

March 16, 2004

## A Biological Dig for the Roots of Language

By NICHOLAS WADE

Once upon a time, there were very few human languages and perhaps only one, and if so, all of the 6,000 or so languages spoken round the world today must be descended from it.

If that family tree of human language could be reconstructed and its branching points dated, a wonderful new window would be opened onto the human past.

Yet in the view of many historical linguists, the chances of drawing up such a tree are virtually nil and those who suppose otherwise are chasing a tiresome delusion.

Languages change so fast, the linguists point out, that their genealogies can be traced back only a few thousand years at best before the signal dissolves completely into noise: witness how hard Chaucer is to read just 600 years later.

But the linguists' problem has recently attracted a new group of researchers who are more hopeful of success. They are biologists who have developed sophisticated mathematical tools for drawing up family trees of genes and species. Because the same problems crop up in both gene trees and language trees, the biologists are confident that their tools will work with languages, too.

The biologists' latest foray onto the linguists' turf is a reconstruction of the Indo-European family of languages by Dr. Russell D. Gray, an evolutionary biologist at the University of Auckland in New Zealand.

The family includes extinct languages like Hittite of ancient Turkey, and Tokharian, once spoken in Central Asia, as well as the Indian languages and Iranian in one major branch and all European languages except Basque in another.

Dr. Gray's results, published in November in *Nature* with his colleague Quentin Atkinson, have major implications, if correct, for archaeology as well as for linguistics. The shape of his tree is unsurprising — it arranges the Indo-European languages in much the same way as linguists do, using conventional methods of comparison. But the dates he puts on the tree are radically older.

Dr. Gray's calculations show that the ancestral tongue known as proto-Indo-European existed some 8,700 years ago (give or take 1,200 years), making it considerably older than linguists have assumed is likely.

The age of proto-Indo-European bears on a longstanding archaeological dispute. Some researchers, following the lead of Dr. Marija Gimbutas, who died in 1994, believe that the Indo-European languages were spread by warriors moving from their homeland in the Russian steppes, north of the Black and Caspian Seas, some time after 6,000 years ago.

A rival theory, proposed by Dr. Colin Renfrew of the University of Cambridge, holds that the Indo-Europeans were the first farmers who lived in ancient Turkey and that their language expanded not by conquest but with the spread of agriculture some 10,000 to 8,000 years ago.

Dr. Gray's date, if accepted, would support the Renfrew position.

Several linguists said Dr. Gray's tree was the right shape, but added that it told them nothing fresh, and that his dates were way off. "This method is not giving anything new," said Dr. Jay Jasanoff, a Harvard expert on Indo-European. As for the dates, Dr. Jasanoff said, "The numbers they have got seem extremely wrong to me."

Dr. Don Ringe, a linguist at the University of Pennsylvania who has taken a particular interest in computer modeling of language, said that Dr. Gray's approach was worth pursuing but that glottochronology, the traditional method of dating languages, had "failed to live up to its promise so often that convincing linguists there is anything there is an uphill battle."

In the biologists' camp, however, there is a feeling that the linguists do not yet fully understand how well the new techniques sidestep the pitfalls of the older method. The lack of novelty in Dr. Gray's tree of Indo-European languages is its best feature, biologists say, because it validates the method he used to construct it.

Most historical linguists know a few languages very well but less often consider the pattern of change affecting many languages, said Dr. Mark Pagel, an evolutionary biologist at the University of Reading.

"The field is being driven by people who are not confronted with the broad sweep of linguistic evolution and is being invaded by people like me who are only interested in the broad sweep," Dr. Pagel said.

Glottochronology was invented by the linguist Morris Swadesh in 1952. It is based on the compiling of a core list of 100 or 200 words that Swadesh believed were particularly resistant to change. Languages could then be compared on the basis of how many cognate words on a Swadesh list they shared in common.

Cognates are verbal cousins, like the Greek *podos* and the English *foot*, both descended from a common ancestor. The more cognates two languages share, the more recently they split apart. Swadesh and others then tried to quantify the method, deriving the date that two languages split from their percentage of shared cognates.

The method gave striking results, considering its simplicity, but not all of the findings were right. Glottochronology suffered from several problems. It assumed that languages changed at a constant rate, and it was vulnerable to unrecognized borrowings of words by one language from another, making them seem closer than they really were.

Because of these and other problems, many linguists have given up on glottochronology, showing more interest in an ingenious dating method known as linguistic paleontology.

The idea is to infer words for items in the material culture of an early language, and to correlate them with the appearance of such items in the archaeological record. Cognates for the word wheel exist in many branches of the Indo-European family tree, and linguists are confident that they can reconstruct the ancestral word in proto-Indo-European. It is, they say, "k'ek'los," the presumed forebear of words like "chakras," meaning wheel or circle in Sanskrit, "kuklos," meaning wheel or circle in Greek, as well as the English word "wheel."

The earliest wheels appear in the archaeological record around 5,500 years ago. So the proto-Indo-European language could not have started to split into its daughter tongues much before that date, some linguists argue. If the wheel was invented after the split, each language would have a different or borrowed word for it.

The dates on the earliest branches of Dr. Gray's tree are some 2,000 years earlier than the dates arrived at by linguistic paleontology.

"Since 'wheel' is shared by Tocharian, Greek, Sanskrit and Germanic," said Bill Darden, an expert on Indo-European linguistic history at the University of Chicago, "and there is no evidence for wheels before the fourth millennium B.C., then having Tokharian split off 7,900 years ago and Balto-Slavic at 6,500 years ago are way out of line."

Dr. Gray, however, defends his dates, and points out a flaw in the wheel argument. What the daughter languages of proto-Indo-European inherited, he says, was not necessarily the word for wheel but the word "k'el," meaning "to rotate," from which each language may independently have derived its word for wheel. If so, the speakers of proto-Indo-European could have lived long before the invention of the wheel.

His tree, Dr. Gray said, was derived with the methods used by biologists to avoid problems identical to those in glottochronology. Genes, like languages, do not mutate at a constant rate. And organisms, particularly bacteria, often borrow genes rather than inheriting them from a common ancestor. Biologists have also learned that trees of any great complexity cannot be drawn up by subjective methods. Mathematical methods are required, like having a computer generate all possible trees — a number that quickly runs way beyond the trillions — and then deciding statistically which class of trees is more probable than the rest.

Dr. Gray based his tree on the Dyen list, a set of Indo-European words judged by linguists to be cognates, and he anchored the tree to 14 known historical dates for splits between Indo-European languages.

Many of the Dyen list cognates are marked uncertain, so Dr. Gray was able to test whether omission of the doubtful cognates made any difference (it did not). He also tested many other possible assumptions, but none of them produced an age for proto-Indo-European anywhere near the date of 6,000 years ago favored by linguists.

"This is why our results should be taken seriously by both linguists and anyone else interested in the origin of the Indo-European languages," he wrote, in a recent reply to his critics.

"We haven't repeated the errors of glottochronology," Dr. Gray said in an interview. "What we are doing is adding value, since we can make inferences about time depths which can't be made reliably in other ways."

Dr. Gray said he had formed collaborations with linguists and hoped they would give his tree a warmer reception once his critics understood that he had not made the errors they cited.

Some linguists are interested in the biologists' approach.

"I think these methods are extremely promising," said Dr. April McMahon of the University of Sheffield and the president of the Linguistics Association of Great Britain, though she expressed concern about Dr. Gray's emphasis on dating language splits.

If the biologists' methods can date languages that existed 9,000 years ago, how much further back can they probe?

"Words exist that can in principle resolve 20,000-year-old linguistic relationships," Dr. Pagel of Reading wrote in a recent symposium volume, "Time Depth in Historical Linguistics," adding that "words that can resolve even deeper linguistic relationships are not out of the question."

Many linguists believe that once two languages have drifted so far apart that they share only 5 percent or so of their vocabulary, chance resemblances will overwhelm the true ones, setting a firm limit on how far back their ancestry can be traced.

"That's a mistaken reasoning which shows the linguists are relying on a model of evolution they trash when they see it written down," Dr. Pagel said.

He added that their argument assumed a constant rate of language change, the very point they know is wrong in glottochronology.

Geneticists believe modern humans may have left Africa as recently as 50,000 years ago, perhaps in a single migration with very small numbers. Reconstructing language of 20,000 years ago would be a big stride toward whatever tongue those first emigrants spoke. But Dr. Gray has no plans in that direction.

"It's hard enough to work out what happened 10,000 years ago, let alone 30,000 years ago," he said.