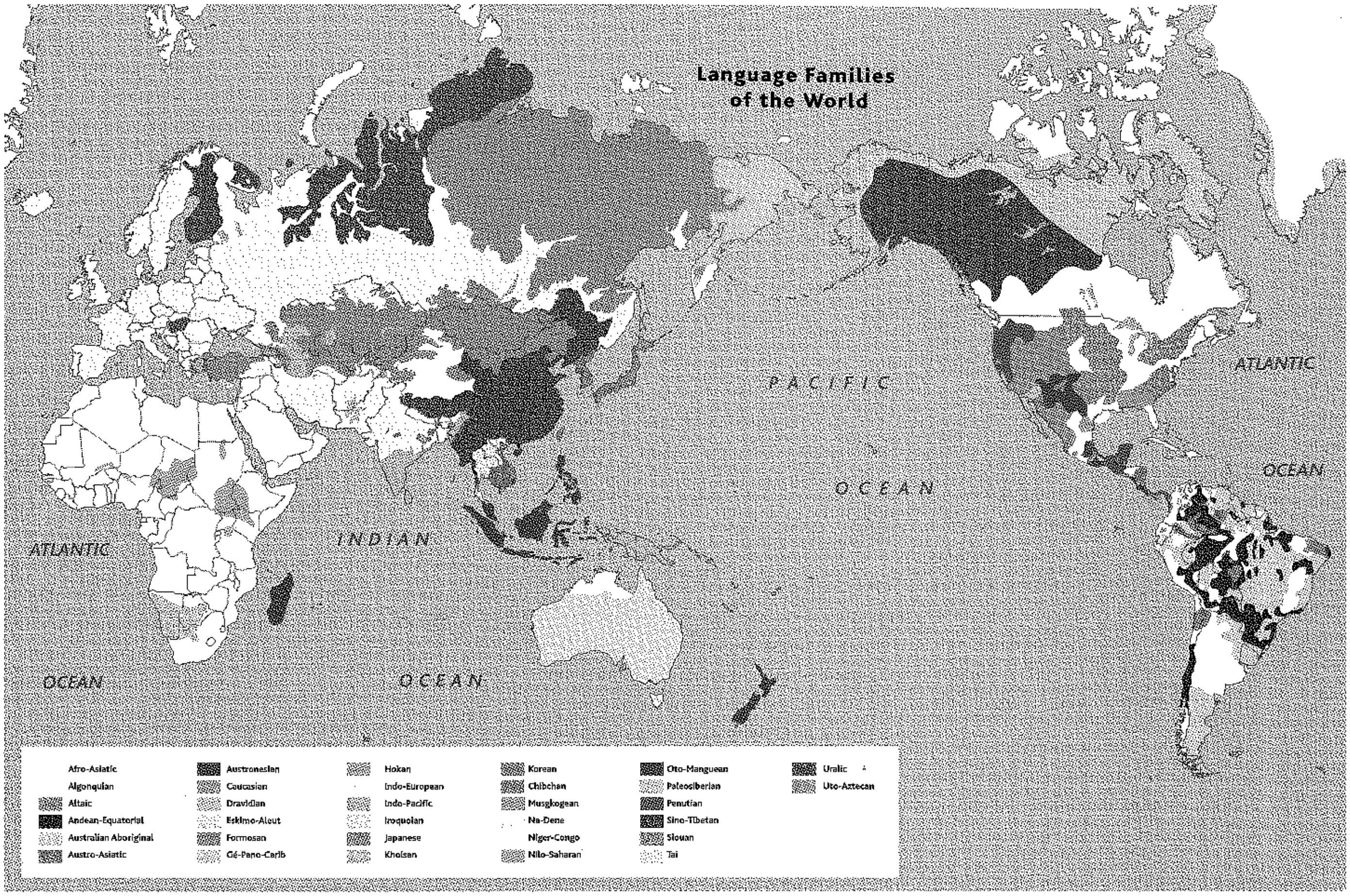
Nonetheless, says April McMahon, a linguist at the University of Sheffield, U.K., the comparative method "seems to work." For example, the method has revealed close connections among the nearly 150 languages of the Indo-European language family, some of them seemingly very distant, such as Hindi, Russian, English, and Iranian (see p. 1323).

But such work is slow, so Greenberg adopted a more broad-brush approach. He usually



Conceptual leap. August Schleicher (*right*) pioneered the idea that languages could be arranged in evolutionary trees.

CREDIT: (TREE) AFTER T. GAMKRELIDZE AND V. IVANOV, *THE INDOEUROPEAN LANGUAGE AND THE INDO-EUROPEANS* (MOUTON DE GRUYTER, 1984)



The talking world. This map shows one view of the complex distribution of language families, although several boundaries remain in dispute.

depended on a "core" vocabulary of a few hundred words, including numbers and pronouns, which he thought were likely to be basic words preserved as new languages arose; changes in those words should reflect the basic alterations involved in any given language. He compared those same words across many languages, an approach that didn't require culling words or even a very deep understanding of the languages under study. In the 1950s, he sorted 2000 African languages into four groups: Khoisan, which includes almost all click speakers; Bantu or Niger-Congo, a collection of 1436 languages; Afro-Asiatic, which includes Egyptian, Hebrew, Arabic, and others; and Nilo-Saharan, languages in Sudan and Central Africa. Many linguists applauded his effort at the time. "What he did was seen as really exciting because nobody had been able to do that," says Barr.

Greenberg eventually concluded that most of the world's 7000 languages can be grouped into about 17 "families." According to Stanford's Merritt Ruhlen, a longtime colleague of Greenberg, his work has made possible "the definitive classification of the world's languages."

But although the simplicity of Greenberg's interpretations is appealing, particularly for nonlinguists, his views in several regions have now come under closer scrutiny, and historical linguists have brought out the heavy guns. Even his acclaimed African study is being questioned, as researchers analyzing the vocabularies and grammar of click languages have concluded that not all belong together (see p. 1319).

Tracking linguistic ghosts

Undeterred by his critics, Greenberg moved into even more controversial territory. His ultimate goal was to group languages into ever bigger entities in search of a true mother tongue. Others had already promoted the existence of the so-called Nostratic superfamily, first proposed at the turn of the last century by Danish linguist Holger Petersen and later by Russian linguists. In their view, Nostratic included the language families of Indo-European, Uralic (spoken in northeastern Europe), North African and Semitic, Dravidian (from Southern India), and Altaic (from Central Asia).

The ghosts of superfamilies subtly haunt modern vocabularies, say Ruhlen and others. Consider the word five, which originated in words representing the hand. The word for fist was *penkwe* and later *pnkwstis* in ancient Indo-European and became *peyngo* in the Uralic root tongue and *p'aynga* in Altaic families such as Turkish and Mongolian. The Proto-Indo-European *penkweros*, finger, became *pente* in Greek, *quinque* in Latin, and *panca* in Sanskrit. From these variants, linguist Manaster Ramer of Wayne State University in Detroit has come up with the ancient Nostratic equivalent, perhaps spoken 12,000 years ago: *payngo* for fist.

Instead of Nostratic, Ruhlen and others support another, very similar superfamily, Eurasiatic, which was proposed by Greenberg 4 years ago and includes languages as disparate as English, Mongolian, Siberian, and Japanese. It encompasses a different subset of language families from Nostratic, although both include Indo-European and Altaic. In this analysis, Greenberg noted quite subtle similarities in the words across families. For example, the words for canines looked alike and often began with a similar sound: the ancestral version of Proto-Indo-European "dog" is thought to be *kwon*; "wolf" in Proto-Uralic was *küjnä*, and in the Russian tongue Gilyak, "dog" was *qan*. Greenberg, along with Ruhlen, also pushed for the recognition of other superfamilies, such as Austric, which would encompass Southeast Asian families.

But all these analyses continue to draw fire from researchers who say the data simply can't support peering so far back in time. "Languages have been evolving for so long that too much has been lost," says Ringe. Many of the similarities Greenberg noted, such as similar first letters, are so subtle that they may be circumstantial, says Ringe.

Going back to the distant past is also difficult because languages change at different rates. Icelandic children can easily read texts written many centuries earlier, but that ability doesn't hold for English students studying Chaucer. For these reasons, many historical linguists generally refuse to set dates for language changes before 5000 B.C.E.; for unwritten languages, they aren't even confident they can go back that far.

Written in the genes

Over the past few decades, genetic analysis has increasingly been brought to bear on this debate. Some studies have offered support for large and ancient family ties. Luigi Luca Cavalli-Sforza of Stanford University, for example, has shown that people with the same genetic makeup also tend to share their language. For example, the Bantu speakers whom Greenberg grouped together do indeed form their own distinct genetic group. In the Americas, Cavalli-Sforza identified four genetic groups; the one that branched off earliest from this line (and is therefore the one that is most distinct genetically) contains speakers of Na-Dene languages, whereas most of the rest use what Greenberg calls Amerind languages.

But as researchers probe the genetics-language connection further, "we've been finding much more noncorrespondence between linguistic trees and biological trees, suggesting a much more important role for language shift in the history of human populations," says Bernard Comrie, a linguist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. Even Cavalli-Sforza himself had found exceptions: Genetically, Ethiopians fit in well with other Africans, but their language is more akin to that of Middle Easterners, for example.

One reason for these differences is that genes and words don't follow the same timetable. It takes many generations for a new genetic variant introduced by invaders to take hold in a population, but a native language can be quickly replaced by that of the conquerors. "You cannot switch your genes, but you can easily switch your language," says Nigel Crawhall, a linguist at the University of Cape Town, South Africa. Nor does changing language require an extreme event such as war; trading and intermarrying can have a similar effect, as may have happened in Ethiopia.

Such a cautious attitude toward piecing together the linguistic past has its own critics, however. It's "like 17th century astronomers who said that the telescope wouldn't get any better," counters Christopher Ehret, a linguistic historian at the University of California, Los Angeles. "There's a group of people who throw up their hands without trying." He thinks Greenberg and DNA experts are prescient in piecing together scenarios that reach much deeper into the past.

Although complete answers remain elusive, the quest to understand linguistic history continues to generate excitement. Geneticists are constantly boosting the diversity of their databases, while linguists race to document disappearing languages. A few researchers are also finding new methods to analyze their data. Ringe, for example, is working with a computer scientist on linguistic models to build plausible scenarios for the evolution of language. McMahon and her colleagues are adapting sophisticated methods used by evolutionary biologists to trace organismal family trees to determine relationships among languages. Ringe expects that these new techniques will reverse a trend of diminishing numbers of historical linguists. After all, says Crawhall, "everyone is curious about how the languages relate to one another."

–ELIZABETH PENNISI

NEWS

Search for the Indo-Europeans

Were Kurgan horsemen or Anatolian farmers responsible for creating and spreading the world's most far-flung language family?

Around 6500 years ago, a group of seminomadic warriors arose on the treeless steppes north of the Black Sea. They herded sheep and goats, and they tamed the wild horse. Their language was rich with words reflecting their pastoral way of life. When one of their warrior-chiefs died, he was buried with great ceremony under a large earth mound called a kurgan. After about 1000 years of restless existence on the barren steppes, the story goes, these nomads went in search of new grazing land, riding out of their homeland between the Dnieper and Volga rivers armed with bows and arrows, spears, and bronze daggers. Over the next 2 millennia, the horsemen swept into eastern and central Europe, Anatolia, and much of western Asia, bringing their culture and colorful language with them. Before long, the hills of Europe and Asia echoed with the gallop of horses' hooves and the strongly enunciated vowels and consonants of a new language, which linguists today call Proto-Indo-European (PIE).

Colorful. Kurgan advocate Marija Gimbutas.

CREDIT: JOSEPH CAMPBELL AND MARIJA GIMBUTAS LIBRARY

The "Kurgan hypothesis," as this dramatic account of the spread of the Indo-European language family during the Early Bronze Age is known, was the dominant paradigm among linguists and archaeologists during much of the 20th century. It is most closely associated with the late Marija Gimbutas, an archaeologist at the University of California, Los Angeles, whose visions of prehistory were often filled with romantic pageantry. She argued that the Kurgans overrode existing matriarchal, Mother Goddess–worshipping societies, imposing their warrior religion as well as their patriarchal culture throughout Europe and western Asia. But the theory caught on for much more pragmatic reasons: It seemed to solve the long-standing mystery of the origins of Indo-European, a closely

Why Anatolia?

ÇATALHÖYÜK, TURKEY—While some archaeologists probe Kurgan earth mounds for the origins of Indo-European languages (see main text), others are focused on the even more ancient cultures of Anatolia, in modern-day Turkey. Here, some see a likely candidate for both the seeds of the Indo-European language and the gateway for the spread of agriculture into Europe some 8000 years ago. Recent research at a number of Anatolian sites has shown that this huge plateau was teeming with human populations, rich with art and culture, and home to the grains, cereals, and legumes key to the expansion of farming. "Everybody agrees that farming came to Europe from Anatolia," says Cambridge University archaeologist Colin Renfrew, chief partisan of the farming-dispersal model of Indo-European origins. "So Anatolia must be the point of departure [for languages too]."

Anatolia's importance in prehistory was firmly established in the 1960s, when excavations here at Çatalhöyük, near Konya in south-central Turkey, unearthed the largest early farming community ever discovered. This vast settlement of up to 10,000 people is also famous for its spectacular wall paintings (*Science*, 20 November 1998, p. 1442). Renfrew and other archaeologists note close cultural continuities in architecture, pottery, and figurines between Çatalhöyük and early Greek sites, which they see as an indication that Anatolia was the springboard for the introduction of agriculture into Greece and the rest of Europe.

Indeed, radiocarbon dating of archaeological sites in western Asia and Europe shows that farming first appeared about 11,000 years ago in the Near East and 10,000 years ago in central Anatolia, and that it had spread to Greece by 8000 years ago. Five hundred years later, farming villages began cropping up in the Balkans and central Europe. The wild ancestors of the seven "founder crops" harvested by the world's first farmers have all been traced to the region of southeastern Turkey and northern Syria (*Science*, 2 June 2000, p. 1602). The origins of einkorn wheat, for example, were recently located in the Karacadağ Mountains of southeastern Turkey, very close to Neolithic sites—dated as far back as 9600 years ago—where archaeologists have found seeds of both wild and domesticated einkorn.

Renfrew believes that Çatalhöyük, as the largest Neolithic settlement in central Anatolia, may have been at least one major source of Indo-European–speaking populations. Surveys in the area around the site suggest that the huge village may have been in close communication with smaller settlements nearby, says Ian Hodder, an archaeologist at Stanford University in California and current director of the Çatalhöyük excavations. "So you have all these people having to create a way of speaking with each other," Hodder says. Nevertheless, he cautions, until the linguistic debates over the origins of Indo-European are resolved, any attempts to correlate Neolithic cultures with the spread of languages may be little more than "clutching at straws." **–M.B.**

related group of 144 tongues that today are spoken on every continent. The family includes English as well as all of the Germanic, Romance, Slavic, Indian, and Iranian languages (see tree diagram).

In 1973, however, Cambridge University archaeologist Colin Renfrew proposed that the driving force behind the propagation of the Indo-European languages was not the fast gallop of horses' hooves but the slow adoption of farming. Renfrew argued that the gradual expansion of the agricultural way of life, which originated in the Near East some 10,000 years ago, carried the language family into new territories together with the seeds of wheat and barley. Because archaeologists widely agreed that farming had spread from Turkey to Greece and southeast Europe, Renfrew's "farming-dispersal hypothesis" pointed to the Anatolian plateau, which makes up most of modern Turkey, as a better candidate for the original Indo-European homeland (see sidebar above and Book Review, p. 1298).

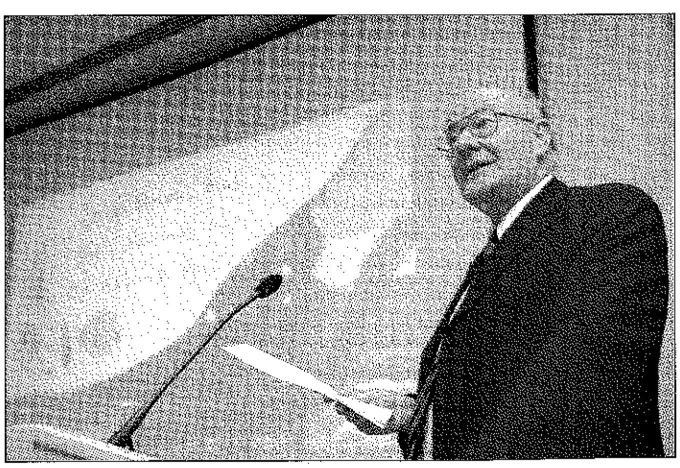
At first, most linguists and many archaeologists reacted with hostility to Renfrew's hypothesis, in part because they thought that it put the initial dispersal of Indo-European languages far too early. But in recent years, an accumulation of new evidence has considerably weakened support for the Kurgan hypothesis. Some archaeologists have challenged the notion that the Kurgans rode horses at all, and others have questioned the original linguistic analyses that put the Indo-European homeland north of the Black Sea. "Confidence in the Kurgan theory is waning," comments historian Robert Drews of Vanderbilt University in Nashville, Tennessee. "But," he adds, "the alternatives are not yet very attractive."

Indeed, Renfrew's analysis has certainly not swept the field. Although new and highly controversial dating of PIE, based on the techniques of evolutionary biology, supports a very ancient origin for the first appearance of the language family—8000 or more years ago—many linguists continue to insist that such early dates cannot be right. Wherever the first Indo-Europeans came from, they argue, reconstructions of the PIE vocabulary indicate that they could not have been the early farmers of Anatolia. "PIE was the language of a society which was very familiar with wheeled vehicles" and copper metallurgy, says Lawrence Trask, a linguist at the University of Sussex, U.K. "This obliges us to date the split of PIE no earlier than about 6000 years ago"—long after Anatolian farmers had dispersed.

Horses, wheels, and wool

Although archaeologists such as Gimbutas and Renfrew have played key roles in promoting the idea of an Indo-European homeland, linguists have often had the last word on the subject. In some cases, the similarities between words in diverse Indo-European languages are so striking that it doesn't take a trained linguist to spot them. The English word brother, for example, translates as *bhrater* in Sanskrit, *brathir* in Old Irish, *frater* in Latin, and *phrater* in Greek. But the field's most notable accomplishments consist of truly heroic efforts over many decades to recreate PIE, a technique sometimes called linguistic paleontology.

The results have caused many linguists to reject the farming-dispersal hypothesis. Take the

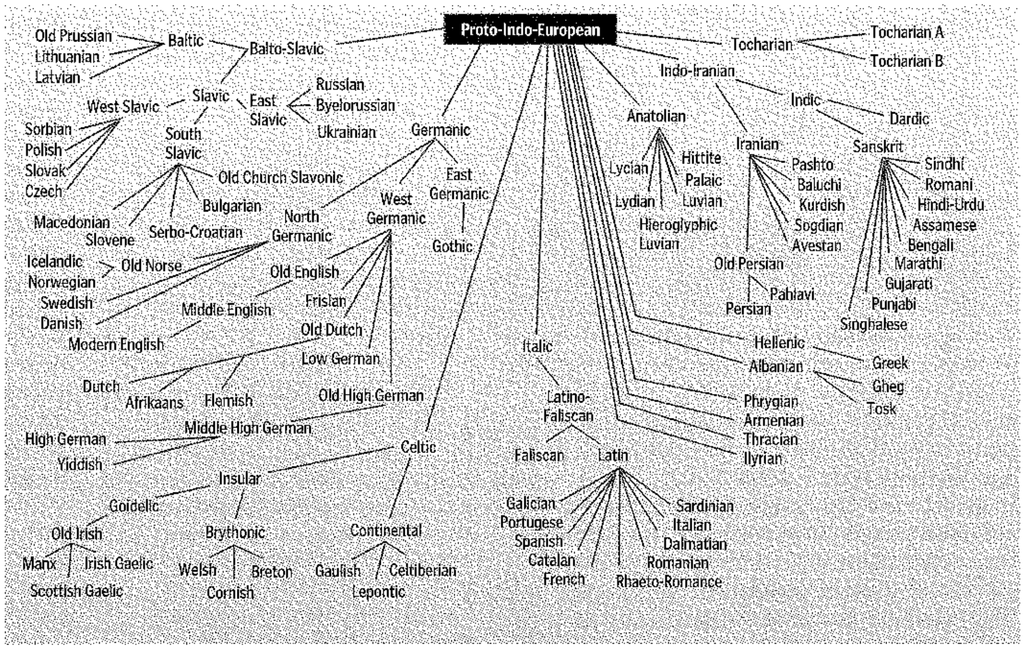


Spreading the word. Colin Renfrew argues that Anatolian farmers were the first Indo-European speakers.

CREDIT: EUROPEAN SCIENCE FOUNDATION

English word wheel, which has its roots in the PIE word $^{*}k^{w}ek^{w}los$. (The asterisk indicates that this word is a reconstruction and has not been actually found in written inscriptions; the superscript is a guide to the way that the word was pronounced.) Its equivalent is *cakras* in Sanskrit, *kuklos* in Greek, and *kukäl* in Tocharian, an extinct Indo-European language once spoken in western China.

"There is no evidence for wheels earlier than about 3500 B.C.," or 5500 years ago, says Bill Darden, a linguist at the University of Chicago. This means, he contends, that the Indo-European languages, all of which apparently share a common root for this word, could have diverged only after wheels were invented. Darden and other linguists have made similar arguments for the words for yoke, horse, and wool. For example, Elizabeth Barber, an archaeologist and linguist at Occidental College in Los Angeles, California, has traced the history of the word wool, which derives from the Indo-European root word **HwlHn-*. Yet when sheep were first domesticated in the Near East, around 9000 years ago, they were hairy rather than woolly. Only after about 6000 years ago did their coats take on the soft, curly pleats that today keep people warm during winter nights. "The farmers who moved from Anatolia to Greece [9000 years ago] did not know about wool, wheels, yokes, or horses," Darden insists. H. Craig Melchert, a linguist at the University of North Carolina, Chapel Hill, agrees: "I personally find the arguments of Darden and Barber quite irrefutable."



Say it in Indo-European. The 144 languages of this family descend from one ancient mother tongue.

CREDIT: ADAPTED FROM L. CAMPBELL, *HISTORICAL LINGUISTICS: AN INTRODUCTION* (MIT PRESS, 1999)

But Renfrew finds such literal readings of PIE too much to swallow. For example, he questions the assumption that the earliest roots of words such as wool and wheel necessarily had the same meanings as they do today. The PIE term for wool, Renfrew says, might originally have referred to the "fleece of the sheep" and only later came to refer to a luxurious coat. Likewise, the root for wheel may derive from earlier words meaning "to turn." And when it comes to the horse, Renfrew and his supporters believe that new evidence pulls the legs right out from under the Kurgan hypothesis.

Gimbutas and other archaeologists had claimed that the Kurgans domesticated the horse more than 5000 years ago, then quickly trotted off to conquer distant lands. Meanwhile, linguists had long been impressed with how rich PIE seemed to be with words for animals such as sheep, cattle, pigs, dogs, and horses, which were abundant on the steppes north of the Black Sea 5000 years ago. (The reconstructed PIE word for horse is **Hek̂wos*, which led to *aśvas* in Sanskrit, *equus* in Latin, and *eoh* in Old English.)

Beginning in the 1960s, excavations at a Kurgan site called Dereivka, located on a tributary of the Dnieper River in modern-day Ukraine, turned up supposed "bit wear" on premolar teeth and horse remains buried with other domestic animals such as sheep and cattle. These were taken as evidence of the earliest known domestication of the horse and dated to about 5500 years ago, right around the time the Kurgans supposedly began to ride forth from their homeland. Recently, for example, archaeologist David Anthony of Hartwick College in Oneonta, New York, concluded from experiments conducted on modern horse teeth that what he interprets as wear marks on the Dereivka horses would have required at least 300 hours of riding with a hard bit. Similar claims have been made for the site of Botai in Kazakhstan, occupied more than 5000 years ago and under excavation since the early 1990s.

But archaeologists disagree about the strength of much of this evidence. Also beginning in the 1990s, Cambridge University archaeologist Marsha Levine carried out a series of studies of horse bones from Dereivka, Botai, and other sites. Levine concludes that there is no credible evidence for horse domestication, and especially for horse riding, at these sites—nor, indeed, at any archaeological site before about 4000 years ago. When Levine examined the horse bones from Botai, for example, she found that the thoracic vertebrae completely lacked the kinds of bone pathology that modern horses suffer when they are ridden. As for bit wear, Levine counters that only a minority of the premolars from the Dereivka and Botai horses show the marks that Anthony interprets as bit wear, and she argues that these marks could have other causes.

And even if horses were mounted as long ago as Anthony and others have claimed, Renfrew argues that there is no direct archaeological evidence that "horses were … ridden for military purposes before about 3500 years ago. This is simply too late to coincide with Indo-European origins." Renfrew also points out that one key assumption long made by linguists—that the horse was not known at all in the Near East—has been overridden by recent evidence of butchered wild horses at the 9000-year-old farming village of Çatalhöyük in south-central Anatolia. This means, he says, that a hypothetical Kurgan homeland "is not a very good alternative" to his Anatolian farming-dispersal hypothesis, although he concedes that this alone "does not mean that the farming hypothesis has to be right."

Genes, words, and trees

Many experts had hoped that the debate between the Kurgan and Anatolian hypotheses might be resolved by studies tracing the genetics of modern Europeans. Indeed, research during the 1970s and 1980s, led by geneticist Luigi Luca Cavalli-Sforza of Stanford University in California and others, gave strong initial support to the notion that a large popula-

Seeking a homeland. The first Indo-Europeans may have been horse lovers from Dereivka or farmers from Çatalhöyük.

tion migrated out of Anatolia during the early days of farming about 8000 years ago (*Science*, 7 July 2000, p. 62). But in more recent years, geneticists have found that this picture is much more nuanced than originally thought and that the original hunter-gatherer populations in Europe may also have contributed significantly to the modern gene pool. As a result of these complications, most partisans in the debate, including Renfrew, believe it is premature to try to directly correlate genetic evidence for population movements with the spread of Indo-European languages (*Science*, 25 April 2003, p. 597).

Although the contribution of genetics to the debate has so far been disappointing, that has not stopped evolutionary biologists from jumping into the fray. In the 22 July 2003 issue of the *Proceedings of the National Academy of Sciences*, geneticist Peter Forster of Cambridge University and linguist Alfred Toth of the University of Zürich in Switzerland reported on an attempt to apply the mathematical and computer techniques of molecular phylogeny —which biologists use to reconstruct evolutionary trees of living organisms—to figure out when the Celtic languages of Europe first went their separate ways. In the process, they also came up with a date for the first divergence of the Indo-European family that is too early for the Kurgan theory but consistent with the Renfrew hypothesis.

Forster and Toth used bilingual Gaulish-Latin inscriptions to date the first split of Celtic from the rest of the Indo-European languages. (Gaulish was the version of Celtic spoken in France.) Their best guess came out at around 5200 years ago, whereas the Kurgan hypothesis puts this split—which occurred on the Atlantic coast, thousands of kilometers to the west of the Kurgan homeland—far later. And when Forster and Toth extrapolated their results backward to the common root of all Indo-European languages, they came up with a date of about 10,100 years ago, plus or minus 1900 years, for the first spread of Indo-European into Europe.

Those dates roughly match a more systematic attempt to use the methods of molecular phylogeny to test the Kurgan and Anatolian hypotheses, reported in the 27 November 2003 issue of *Nature* by evolutionary biologists Russell Gray and Quentin Atkinson of the University of Auckland, New Zealand. Their best estimate for the initial split came out at about 8700 years ago, which also coincides very closely with the first spread of farming from Anatolia into Greece.

However, many linguists remain unconvinced by such analyses, questioning the relevance of evolutionary biology techniques to linguistic problems (*Science*, 28 November 2003, p. 1490). "There is no reason whatsoever to assume that vocabulary would behave the same way that organisms do," says Alexander Lehrman, a linguist at the University of Delaware in Newark.

At the very least, the new evidence is making partisans on both sides of the debate think twice about their assumptions. And although a meeting of the minds is hard to envision in this contentious field, there may be grounds for compromise. Gray and Atkinson identified a rapid divergence of languages around 6500 years ago that gave rise to the Romance, Celtic, and Balto-Slavic language families. Because this date matches the first evidence for Kurgan occupation of the Black Sea steppes, Gray and Atkinson say, both camps could be partly right: The farmers spread PIE initially, but the Kurgans spurred the later burst of languages. "There is no need to set up the Kurgan and farming hypotheses at variance with one another," says April McMahon, a linguist at the University of Sheffield, U.K. "But sadly, this is something that [people] have a tendency to do."

–MICHAEL BALTER

NEWS

From Heofonum to Heavens

Ancient texts and computer simulations help linguists explore how words and grammar evolve over the time scale of centuries

If a modern-day priest were to chance upon an 11th century manuscript of *The Lord's Prayer* in English, he would need the Lord's help to decipher its meaning. Much of the text would be gobbledygook to him, apart from a few words that might have a recognizable ring, such as heofonum (heavens) and yfele (evil). And even after a word-for-word translation, the priest would be left with the puzzling grammatical structure of sentences like "Our daily bread give us today."

Although researchers generally think of languages as having evolved slowly over many millennia, language change occurring over time spans of a few centuries has confounded scholars since medieval times. After trying to read a 600-year-old document, the first known printer of English works, William Caxton, lamented in 1490, "And certainly it was written in such a way that it was more like German than English. I could not recover it or make it understandable" (translated from Old English).

The comparative analysis of such texts is the closest that researchers can get to tracing the evolutionary path of a language. By studying the evolution of words and grammar over the past 1200 years of recorded history, linguists hope to understand the general principles underlying the development of languages. "Since we can assume that language and language change have operated in the same way for the past 50,000 years, modern language change can offer insights into earlier changes that led to the diversification of languages," says Anthony Kroch, a linguist at the University of Pennsylvania in Philadelphia.

Since the mid-19th century, that hope has driven scholars to document a variety of grammatical, morphological, and phonological changes in French, English, and other languages. In the past 3 decades, more and

MAP: A. STONEBRAKER/SCIENCE

more theoretical and historical linguists have turned their attention to analyzing these changes, and sociolinguists have explored the social and historical forces at work. Now researchers in the growing field of computational linguistics are using computer models of speech communities to explore how such changes spread through a population and how language changes emerge in multilingual populations.

The simulations are infusing precision into the study of a phenomenon once thought to be the exclusive domain of humanistic inquiry. "Computational modeling of language change is in its infancy, but it is already helping us to reason more clearly about the factors underlying the process," says Ian Roberts, a linguist at the University of Cambridge, U.K.

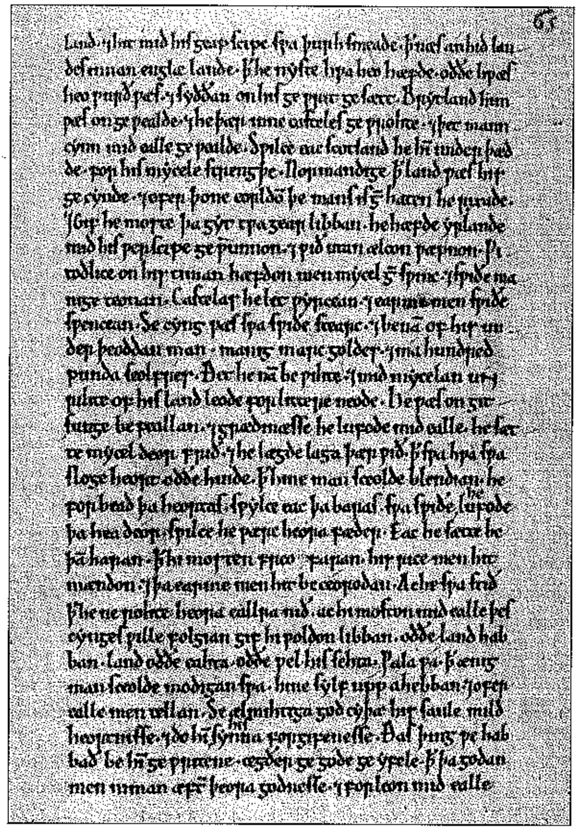
Voice of the Vikings

Linguists view language change as something of a paradox. Because children learn the language of their parents faithfully enough to be able to communicate with them, there seems no reason for language to change at all. But historical texts show that change is common, although the trajectory and rate of change may be unique for any given language. In the 10th century, to consider a classic example, English had an object-verb grammar like that used today in Modern German, requiring sentence constructions such as "Hans must the horse tame." By 1400 C.E., the English were using the familiar verb-object grammar of "Hans must tame the horse." French underwent a similar change before the 16th century, whereas German retained its basic grammar.

To find out why such changes happen, researchers explore the historical circumstances surrounding them. In the past few years, based on a comparative analysis of religious texts from northern and southern England, Kroch and his colleagues at the University of Pennsylvania have suggested that northern English was transformed during the 11th and 12th century as Viking conquerors married native Anglo-Saxon women, who spoke Old English. The resulting bilingual households became crucibles for linguistic change. For example, whereas Old English had distinct verb endings to mark differences in person, number, and tense, the speakers of what is now called Early Middle English began using simpler verbs—perhaps because the Scandinavians had difficulties keeping track of all the verb forms—and settled on a simplified system closer to what we use today.

In the absence of invasions and other external influences, languages can remain stable for long periods. Japanese and Icelandic, for instance, have hardly changed since 800 C.E. But researchers point out that isolation does not guarantee the status quo; grammatical shifts can also be triggered by internal forces such as minor changes in the way a language is spoken.

French is a case in point. In the 16th century, the language changed from a system in which the verb always had to be in second place (known as a verb-second structure) to one in which the verb (V) could be in any position as long as it came after the subject (S) and before the object (O); Modern French and Modern English both have this SVO structure. For example, "Lors oirent ils venir un escoiz de tonnere" (Then heard they come a clap of thunder) became "Lors ils oirent venir un escoiz de tonnere" (Then they heard come a clap of thunder). Roberts, who documented the transition by comparing a representative text from each century between the 13th and the 17th, believes that the change arose because speakers of Middle French reduced the emphasis on subject pronouns—"they" in this example—to the point where children learning the language barely heard the pronouns. Roberts inferred this decline in phonetic stress from usage changes in the written language. For example, subject pronouns were earlier used with modifiers, such as "I only," but later they did not carry such modifiers. The result of this reduced emphasis, says Roberts, "was that for sentences beginning with a subject pronoun, the verb sounded like the first word of the sentence to the listener." That ambiguity dealt a fatal blow to the verb-second rule, paving the way for the emergence of an SVO grammar.



Great-great-grandfather's English. This poem, written in 937 C.E. about a battle, shows the rapid evolution of English.

John the book buys

But a new grammatical feature cannot emerge overnight. For a variant such as an innovative construction by a single speaker or a novel form of syntax produced by a new adult learner to become part of the language, it must get picked up by other speakers and be transmitted to the next generation. Historical texts show that it can take centuries for a change to sweep through the entire community.

David Lightfoot, a linguist at Georgetown University in Washington, D.C., says the key to understanding large-scale linguistic transformation lies in the link between the diffusion of novel forms through one generation and large grammatical shifts occurring across generations—changes he calls "catastrophic." And this link, according to him and many others, is language acquisition. Children may simply carry forward a variant that arose in the preceding generation. But more significantly, says Lightfoot, children may themselves serve as agents of change "by reinterpreting a grammatical rule because of exposure to a variant during their learning experience." As adults, they may end up using a somewhat different grammatical system from that of their parents. Repeated over generations, this can lead to a dramatic makeover in the language.

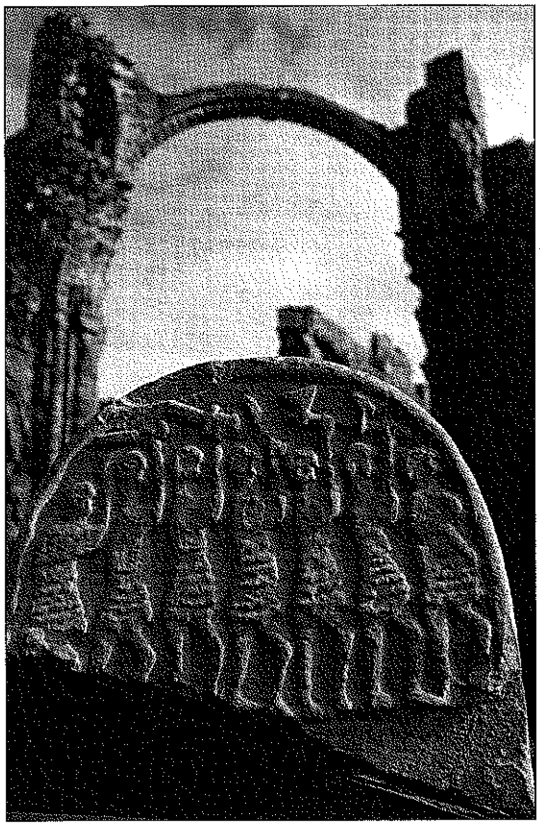
Computational linguists such as Partha Niyogi of the University of Chicago have built computer models to understand the dynamics of such evolution. Their goal is to map out the relationship between learning by the individual and language change in the population, which Niyogi calls the "main plot in the story of language change."

In one of the first attempts to unravel that plot's outline, Niyogi and Robert Berwick, a computer scientist at the Massachusetts Institute of Technology, came up with a class of models simulating the transmission of language across generations. They started out by considering a virtual population with two types of adult speakers. The first type uses one set of grammatical rules—say, one that, like English, mandates a verb-object order for all constructions, generating sentences such as "John buys the book," and "I know that John buys the book." The rest of the speakers use a different grammar, for example one similar to German, in which the first verb goes in the second position but the second verb goes after the object. The speakers of the second grammar produce some sentences exactly like speakers of the first—"John buys the book"—but they also produce other kinds of sentences such as "I know that John the book buys." The researchers spelled out a learning algorithm for children in this population, providing each learner with logical steps for acquiring grammatical rules from linguistic encounters with adults.

Following the linguistic behavior of this virtual community over generations led Niyogi and Berwick to a startling conclusion. They found that contrary to expectation, the population does not inevitably converge on the grammar spoken by the majority, nor on the simpler of the two grammars. Instead, the winning grammar is the one with fewer grammatically ambiguous sentences like "John buys the book," which, although simple, might be analyzed as belonging to either grammatical type. In other words, if minority speakers consistently produce a smaller proportion of grammatically ambiguous sentences as compared to the majority, the population will over time shift completely to the minority grammar.

Niyogi, who first presented the work at the International Conference on the Evolution of Language at Harvard in April 2002 and is publishing it in a forthcoming book, says the finding makes it possible to imagine how a grammatical variant spoken by a handful of individuals might replace an entrenched grammar. It's conceivable for the variant to pose no threat to the established grammar for many generations, he says, until the proportion of grammatically ambiguous sentences produced by speakers of the variant drops below the corresponding proportion for the dominant grammar. "For instance, sociocultural factors might change the content of conversations among minority, English-type speakers in a way that they stop using single-clause sentences like 'John buys the book,' " says Niyogi. That would make their speech more complex—but less grammatically ambiguous. Then learners would hear a higher proportion of the multiple-clause, uniquely English constructions in English speech than they would hear uniquely German constructions in German speech. This would make them more likely to infer the English grammatical system from what they heard, even though their overall exposure to German and even uniquely German constructions would be greater. Suddenly, the mainstream German grammar would become unstable and the English grammar would begin to take over.

"That a little good info should be able to trump a lot of bad [ambiguous] info makes sense," says Norbert Hornstein, a linguist at the University of Maryland, College Park, who sees the mechanism of change suggested by Niyogi as "a good fit with our understanding of language acquisition." He says it provides a possible explanation for how small local changes—for instance, a simplification of the verb system by mixed households in 13th century northern England—may have spread through the entire population. Confirming this account of change would require testing computational models with real-world data such as the proportion of specific syntactical forms in historical texts, assuming written language to be a faithful impression of speech. Niyogi admits that the task could take years.



Lasting legacy. Viking marauders, like those depicted in this English sculpture, left their mark on the English language.

In a broader sense, however, researchers have already validated the computational approach by matching the outlines of models to real-world situations. For example, University of Cambridge linguist Ted Briscoe modeled the birth of a creole, a linguistic patois that arises from prolonged contact between two or more groups. He specifically considered Hawaiian English, which developed between 1860 and 1930 through contact between Europeans, native Hawaiians, and laborers shipped in from China, Portugal, and other countries. Briscoe's simulation started out with a small but diverse group of speakers and factored in the periodic influx of adult immigrants. He found that a population with the right mix of children and new adult learners converged on an SVO grammar after two generations. That matches empirical studies showing that many features of Hawaiian Creole, including an SVO word order, did not stabilize until the second generation of learners.

Salikoko Mufwene, a sociolinguist at the University of Chicago, says that a detailed picture of mechanisms of language change could emerge if computational researchers succeed in modeling very specific contexts. For instance, he says, modeling spoken exchanges on a homestead of eight Europeans and two African slaves could help illuminate the linguistic evolution of the larger community. "The two Africans in this example are likely to be so well immersed that after a few months they would be speaking a second language variety of the European language. Say one of the Africans is a woman and bears a child with one of the white colonists. The child is likely to speak like the father because the father's language happens to be dominant at the homestead. Growing up, this child will serve as a model for children of new slaves," explains Mufwene. "Nonnative speakers will exert only a marginal influence on the emergent language of the community," in this case the native European variety.

But if the population increases significantly through a large influx of new slaves, he says, the dynamics of interaction change, and more adult nonnative speakers of the European language serve as models. Children now have a greater likelihood of acquiring some of the features spoken by adult nonnatives and transmitting them to future learners; over time, a new variety of the European language will emerge.

Detailed modeling along these lines, Mufwene says, could unveil the significance of factors that researchers may have missed, such as the pattern of population growth and the pace of demographic shifts. "Even without real-world number crunching," he says, "the exercise would suggest what questions we should be asking and what kinds of evidence we should be looking for."

–YUDHIJIT BHATTACHARJEE

CREDIT: TED SPIEGEL/CORBIS

The Future of Language

David Graddol*

The world's language system is undergoing rapid change because of demographic trends, new technology, and international communication. These changes will affect both written and spoken communication. English may not be the dominant language of the future, and the need to be multilingual will be enhanced. Although many languages are going extinct, new ones are emerging in cities and extended social groups.

We are living through an extraordinary moment of linguistic history. The world's language system, having evolved over centuries, has reached a point of crisis and is now rapidly restructuring. We will experience some decades of rapid, and perhaps disorienting change, after which a new linguistic world order will emerge. Precise predictions may be difficult, but the general shape of things to come is clear (*1*).

The Demographic Future

Global demography is one cause of the language crisis. The world's population rose rapidly during the 20th century, but the major increase took place in less developed countries. This trend, decade on decade, is transforming the global "league table" of languages, as based on native speaker numbers. The "top 10" languages at the end of the 20th century (Table 1) are not representative of the estimated usage of young people in 2050 (Table 2).

Estimating native speaker numbers for the larger—and otherwise best documented—languages such as English is surprisingly difficult. The numbers provided in Tables 1 and 2 are based on United Nations population projections and estimates of the linguistic demography of each country—a technique that is approximate but that allows principled projections of future language usage.

One perhaps unexpected trend is a relative decline of English, as projected in Fig. 1. In the mid-20th century, nearly 9% of the global population grew up speaking English as their first language, but that proportion is declining—toward nearer 5% by 2050 (*2*).

Chinese (whether one counts only Mandarin or all Chinese dialects, which share a common writing system) is well established as the world's largest language (in terms of native speakers), and its position will remain unchallenged. The next four major languages, however, are gradually converging and are likely to be equally ranked by 2050, with Arabic rising as English declines. But the combined "market share" of these larger languages taken together is unlikely to change much over the coming decades. It is the languages of the next rank—such as Bengali, Tamil, and Malay—which are growing most rapidly.

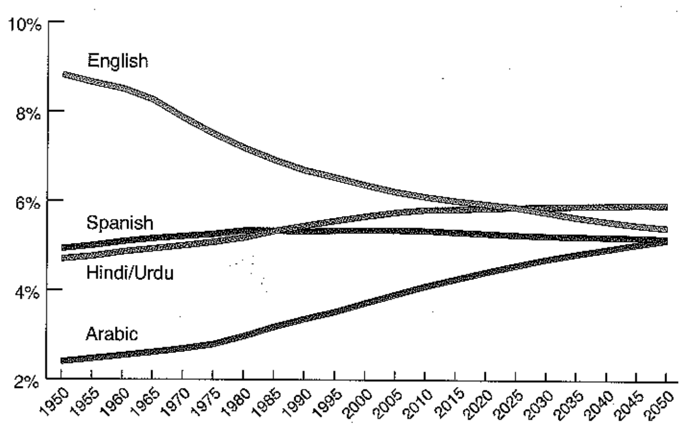


Fig. 1. The changing percentage of the world's population speaking English, Spanish, Hindi/Urdu, and Arabic.

The Future of Diversity

While a few languages compete for position at the top of the world hierarchy, there is devastation at the base. Most linguists agree that roughly 6000 languages exist in the world today. Yet 90% of these may be doomed to extinction, with much of this loss happening in the coming century. We may now be losing a language every day.

Public awareness of such language loss will grow, as will an appreciation of its wider implications. Regret at the passing of quaint and linguistically interesting languages may be replaced by concern about their strategic and economic importance. In 2001 the United Nations Environmental Programme (UNEP) concluded that "Losing a language and its cultural context is like burning a unique reference book of the natural world" and linked indigenous languages with "a vital understanding of sustainable land management and of cultivation practices which exploit diversity (*3*)."

However, while we lose older, rural languages, new urban hybrid forms may help maintain global diversity. Cities are places where languages mingle and where language change speeds up. And the fast growing urban areas of the world are breeding grounds for new hybrid languages—just as hundreds of new forms of English have already been spawned around the world (*4*).

Paradoxically, cities of the future will also allow immigrant languages to survive. Ethnic minorities often now belong to diasporic communities, within which members travel, watch the same films and satellite television channels, and communicate daily by telephone or e-mail. Everywhere, the social identities and networks that languages reflect and construct are becoming dispersed and less geographically tied. We can expect the continued decline of traditional geographically based dialects.

The End of Modern Languages

Many of these trends will challenge our sense of what is normal in language matters, shaped as it has been by a centuries-long experience of modernity. Modernity arose from complex historical factors including the emergence of sovereign nation states, capitalist societies, the protestant reformation in northern Europe, and the development of printing, which disseminated identical copies of standard texts. Modernity also gave us "modern languages," each a national language that has benefited from centuries of development in its grammatical and lexical resources.

In English, for example, the "national language project" began with literature in the 16th century (poets and dramatists such as Dryden and Shakespeare attempted to remedy the defects of English as compared with Latin and Greek); science was added in the 17th century (Sir Isaac Newton published first in Latin, later in English); dictionaries and grammars were created in the 18th century (Samuel Johnson); and the 19th century brought corporate affairs, modern advertis-

The English Company (UK) Ltd., 2 Western Road, Wolverton, Milton Keynes, MK12 5AF, UK.

*To whom correspondence should be addressed. E-mail: david@english.co.uk

Table 1. Estimates of numbers of native speakers globally in 1995 for the top 10 languages (*1*).

Language	No. of native speakers (millions)
1. Chinese	1113
2. English	372
3. Hindi/Urdu	316
4. Spanish	304
5. Arabic	201
6. Portuguese	165
7. Russian	155
8. Bengali	125
9. Japanese	123
10. German	102

ing, international diplomacy, and many other new forms of communication (*5*).

But the whole modernity project may now be unraveling, taking us into new linguistic landscapes. The "old" national languages are losing functionality as much communication—economic, cultural, and political—becomes international. Swedish, like many smaller European languages, is now positioned more as a local language of solidarity than one for science, university education, or European communication.

Big languages like English, meanwhile, have lost armies of linguistic gatekeepers who used to ensure that only the language of a social elite—sanitized by copy editors—reached public consumption. A combination of new technology, new skills [anyone can print a magazine or publish a blog (Web log)], changing public attitudes to correctness, and economics of publication (most copy editors are now freelance) have led to "destandardization." Written language now much more closely reflects the norms of speech. Dictionaries include the latest slang expressions because they appear in newspapers. Is e-mail best thought of as spoken language written down? Or as a new kind of informal writing?

A Multilingual Future

Any look into the future must entertain the idea that soon the entire world will speak English. Many believe English will become the world language to the exclusion of all others. But this idea, which first took root in the 19th century, is past its sell-by date. English will indeed play a crucial role in shaping the new world linguistic order, but its major impact will be in creating new generations of bilingual and multilingual speakers across the world.

The growth of Spanish in the United States can be understood as part of a much wider global trend toward bilingualism. In Europe, a wave of English has spread from North to South (*6*). In Sweden, Denmark, and Netherlands, nearly 80% of the population now claim fluency in the language (Fig. 2); France is a state of transition; in Italy, Spain, Greece, and Portugal, learning English is now big business. Indeed, students and employees may be assumed to speak English—it is regarded as a basic skill taught in elementary school alongside computer skills. Employers in parts of Asia are already looking beyond English—in the next decade, the new "must-learn" language is likely to be Mandarin.

The spread of English and other major languages beyond their traditional territories has eroded the idea that "one country, one language" is the norm. In the new world order, most people will speak more than one language and will switch between languages for routine tasks. Monolingual English speakers may find it difficult to fully participate in a multilingual society. We also must think differently about what it means to speak a language, or to learn and teach it. The expectation that someone should always aspire to native speaker competence when learning a foreign language is under challenge, as is the notion of "native speaker" itself.

Future of Grammar

Theoretical perspectives in linguistics will shift to reflect these trends. In the 19th century, scholarly study of language focused on vocabulary—particularly its historical subdiscipline known as "etymology"—and phonetics, which provided a new, laboratory-based dimension. During the 20th century, scholars became more interested in grammar—especially the problems related to word order and syntax that a language like English presented (the heavily inflected classical languages of Latin and Greek had allowed grammar to remain a branch of word study).

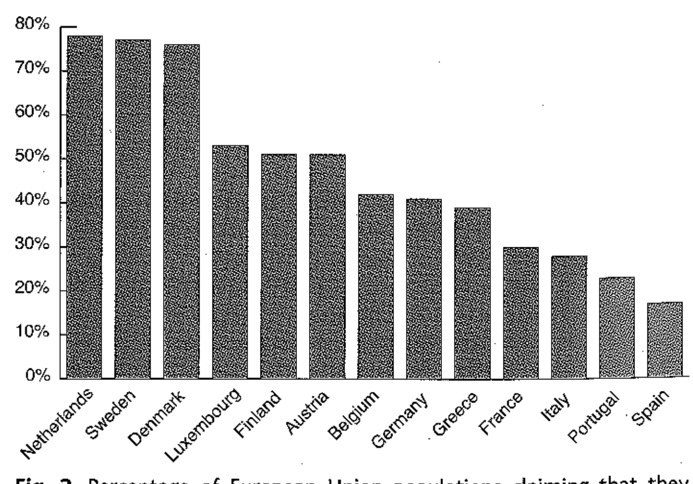


Fig. 2. Percentage of European Union populations claiming that they speak English (*6*).

Table 2. Estimates of numbers of native speakers globally aged 15 to 24 in 2050 (*1*).

Language	No. of native speakers globally (millions)
1. Chinese	166.0
2. Hindi/Urdu	73.7
3. Arabic	72.2
4. English	65.0
5. Spanish	62.8
6. Portuguese	32.5
7. Bengali	31.6
8. Russian	14.8
9. Japanese	11.3
10. Malay	10.5

When Noam Chomsky published his groundbreaking book on *Syntactic Structures* in 1957 (*7*), syntax became regarded as the central problem in linguistic description. But in the future, we may come to appreciate how far the Chomskyan approach has led linguists down a blind alley. Over the last half-century, theories of syntax have lost touch with language as spoken by people in the real world, and have retreated into abstract studies of universal features of human cognition.

Linguists keen to develop theories applicable to real-world problems of our age—such as in education, machine translation, information retrieval, national security, and even forensic law—have begun to exploit "data mining" techniques made possible by the power of modern computers. They have scrutinized patterns of language in huge collections of real-world texts and conversations—hundreds of millions of words at a time. Such corpus-based analysis already suggests an answer to something that has puzzled grammarians for hundreds of years: No one has ever successfully produced a comprehensive and accurate grammar of any language. In the words of the early 20th-century anthropological linguist, Edward Sapir, "all grammars leak" (*8*). Some emerging text-based grammars suggest that such an attempt is unnecessary—there need be no more endless arguments over taxonomies of subordinate clause. It seems that much of what we have expected of grammars can be better explained by focusing on words and the complex way in which they keep each other's company. Some words tend to be used as the subject rather than object of a clause, others may typically appear in prepositional phrases. The human brain is able to store experience of how words pattern, what kinds of text they appear in, what kinds of rhetorical structure will follow them.

This is the new science of collocation and colligation that illuminates how texts work.

Future of Texts

Corpus linguists will have to work fast to keep up with the changing nature of texts. As texts become shorter, more fragmentary, and multimodal (using pictures, color, sound, kinetics as well as words), so strategies of interpretation and ways of reading will change.

A struggle is brewing too between author and reader, the producer and consumer of texts, which has many of the dimensions—political, economic, social, technological—that characterize postmodernity. On the one hand, multimodal texts need more attention by designers and editors to marshal disparate forms of information into a coherent whole. But against them is a movement—at times fundamentalist in fervor—that demands free access to "content," and argues that publishers, editors, and designers are part of a capitalist conspiracy to add cost and control access to knowledge. Digital texts may mark the death of design—which will become a matter of a reader's preference setting. But technology also gives publishers new freedom to reversion intellectual property, to make it look different to different categories of reader, and to sell text by the paragraph. The linguistic resources required to construct and interpret longer, unified texts—which collectively form institutionalized genres—may be lost in all but specialized domains such as the scientific article. Readers will be left to make sense of fragmentary, often contradictory information dispersed across different channels.

Will the Future Understand Us?

When Thomas Sebeok, an American specialist in semiotics, was asked in the 1980s to advise on a method of communicating the whereabouts of dangerous repositories of radioactive waste to generations 10,000 years hence, he concluded that there was no secure means of transmitting such knowledge over 300 generations. Instead, he recommended putting in place a relay system which ensured that "as the information begins to decay, it should be updated" and argued that any messages written in English should be designed for only three generations ahead—that is, 100 years (*9*). This may seem a short horizon—if a linguist were faced now with a typical text from the 22nd century, he or she would be unlikely to conclude that the language has radically changed in its core vocabulary or grammar. But we might not be able to make much sense of it.

References

1. D. Graddol, *The Future of English?* (British Council, London, 1997).
2. D. Graddol, in *English in a Changing World*, D. Graddol, U. Meinhof, Eds. AILA, Milton Keynes, 1999), pp. 57–68.
3. United Nations Environment Programme, Press Release, February 2001. Also at www.unep.org/Documents/Default.asp?DocumentID=192&ArticleID=2765.
4. D. Crystal, *Cambridge Encyclopedia of the English Language* (Cambridge Univ. Press, Cambridge, ed. 2, 2003).
5. D. Graddol, D. Leith, in *English: History, Diversity and Change*, D. Graddol, D. Leith, J. Swann, Eds. (Routledge, London, 1994), pp. 136–166.
6. European Commission, *Standard Eurobarometer 52*, http://europa.eu.int/comm/public_opinion/archives/eb/eb52/eb52_en.htm (2000).
7. N. Chomsky, *Syntactic Structures* (Mouton, The Hague, 1957).
8. E. Sapir, *Language* (Harcourt, Brace & World, New York, 1921), p. 39.
9. T. A. Sebeok, in *On Signs*, M. Blonsky, Ed. (Blackwell, Oxford, 1985), pp. 448–466.

VIEWPOINT

Software and the Future of Programming Languages

Alfred V. Aho

Although software is the key enabler of the global information infrastructure, the amount and extent of software in use in the world today are not widely understood, nor are the programming languages and paradigms that have been used to create the software. The vast size of the embedded base of existing software and the increasing costs of software maintenance, poor security, and limited functionality are posing significant challenges for the software R&D community.

We are living in a rapidly evolving information age. Computers, networks, and information pervade modern society. Some of the components are visible: Virtually every office and home is equipped with information devices such as personal computers (PCs), printers, and network connection devices. An increasing fraction of the population is using the Internet for tasks as varied as e-mail, messaging, searching for information, entertainment, and electronic shopping. The amount of information on the Internet is measured in exabytes.

Most of the infrastructure supporting the information age, however, is not evident. Today's information appliances such as TVs, organizers, and phones contain microprocessors and other forms of embedded computer systems. Telecommunications and Internet access systems are all controlled by networked computers. Wireless networks with voice and data capabilities are found the world over.

Department of Computer Science, Columbia University, New York, NY 10027, USA. E-mail: aho@cs.columbia.edu

The information age has been thrust upon society, and everyone is being affected by the new technology. The information infrastructure is creating new opportunities for improving all aspects of life from childhood to old age. But the technology is also creating new challenges, especially in areas such as the security and privacy of information systems.

The Unappreciated Importance of Software

Few people appreciate the importance of software—until it breaks! The amount of software used by governments, companies, educational institutions, and people throughout the world is staggering. An individual system, such as a PC operating system, can consist of many tens of millions of lines of code. If we assume that there are 5 million programmers worldwide, each producing 5000 lines of new software a year (the industry average), then a conservative estimate is that the world is already using hundreds of billions of lines of software to conduct its affairs. Assuming that it costs somewhere between $10 and $100 to produce a line of working software, we see that the worldwide investment in software is in the trillions of dollars. A software system requiring tens of millions of lines of code would cost hundreds of millions of dollars to develop from scratch. The high cost of new software development is one of the principal drivers of the creation of open-source software, whose system development is essentially done for free by volunteer software specialists throughout the world. But open-source software has created another market opportunity: companies that provide maintenance and customization services for users of open-source systems.

A more sobering figure lies in the number of defects in this huge embedded software base. Software managers estimate that the number of defects per million lines of delivered software ranges between 10 and 10,000. Assuming an embedded base of 500 billion lines of software, this would mean there are somewhere between 5 million and 50 billion defective lines lurking in the embedded base waiting to be triggered. It is not known how many of these are showstopper defects that result in the failure of a system, but all too often major critical systems fail because of software defects. It has been estimated that the Y2K software problem cost hundreds of billions of dollars to fix. Because software defects can result in huge economic losses, or in some cases the catastrophic failure of a system, a lot of software research is being devoted to more effective methods than testing for making reliable software.

How Programming Languages Have Evolved

The first programming languages were machine languages, consisting of commands expressed as sequences of 0's and 1's that told the computer what operations to execute. It was difficult and time-consuming to write programs in machine language, and once written, the programs were hard to decipher and modify. The first significant step toward a more human-friendly programming language was the invention of assembly languages in the 1950s. Assembly languages are still close to the machine, but when sequences of 0's and 1's were replaced by mnemonic commands, such as "load register X from memory location Y" or "store the contents of register X into memory location Y," programs became easier for humans to write and modify.

In the latter half of the 1950s, another major step was made in the design of more human-friendly languages. The higher-level languages Fortran, Cobol, and Lisp were invented: Fortran for scientific computation, Cobol for business data processing, and Lisp for symbol processing. After some initial resistance by hard-core assembly language programmers who felt that programs written in these languages would not run as fast as their assembly language programs, these languages became immensely popular in their respective communities. Fortran became the lingua franca of scientific computing, Cobol was for many years the world's most popular programming language, and Lisp became the mainstay of the artificial intelligence community. Their popularity stemmed from the fact that programs for the intended application areas could be written more rapidly, concisely, and easily in these languages than in assembler language. Indeed, these languages and their descendants continue to be widely used even today.

The 1960s saw the development of Algol 60, the first programming language with block structure. Although Algol 60 was never widely used, it is notable for the languages it influenced, including Pascal and Modula. Algol 60 also introduced BNF (for Backus-Naur Form), now a widely used grammatical notation for expressing the syntax of a language. For many years, Pascal was used as the introductory programming language in computer science departments.

John Kemeny, while at Dartmouth College, felt that every college student should know how to program; and in the early 1960s, he and Thomaz Kurtz created a simple imperative language called Basic, which was easy to learn. Basic spawned many dialects, the most notable of which is Visual Basic, arguably the world's dominant programming language today.

Another major development during the 1960s was the introduction of object orientation into programming through the creation of Simula 67 (*1*). Like Algol 60, Simula 67 itself was not widely used, but its influence was profound. Nearly every modern general-purpose programming language today supports object-oriented programming. Object orientation allows a programmer to focus on the design of the data (objects) and the interfaces to the data. It facilitates "plug and play" among software modules.

In the early 1970s, Dennis Ritchie at Bell Labs created the systems programming language C to implement the third version of the Unix operating system that he was codeveloping with Ken Thompson. In the 1980s, Bjarne Stroustrup, also at Bell Labs, created C++, an extension of C with object orientation. Because of their efficiency and early association with Unix, C and C++ became the most widely used systems programming languages.

Another genre of popular programming languages is the scripting languages: typeless languages with high-level primitives for manipulating data (*2*). Some of the most popular languages of today, such as awk, javascript, perl, php, python, sh, and tcl, are scripting languages. With these languages it is possible to specify programming tasks in a few lines of code that would otherwise take hundreds of lines in a lower-level language such as C or C++.

Trends in Modern Language Design

The newest major programming languages, Java and C#, build on the legacy of C and C++ but incorporate features that support the notion that the information infrastructure must be a robust distributed system. Many of the features in these languages are based on strong theoretical underpinnings, such as strong type systems and monitors, and on software design techniques that have been extensively explored by the computer science research community. Here are some of the design principles that the creators of these new languages have espoused (*3*).

Simplicity. Programmers want languages that are easy to learn, use, and understand. The newer languages tend to have support for features that make programming easier, such as automatic garbage collection; yet have a syntax that is familiar to C and C++ programmers.

Robustness. Because security and safety are of paramount importance in modern software systems, the new languages have strong type systems that allow more errors to be caught at compile time, and they restrict the use of pointers, which account for many of the vulnerabilities that hackers tend to exploit in C and C++ programs.

Portability. Programmers would like their programs to run on as wide a variety of machine architectures as possible, producing the same results. The newer languages have introduced mechanisms such as run-time bytecode interpreters, which allow the same program to be run, producing the same results, on different machine architectures.

Internet compatibility. Everyone would like to access software applications from anywhere on the Internet. The new languages either have class libraries or access to class libraries that facilitate the connection of programs with Internet protocols and applications.

Concurrency. Modern applications need to interact with many systems at the same time. The newer languages have multithreading and concurrency primitives that support the building of applications requiring simultaneous multiple threads of execution.

Languages of the Future

Software can be used in virtually any field of human endeavor. Just as there are different notations for dealing with chemistry, dance, law, mathematics, music, and so on, there will never be a single language for creating all forms of software. There are already thousands, perhaps tens of thousands, of programming languages in use today. Each field has at least one language that is used primarily by the practitioners of that field. Most college students today are familiar with languages for editing documents, formatting papers, creating presentations, and performing calculations. Each language is universal in the sense that it is capable of expressing any computation, but each language has been designed so it is easy to use for the applications it was designed for, be it hardware design or music synthesis. Naturalness and ease of use will most likely continue to be dominant influences in language design.

Some researchers are currently exploring more exotic forms of languages involving speech, gestures, pictures, and templates. It is not clear at this point whether multimedia will become a dominant feature in programming languages of the future, but it is quite clear that in certain application areas, these new forms of communication with computers are compelling.

Making Software Systems More Reliable

A fundamental question for the field of software development is whether there is a scientific basis for making reliable software. In 1956, von Neumann showed how more reliable hardware could be made from unreliable components by using redundancy (*4*), and earlier Shannon showed that unreliable communication over a noisy channel could be made more reliable by using error-detecting and -correcting codes (*5*). Today, redundancy and error-detecting and -correcting codes are routinely used.

No analogous technique is known for making reliable software. Although the production of software involves people as well as process and technology, the human component in the production of software has not been adequately understood or modeled. Some software researchers have advocated *N*-version programming, a technique where two or more people independently write a program from the same specification, but subsequently researchers have discovered that programmers tend to make the same kinds of mistakes even if they don't communicate with one another (*6*).

It is unlikely that humans will ever write software with zero defects. Researchers are actively exploring many techniques to make more reliable software systems, keeping the frailties of human programmers in mind. Static type checking and model checking provide promising avenues for detecting errors earlier in the software life cycle. A more ambitious approach is to see whether software systems can be designed to be resilient to errors. An even more ambitious approach is to design systems that automatically correct errors when they are detected. The goals of this research are laudable, but most likely it will be some time before we will see self-correcting software systems widely deployed in practice. What we can be certain about is that the embedded base of software will continue to grow in size, diversity, and functionality.

References and Notes

1. K. Nygaard, O. Dahl, in *Proceedings of the ACM SIGPLAN History of Programming Languages Conference*, R. L. Wexelblat, Ed. (ACM Monograph Series, New York, 1981), pp. 439–493.
2. J. Osterhout, *Higher Level Programming Languages for the 21st Century* (IEEE Computer, IEEE, New York, 1998), pp. 23–30.
3. For an overview of these principles, see http://java.sun.com/docs/overviews/java/java-overview-1.html.
4. J. von Neumann, in *Automata Studies*, C. E. Shannon, J. McCarthy, Eds. (Princeton Univ. Press, Princeton, NJ, 1956), pp. 43–98.
5. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
6. S. Brilliant, J. Knight, N. Leveson, *IEEE Trans. Software Eng.* **16**, 238 (1990).

VIEWPOINT

Of Towers, Walls, and Fields: Perspectives on Language in Science

Scott Montgomery

Language in science is in the midst of change and appears dominated by two contradictory trends. Globalization of scientific English seems to promise greater international unity, while growth of field-specific jargon suggests communicational diaspora. Real in part, each trend is complex and multileveled, and includes elements of convergence and divergence, along with important implications for the present and future of technical knowledge.

Science, it appears, has come to a historical crossroads. On the one hand, it would seem to have completed the Tower of Babel, its knowledge now reaching far beyond the heavens and, through the global spread of English, recovering the ancient dream of a single language for the wisdom of the nations. Yet, from another vantage, the very opposite is suggested: this great tower of unanimity broken and rebuilt into a thousand walls by the power of jargon, dividing the disciplines by the arcanity of specialist speech.

Two great trends of opposing force, two linguistic movements that annul each other's action. Is such a state of affairs real, and is it prevalent? What do the facts say, as far as we can discern them, and what are their implications?

These are not mere academic questions. Scientific knowledge exists because scientists are writers and speakers, users and sharers of language that, like all language, is constantly evolving. Words are the primary medium by which technical work is embodied, added to the corpus of professional understanding, and passed on. Whatever directly affects the speech of science and its development affects scientific endeavor near its core—its ability to express and render available its nuclear substance.

Take the case of English, then. How true is the claim that it constitutes an international language for science, an ever-expanding one? The answer is "very true," indeed, but with certain limits and qualifications.

The Role of English

Dominant use of English in science must be understood within a larger frame. First, there is the advent of this tongue as a global language generally. British colonialism sowed the seeds early on, in North America, India, Australia, Hong Kong, and other centers of influence. Simultaneously, the Industrial Revolution gave English prominence in technological matters crucial to modernization. However, it has really been since World War II, which so greatly advanced U.S. military, economic, technological, and political sway (and thereby, cultural impact), that English has become linguistic capital for the larger world. Today, this tongue serves as lingua franca for a wide range of domains in daily experience: entertainment, advertising, travel and tourism, international business, telecommunications, the news media, computer technology. English is now the most popular, and most required, foreign language to be studied anywhere (*1*). Its uptake in technical circles, meanwhile, has been aided by the rise of "big science" in the United States and the resulting vast increase in scientific output. English, in a sense, has ridden a great wave of cultural and intellectual affluence.

1511 18th Avenue East, Seattle, WA 98112, USA. E-mail: scott.montgomery@prodigy.net

Second, there has been the globalization of science itself. Industrial development in Asia, portions of Africa, the Middle East, and Latin America has motivated the spread of research in many fields. Today, important conferences and symposia are held regularly on every continent, thereby providing demand for a common medium of speech. Part of this, too, has been the Internet, developed in the United States and dominated early on by the English language. Although the Net has become more linguistically diverse with each passing year, the precincts of scholarly writing—science above all—continue to favor English to a high degree (*2*).

Linguistic studies suggest that by the 1980s, more than 60% of the journal literature in science was being printed in English (*3*). Twenty years later, the figure is likely closer to 80% (for some fields, over 90%). English has become the language chosen for international meetings of all types, for corporate science, multinational research programs, official Web sites, and much more. On the informal side, "invisible colleges" made possible by the Internet also rely on this tongue—if a nuclear chemist from China wants to contact or collaborate with a colleague in Brazil or Germany (or both), they will use English to communicate. Less apparent is the importance all this has given to English language training, one of the major growth industries today and part of the regular technical curriculum in most universities, with upper-level courses in some countries now actually being taught in English.

Implications

There are implications to all this, of course, and they are potent. The role of English greatly increases the possibilities for international commerce in scientific work, on many levels. Agricultural biologists in Ecuador can make direct use of new fertilizer studies in Sri Lanka or Taiwan. Having an international language lubricates and encourages exchange. It also adds to the career options (and mobility) of scientists, and it can help provide a way for technical communities to advance themselves and to join their specific talents and subject matter to the contemporary scientific enterprise.

However, executive use of English can have other effects as well. Languages of intellectual power—such as Greek in the Hellenistic period, Arabic in the 8th to 12th centuries, and Latin in medieval Europe—can divide the world into "haves" and "have nots." This means the possibility for severe intellectual isolation. Scientists with little or only moderate command of English may find it more difficult to publish in international journals and thus to reach a corresponding audience. English dominance can weaken the relevance of other tongues, if these are unable to recreate the expanding technical vocabularies now preferentially coined in the global tongue. Such dominance can yield the illusion of a communicational tyranny by English and the underdog psychology that goes with it.

Realities of this type have been much decried, with no small justice (*4–6*), but two points should be made. First, any idea of linguistic imperialism, applied to science, misses the mark. Yes, these effects are serious and potentially damaging, for individuals and for science too. Yet no central hand is at work here and no reach for empire. There are no standardizing efforts on the global scene led with efficacy by one or another self-appointed authority (*7*). This leads to the second point: No one "owns" scientific English. No single country or culture possesses control over its qualities or direction (it is a true international language). Instead, the process of globalizing English seems a rather wild and woolly affair, a cumulative effect of myriad decisions by editors, teachers, students, parents, writers, publishers, translators, officials, scholarly associations, corporations, schools, and so on, with an equally wide array of motives. Language change, in general, has never been a phenomenon that can be strictly legislated. Will the spread of English continue? Yes, without doubt; there is much momentum in this direction. However, this is no cause for fatalism.

Limits and Variations

English may well govern many areas of technical communication, but it does not rule all of science, by any means. Its realm has boundaries, which prevent it from sweeping aside other tongues in the great conversation of science. Much technical literature and (lest we forget) spoken exchanges of all types remain entirely dependent on national languages. The prevalence of English, globally speaking, is mainly confined to formal and international settings—a major part of science, to be sure, but not the only part. Another question also arises: Do scientists and engineers everywhere speak the same English? Do Anglo-American forms of the language, for example, rule the planet?

The answer, it turns out, is no. Linguists today speak less "world English" than "world Englishes" or "varieties of English" (*8–10*). A broad literature reveals how, as an imported tongue, English is modified when it enters new linguistic communities and becomes nativized. Adopted languages, after all, are never received as final coin. They are cultural material, and as such, undergo adaptive re-casting. The English of Hong Kong is not precisely that of West Africa or the United States or Singapore (*11*). The same holds true for scientific English, though in restrained fashion. A term like "lithoassociation" or a phrase such as "wholesome models of crustal structure," found in Indian geologic writing, lie outside the American or British literature. Selective elimination of articles (such as "a," "an," and "the"), use of different plural forms ("sedimentaries" rather than "sedimentary strata"), and other phenomena of this type can be routinely seen in other scientific Englishes—they are not mere errors, in other words, but part of a system of alternative usage (*12*). Do they threaten, however, to erode shared comprehension? This can be the case for colloquial varieties of English (*13*). In science, however, with its emphasis on conservative usage and international intelligibility, experience indicates they do not. What of the future? Scientific Englishes would lose all purpose were they to diverge too far; there would be no reason to employ them.

In the meantime, a few figures are worth contemplating: ~400 million people have English as their mother tongue, >430 million have it as a second language, and ~750 million more use it as a foreign language (*1*). Overall, the proportion of native speakers is decreasing worldwide, from over 8% in 1950 to ~6.5% today, projected to 5% or less by 2050 (*14*). The pressures for continued adaptation, posed by fewer and fewer native speakers of a more and more global language, would appear impressive.

Jargon: Quo Vadis?

What, then, of our second theme, the growth and consequence of jargon? Nearly every scientist has familiarity in this area too. Consider the entomologist faced with an issue of *Physics Today* or *Cell*, the graduate student in oceanography doing battle with an article in *Atmospheric Research*. These are not impossible or even improbable encounters, but they would likely be difficult ones. Even within single fields, boundaries of terminology may seem to hopelessly divide specialties and subspecialties, whose numbers grow greatly with each passing decade. How far has this process gone? Calling oneself a "gravitational wave physicist" or an "expert on leg anatomy of Early Cretaceous sauropods" is not at all extraordinary. In the meantime, "biology" and "geology" have evolved into the "life sciences" and "earth and planetary sciences." Ours is the era of such pluralizations. They, too, are part of the language of science.

The birth of new fields, and thus new vocabularies, has been a defining aspect of scientific progress. Specialization reveals itself as a mark of intellectual vigor, the historical sign that in order to expand and deepen, natural science has had to diversify and concentrate. It has had to pursue new subject matter, engage in greater precision, work at smaller or larger levels of observational and analytical scale, take up higher mathematics, and develop improved laboratory technologies, all the while inventing new terms and phrases to express the new knowledge and new practices gained. On the surface, these new vocabularies have seemed to turn science into a glitter of disconnected realms, self-contained linguistic galaxies spinning outward, ever apart.

Yet this perception, however common, misses something critical about the nature of each field's dilemma—and, perhaps, the dilemma of its nature. Increasing specialization, rather than causing only a spiralling dispersal has resulted in new connections of its own, new cross-over. Growing specialization has generated an ever-greater range of opportunities—even demands—for the sharing of language. For instance, the power to examine, analyze, and manipulate phenomena at smaller and smaller scales has brought the province of the molecular, once reserved for chemists, into immediate relevance for botany, zoology, medicine, meteorology, many areas of geology, engineering, and so on. This has meant the adoption of terminologies appropriate to such scales of observation and analysis.

Commingling has a number of sources. Integration of computer technology into nearly every aspect of science is one. The adaptive use of other technologies (e.g., nuclear magnetic resonance, laser optics, and neural network applications) is another. Exploring phenomena from a multidisciplinary vantage—the human genome, for instance, or the surface of Mars—continues to be a major part of science. "Transdisciplinary research," as often said, has brought options and opportunities to every field. Formerly separate areas have been united: biopaleogeography, psychoneuroimmunology, planetary geophysics, and chemical anthropology, among many others. At every step is the increased sharing of terminologies. Ours is indeed an era of pluralisms, but a fruitful era as well.

The language of science, in consequence, reveals patterns of divergence and convergence both. This language, as it evolves, is headed neither toward ultimate unity nor utter diaspora. Barriers set up by specialized jargon exist, without doubt, as they have for some time. Yet many have become increasingly porous, allowing flow in both directions. Such will undoubtedly continue—science is today the most active area of language creation.

The truer dilemma would seem to involve barriers between scientific discourse in general and public understanding. Here, porosities are lacking; on the contrary, the walls built by arcane terminologies have become thicker and more solid. Scholars of English state that scientific terms make up the greater part of this most global of languages (*15*), suggesting that technical Englishes everywhere face a similar fate of inaccessibility. The problem of scientific literacy has a decided international dimension. Making science comprehensible to general audiences has literally become an effort of translation, and the translator who accepts this task in English must be counted a global voice of no small influence.

Possibilities and Goals

Even in so brief and selective a discussion as this, some clear tasks emerge for science and its language. To help ensure expanding access to English for all of science's intellectual citizens would appear a worthy goal. Language training is now a critical fact and compelling factor in modern science; how this is done and by whom become questions of considerable import, with sociopolitical dimensions. Another, no less important, goal would be to increase tolerance toward variation in scientific English—to avoid the imperial attitude that one standard must be obeyed in order that any and all threats of semantic chaos be met. Success in communicating science to general audiences will rise and fall with the quality of its mediators—teachers, science writers, scientists themselves, and any others. It may help, in this arena of work, to recognize the linguistic context involved: Explaining science qualifies as a form of translation, the movement of knowledge from one linguistic context to another.

Finally, a caveat. Language in science is a historical reality, evolving at every turn. To recognize English as an international or global tongue is to state a fact of linguistic geography; to call it "the universal language of science," evoking Babel once more, is to indulge a very different, indeed mistaken, sense of innateness and the inevitable. Similarly, to lament the use of jargon as an imprisoning effect of doing science is to overlook the changing complexities of its role vis à vis knowledge and progress. Using statements of finality to comprehend these areas of science is like attempting to unlock one thing with the key to something else. Towers and walls can certainly be seen across the fields of technical speech today. Yet the full landscape, when revealed to the interested eye, shows a domain that directly reflects, as it must, the complexity and the fertile change that are central to scientific practice itself.

References and Notes

1. D. Crystal, *English as a Global Language* (Cambridge Univ. Press, Cambridge, ed. 2, 2003).
2. Reliable numbers are hard to come by. Crystal (*1*) proposes an estimate of 80% English sites for the World Wide Web during the 1990s, dropping steadily, possibly to 40% or less by 2005, as all portions of the globe come online. However, science is a different matter. Electronic journals in science, which have increased greatly since 1998, are predominantly in English. Preprint archives, an especially influential and unique offering of Internet science, are also in this language. Institutional sites, such as those for research organizations, universities, international programs, and so on, are also usually in English, and where not—a key fact—they often provide a link to an English version of the site.
3. A. Large, *The Foreign-Language Barrier* (Deutsch, London, 1983).
4. R. Phillipson, *Linguistic Imperialism* (Oxford Univ. Press, Oxford, 1992).
5. A. S. Canagarajah, *Resisting Linguistic Imperialism in English Teaching* (Oxford Univ. Press, Oxford, 1999).
6. A. Pennycook, *The Cultural Politics of English as an International Language* (Longman, London, 1994).
7. M. Görlach, *Still More Englishes* (John Benjamins, Amsterdam, 2002).
8. T. McArthur, *The English Languages* (Cambridge Univ. Press, Cambridge, 1998).
9. International Association for World Englishes, www.iaweworks.org/
10. B. Kachru, Ed., *The Other Tongue: English Across Cultures* (Univ. of Illinois Press, Urbana, IL, ed. 2, 1992).
11. B. Kachru, C. Nelson, in *Analysing English in a Global Context: A Reader*, A. Burns, C. Coffin, Eds. (Routledge, London, 2001), pp. 9–25.
12. S. L. Montgomery, *Science in Translation: Movements of Knowledge through Cultures and Time* (Univ. of Chicago Press, Chicago, IL, 2000), pp. 258–263.
13. L. E. Smith, in (9), pp. 75–90.
14. D. Graddol, in *English in a Changing World* (AILA Review 13), D. Graddol, U. H. Meinholf, Eds. (Association Internationale de Linguistique Appliqué, London, 1999), pp. 57–68.
15. D. Crystal, *The Cambridge Encyclopedia of the English Language* (Cambridge Univ. Press, Cambridge, 1995), p. 119.