Linguistic Origins of Native Americans

Scholars have long wondered how Native Americans settled the New World. Recent research indicates that their many tongues belong to just three families, implying three waves of immigration from Asia

by Joseph H. Greenberg and Merritt Ruhlen

A little over two centuries ago Sir William Jones, an English jurist serving in India, was struck by remarkable similarities among Sanskrit, Classical Greek and Latin. He proposed that these languages, and probably Gothic and Celtic as well, had "sprung from some common source, which, perhaps, no longer exists." This source became known, in the following century, as Proto-Indo-European—a protolanguage that linguists have since labored to reconstruct [see "The Origins of Indo-European Languages," by Colin Renfrew; SCIENTIFIC AMERICAN, October 1989, and "The Early History of Indo-European Languages," by Thomas V. Gamkrelidze and V. V. Ivanov; SCIENTIFIC AMERICAN, March 1990].

Jones, however, did not reconstruct a syllable. He reached his conclusions by observing, as he put it, "a stronger affinity, both in the roots of verbs, and in the forms of grammar, than could possibly have been produced by accident" [see bottom illustration on page 96]. This evolutionary hypothesis was not lost on scholars interested in New World languages. In 1789, only three years after Jones's celebrated discourse, Thomas Jefferson wrote: "I endeavor to collect all the vocabularies I can, of American Indians, as of those of Asia, persuaded, that if they ever had a common parentage, it will appear in their languages."

Yet although 19th-century scholars identified hundreds of American languages and grouped them into families, none of them ventured the more comprehensive taxonomy that Jefferson had envisaged. The traditional account instead multiplied families, until the number reached about 60 in North America and about 100 in South America, far greater than the number in the Old World, where, for example, Africa has but four.

These estimates are puzzling because taxonomic diversity normally increases with time. Yet most archaeologists have long agreed that human settlement in the Old World substantially predates that in the New. The current consensus is that modern humans emerged at least 100,000 years ago, probably in Africa, and did not reach the Americas until about 12,000 to 20,000 years ago. How could the American languages have diversified to such a great extent?

The difficulty called for a more comprehensive classification. But in the early years of this century, when Alfred L. Kroeber and Edward Sapir first attempted to reduce the many American languages to a handful of larger families, they met with vigorous opposition from such anthropologists as Franz Boas, Pliny Goddard and Truman Michelson. These opponents did not seriously doubt that there were similarities among the American language groups. What they disputed—and what many dispute even today—was the origin of these similarities. Whereas Kroeber and Sapir insisted that the similarities stemmed from a common heritage, and were thus genetic in nature, Boas and his followers attributed the similarities to the diffusion of words from one language family to another.

Our research confirms the genetic approach. By comparing the most conservative elements in the vocabularies of hundreds of languages of North and South America, one of us (Greenberg) found just three families. Because each family shows closer affiliation with Asian families than with the other two American groupings, the tripartite division implies there were at least three distinct migrations from Asia. This hypothesis finds confirmation in the research of physical anthropologists.

The traditionalists who oppose our classification do not offer a better one. Instead they assert that by comparing languages two at a time and in great depth they will arrive at the true system—in another 50 to 100 years. We believe such work is misguided. To systematize a jumble of languages—or rocks, or animals—one must compare them as a group. Moreover, the multilateral approach has worked before. When Greenberg used it to classify the African languages some 40 years ago, traditionalists in that field opposed the method. Today everyone—even the traditionalists—embraces its results.

Those who compare languages two by two are simply ignoring much relevant evidence. Scholars related Albanian to English not by making a systematic comparison of the two languages in isolation but by establishing that each belonged to the Indo-European family. Indeed, Indo-Europeanists have never used a binary approach.

Our system of multilateral analysis uncovers precisely those relations that tend to escape notice in the binary ap-

JOSEPH H. GREENBERG and MERRITT RUHLEN collaborate in the comparative study of languages. Greenberg is professor emeritus of anthropology and linguistics at Stanford University. His chief interests are the universal features of language and the historical classification of language. He is past president of the Linguistic Society of America and has been elected a member of the American Philosophical Society, the American Academy of Arts and Sciences and the National Academy of Sciences. Ruhlen is an independent researcher based in Palo Alto, Calif. He earned a doctorate in linguistics from Stanford in 1973 and worked on the Language Universals Project. He is author of *A Guide to the World's Languages*.

GREENBERG'S CLASSIFICATION places America's many languages into just three families. Eskimo-Aleut (*purple*) and Na-Dene (*orange*) belong to the Old World groups known as Eurasiatic and Dene-Caucasian, respectively (*inset*). Amerind (*yellow*) is related to Eurasiatic. Amerind was the first family to enter the New World, Eskimo-Aleut, the last.

proach. We compare hundreds of languages at a time—a search in breadth rather than an analysis in depth—by examining a list of several hundred words. This list contains words that denote universal concepts, such as personal pronouns, body parts and aspects of nature (water and fire, for example). Because such concepts are rarely borrowed, languages seldom have occasion to borrow their names. English provides an illustration of this rule. Although it has borrowed many words from many languages, most of its basic vocabulary derives from Proto-Germanic. English has "one, two, three, I, mine, father, water"; German has "*ein, zwei, drei, ich, mein, Vater, Wasser.*"

A comparison of the basic vocabularies of hundreds of languages from North and South America led Greenberg to group the many postulated families into just three: Eskimo-Aleut, Na-Dene and Amerind. The first two—Eskimo-Aleut in the Arctic and Na-Dene in Canada and the southwestern U.S.—had long been accepted, and so the innovation consisted in grouping all the other American languages under Amerind. It contains 11 subfamilies, distributed throughout much of North America and all of South America [*see illustration on this page*].

In support of Amerind, Greenberg

Defining a Family by a Single Linguistic Innovation: T'ANA

Evidence in its daughter languages implies that Proto-Amerind had a root that sounded like T'ANA, meant "child" and assumed three vocalizations that indicated gender. Because the etymology runs through all of Amerind's 11 branches but is not found in any other group, it ties the family together and distinguishes it from others. Branches appear in the first column. Almosan-Keresiouan and Chibchan-Paezan are divided, and each thus occupies two rows. All daughter languages are modern save Proto-Uto-Aztecan, which is reconstructed.

LANGUAGE FAMILY AMERIND	LANGUAGE PROTO-AMERIND	FORM T'ANA	MEANING "CHILD, SIBLING"
Almosan	Nootka	t'an'a	"child"
Keresiouan	Yuchi	tane	"brother"
Penutian	Totonac	t'ána-t	"grandchild"
Hokan	Coahuilteco	t'an-pam	"child"
Central Amerind	Proto-Uto-Aztecan	*tana	"daughter, son"
Chibchan	Miskito	tuk-tan	"child, boy"
Paezan	Warrau	dani-	"mother's sister"
Andean	Aymara	tayna	"firstborn child"
Macro-Tucanoan	Masaca	tani-mai	"younger sister"
Equatorial	Urubu-Kaapor	ta'in	"child"
Macro-Carib	Pavishana	tane	"my son"
Macro-Panoan	Lengua	tawin	"grandchild"
Macro-Ge	Tibagi	tog-tan	"girl"

proposed about 300 etymologies, or groups of words that he believes have all evolved from a single ancestral word. The members of each such group are called cognates. Recent work by one of us (Ruhlen) has raised the number of etymologies to about 500.

Some of these roots are distributed so broadly that it is difficult to understand how they were overlooked for so long. The main reason, no doubt, is that specialists in American languages have each tended to focus on one language family. Thus, even if there were similar words running through family after family, nobody would notice them.

A good example is furnished by an Amerind root whose sounds were roughly TANA, TINA or TUNA and whose meaning fell somewhere in the range of "child, son, daughter" (the capital letters signify that the sounds are approximations). No one who careful-ly compares the vocabulary of Amerind languages from North and South America can fail to be impressed by the very high frequency of such terms.

How should we explain this broad distribution? One possibility might be that such terms appear around the world, as do words resembling "mama" and "papa." Unfortunately for this hypothesis, forms such as TANA and TUNA, with the meaning "son" or "daughter," are as rare outside Amerind as they are abundant within it. This root not only ties Amerind together but also distinguishes Amerind from other language families. It is, as linguists say, an exclusive innovation of the Amerind family.

Recent research by Ruhlen appears to explain why the first vowel of the root varies and why the root finds widespread use in words denoting both the sexes (son/brother and daughter/sister) and the neutral form (child/sibling). The reason is that Proto-Amerind, the original language from which all modern Amerind languages derive, had three forms, or grades, of the root in question, in which the first vowel was correlated with sex as follows: T'ANA "child, sibling," T'INA "son, brother, boy" and T'UNA "daughter, sister, girl." (The apostrophe represents a glottal stop after the "T"—a sound heard in the Cockney pronunciation of "bottle.")

As might be expected, in the 12,000 or more years since Amerind began to divide into subfamilies, the correlation between the initial vowel and the original gender has often been lost. As a result, many forms that are clearly cognates of the others now show the "wrong" vowel. One example of this kind is Proto-Algonquian *tána "daughter," where the first vowel is *á rather than *ú. (The asterisk signifies that the form has been reconstructed on the basis of the modern daughter languages.) Most likely this discrepancy is the result either of the first vowel assimilating the timbre of the second vowel or of the a-form of the root being extended, by analogy, throughout the language at the expense of the i- and u-forms. Such analogical extension is common in linguistic history. In English, for instance, the -ed form of the past tense of regular verbs (as in "kick/kicked") is extended by some speakers to the past tense of irregular verbs (as in "see/see'd").

It is noteworthy that the vowels i and u proposed for these masculine and feminine kinship terms coincide with the gender system in two major Amerind subgroups of South America and also in the Chinook language of Oregon. These agreements are too numerous to

	SANSKRIT	CLASSICAL GREEK	LATIN	OLD IRISH	GOTHIC
I carry	bhár-āmi	phér-ō	fer-ō	bir-u	baír-a
thou carriest	bhár-asi	phér-eis	fer-s	bir-i	baír-is
he carries	bhár-ati	phér-ei	fer-t	ber-id	baír-ith
we carry	bhár-āmas	phér-omen	fer-imus	ber-mi	baír-am
you carry	bhár-atha	phér-ete	fer-tis	ber-the	baír-ith
they carry	bhár-anti	phér-ousi	fer-unt	ber-it	baír-and

VERBAL VESTIGES of a common ancestor led William Jones, an 18th-century English jurist, to place these five ancient languages in one family, now called Indo-European. English is most closely related to Gothic.

LANGUAGE	FORM	MEANING	LANGUAGE	FORM	MEANING
PROTO-AMERIND	T'INA	"SON, BROTHER, BOY"	PROTO-AMERIND	T'U'NA	"DAUGHTER, SISTER, GIRL"
Yurok	t^{s}in	"young man"	Coeur d'Alene	tune	"niece"
Mohawk	-'tsin	"male, boy"	Yuchi	t^s'one	"daughter, son"
Molale	pnē-t'in	"my elder brother"	Central Sierra Miwok	tūne-	"daughter"
Yana	t'inī-si	"child, son, daughter"	Salinan	a-t'on	"younger sister"
Cuicatec	'díínó	"brother"	Taos	-t'út'ina	"older sister"
Changuena	sin	"brother"	Lenca	tuntu-rusko	"younger sister"
Millcayac	tzhœng	"son"	Cayapa	t^{s}uh-ki	"sister"
Tehuelche	den	"brother"	Tehuelche	thaun	"sister"
Tiquie	ten	"son"	Tiquie	ton	"daughter"
Mocochi	tin-gwa	"son, boy"	Morotoko	a-tune-sas	"girl"
Yagua	dēnu	"male child"	Nonuya	-tona	"sister"
Tacana	u-tse-kwa	"grandchild"	Tacana	-tóna	"younger sister"
Guaío	china	"older brother"	Piokobyé	a-ton-kä	"younger sister"

be accidental and too widespread to reflect linguistic borrowing. Indeed, many of them fall on either side of clear geographic discontinuities.

Just as Jones was impressed by the conjunction of roots and affixes, so too do we find in Amerind an equally impressive conjunction of the root in question and various grammatical affixes. Those that may modify the root T'ANA include the pronominal prefixes *na-* "my" and *ma-* "your," both of which appear in all 11 Amerind subgroups. The former appears in forms such as Proto-Algonquian *ne-tāna* "my daughter," Kiowa *nō-tōⁿ* "my brother," Paez *ne-tson* "my brother-in-law" and Manao *no-tany* "my son." Such pronominal affixes are among the most stable elements in language: they are almost never borrowed. That entire systems of them could have been systematically transmitted from one language to the next, from British Columbia to Tierra del Fuego, defies the imagination.

Amerind suffixes include diminutive forms that one naturally associates with words denoting children. The Proto-Amerind diminutive *-i'sa* is found in Proto-Algonquian *ne-tān-ehsa* "my daughter," Mixtec *tá'nù i'sá* "younger sister," Esmeralda *tini-usa* "daughter," and Suhin *tino-ice* "young woman." The Proto-Amerind diminutive *-mai* is seen in Luiseño *tu'-mai* "woman's daughter's child," Masaca *tani-mai* "younger sister" and Chapacura *tana-muy* "daughter."

Proto-Amerind deployed an intricate system of suffixes. One such suffix, *-ki*, indicated a reciprocal relation, such as that which makes a single word mean either a man's sister's son or a boy's mother's brother. This suffix—in conjunction with various roots to which it attaches—has been reconstructed for Proto-Siouan as *-thā-ki* "man's sister" and is seen in such modern languages as Pawnee *t'i-'i* "boy, son," Southern Pomo *t'i-ki* "younger sibling," Mazahua *t'i-'i* "boy," Amaguaje *-tsen-ke* "son" and Aponegicran *-thon-ghi* "sister."

The threefold classification of languages implies that no more than three Asian migrations left linguistic traces. Fewer migrations are possible if they gave rise to communities that split on the eastern side of the Bering Strait. To decide on the precise number, one must compare the language families of America and Asia.

Recent work by Russian and American linguists indicates that there probably were exactly three migrations. Eskimo-Aleut is the easternmost member of a vast family that we call Eurasiatic and that Russian scholars call Nostratic. (The two classifications differ slightly. Eurasiatic includes Indo-European, Uralic-Yukaghir, Turkic, Mongolian, Tungus, Korean, Japanese, Ainu, Gilyak, Chukchi-Kamchatkan and Eskimo-Aleut. Nostratic is broader, including also the Dravidian family of southern India, the Kartvelian family of the Caucasus and the Afro-Asiatic family of North Africa and the Middle East.)

Na-Dene's relatives in Asia were recently identified by Sergei Starostin of the Institute of Oriental Studies, Sergei Nikolaev of the Institute of Slavic Studies in Moscow and John Bengtson, an independent linguist in Minneapolis. Starostin began by connecting three Old World families that had hitherto been considered independent: Caucasian, Sino-Tibetan and Yeniseian (a family of central Siberia that has only a single surviving language). Nikolaev then showed that Na-Dene was unmistakably related to Caucasian (which he and Starostin had together reconstructed) and hence by extension to Sino-Tibetan and Yeniseian as well.

In a more comprehensive comparison of all relevant families, Bengtson added Basque (an isolated language of northern Spain) and Burushaski (an isolated language of northern Pakistan) to this family, which has come to be called Dene-Caucasian. Na-Dene proves to be the easternmost extension of Dene-Caucasian. Because that family is distinct from Eurasiatic, Na-Dene could not have split from Eskimo-Aleut in the Western Hemisphere. It must have reached the Americas by means of a separate migration.

Over the past few years, we have compared Amerind with the world's other language families and found that it is most closely related to Eurasiatic. The taxonomic relation is quite distant: whereas Eskimo-Aleut is a member of the Eurasiatic family, Amerind is related to Eurasiatic as a whole. That is, its genetic connection reaches much further back in time.

The first migration, known on archaeological grounds to have occurred some time before 12,000 B.P., gave rise to the Amerind family, which occupied most of the New World at the time of Columbus's arrival in 1492. The second migration, somewhat later, gave rise to the Na-Dene family. Finally, perhaps 4,000 to 5,000 years ago, the final migration took place, bringing the ancestors of the Eskimo and Aleut first to

southwestern Alaska and later across the northern perimeter of North America to Greenland.

A single etymology can illustrate both the unity of Amerind and its ties to the Eurasiatic/Nostratic constellation. The Proto-Amerind root MALIQ'A, meaning "swallow, throat," has left its mark in no fewer than eight of the 11 Amerind subfamilies from Canada to the tip of South America [*see illustration below*]. In Canada's Salish subfamily we find Halkomelem *məlqw* "throat." Down the coast in Oregon we find in Tfaltik, an extinct language of the Penutian subfamily, *milq,* which means "swallow." In Yuman, a subdivision of the Hokan subfamily, this root has become the general word for "throat." In Arizona we find Mohave

mal^yaqe "throat," whereas Akwa'ala, in Baja California, has *milqi* "neck." In Panama, Cuna has *murki* "swallow," where the original *l* has apparently changed to *r,* a very common replacement. In the Andean subfamily the Quechua language has *malq'a* "throat"; in the Equatorial subfamily, the Guamo language has *mirko* "drink."

What is the probability that these similar forms arose independently? One can make a rough estimate by holding the meaning within the narrow semantic range "swallow-throat" and making a number of phonological assumptions. Let us begin by assessing the probability that the Halkomelem and Tfaltik forms resemble each other by accident. Disregard the vowels as less stable than consonants and calculate the chances that the three consonants will accidentally

match. Next, limit both languages to only the following consonants: $p, t, t',$ $k, k', q, q', s, m, n, l, r, y, w.$ Accept only m for the first consonant, l or r for the second position and k, k', q, q' for the third consonant.

Under these assumptions, the chances of an accidental match are $(1/13)(2/13)(4/13) = 0.0036413291$. If we then round this off to 0.004 and calculate the probability for a random similarity among six families, we obtain $(0.004)^5 = 0.0000000001024$, or about one chance in 10 billion. These rough calculations assume an equal probability of all consonant types. Because this assumption does not hold, the figure will actually be somewhat larger, yet it will still be of the same infinitesimal order of magnitude. So much for accidental resemblances.

LANGUAGE FAMILY		LANGUAGE	FORM	MEANING
AFRO-ASIATIC		Proto-Afro-Asiatic	*mlg	"to suck, breast, udder"
		Arabic	mlj	"to suck the breast"
		Old Egyptian	mnd^y	"woman's breast, udder"
INDO-EUROPEAN		Proto-Indo-European	*melg-	"to milk"
		English	milk	"to milk, milk"
		Latin	mulg-ēre	"to milk"
URALIC		Proto-Finno-Ugric	*mälke	"breast"
		Saami	mielga	"breast"
		Hungarian	mell	"breast"
DRAVIDIAN		Tamil	melku	"to chew"
		Malayalam	melluka	"to chew"
		Kurux	melkhā	"throat"
ESKIMO-ALEUT		Central Yupik	melug-	"to suck"
AMERIND		Proto-Amerind	*maliq'a	"to swallow, throat"
	Almosan	Halkomelem	melqw	"throat"
		Kwakwala	m'lXw-'id	to chew food for the baby"
		Kutenai	u'mqolh	"to swallow"
	Penutian	Chinook	mlqw-tan	"cheek"
		Takelma	mülk'	"to swallow"
		Tfaltik	milq	"to swallow"
		Mixe	amu'ul	"to suck"
	Hokan	Mohave	mal^yaqé	"throat"
		Walapei	malqi'	"throat, neck"
		Akwa'ala	milqi	"neck"
	Chibchan	Cuna	murki-	"to swallow"
	Andean	Quechua	malq'a	"throat"
		Aymara	mal^yq'a	"to swallow, throat"
	Macro-Tucanoan	Iranshe	moke'i	"neck"
	Equatorial	Guamo	mirko	"to drink"
	Macro-Carib	Surinam	e'mōkï	"to swallow"
		Faai	mekeli	"nape of the neck"
		Kaliana	imukulali	"throat"

OLD WORLD TIES appear in the etymology of the extremely ancient root MALIQ'A, whose meaning was close to "swallow" or "throat." Cognates appear in eight Amerind branches and in more than one language from each of the listed Old World families. The chances that such resemblances could have occurred by accident are vanishingly small.
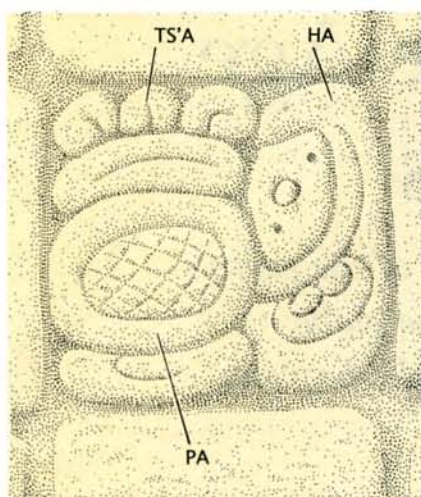
AMERICAN LANGUAGES AND PEOPLES form two correlated family trees, the one based on etymologies, the other on genes. Amerind speakers include the Maya, who carved the glyph for *ts'apah,* meaning "was set upright," more than 1,000 years ago (*left*). Among the Na-Dene speakers are the Apaches, who were led in the 19th century by Geronimo (*center*). Eskimo-Aleut speakers, including these Inuit from Canada's Northwest Territory (*right*), range from Siberia to Greenland.

Let us turn now to the question of whether this root can be found in the Old World. As we saw earlier in the case of T'ANA "child," there is no guarantee that elements widespread in Amerind will be found outside that family. In this case, however, cognate forms of this root are scattered through the Old World. The original Russian Nostraticists, the late Vladislav Illich-Svitych and Aaron B. Dolgopolsky (now at the University of Haifa), have reconstructed a Nostratic root *mälgi "to suck the breast, to nurse." This root connects Proto-Afro-Asiatic *mlg "to suck the breast" (as in the Arabic *mlj*), Proto-Indo-European *melg- "to milk," as well as the noun "milk" and Proto-Finno-Ugric *mälke "breast" (as in Saami *mielgâ*). We have found cognate forms in Eskimo-Aleut such as Central Yupik *melug-* "to suck." Finally, the Dravidian family displays apparent cognates in forms such as Kurux *melkhā-* "throat" and Tamil *melku* "to chew."

The range in meaning displayed by these families suggests that the ultimate ancestor of this root meant "to nurse, to suck the breast," a meaning preserved in Afro-Asiatic. In Indo-European there was a slight semantic shift from the notion of nursing to that of milking, whereas Uralic shows a different shift: to the noun "breast." In Dravidian the meaning has shifted to "chew," a natural semantic connection for anyone who has ever watched a baby nursing, and "throat." In Eskimo the meaning has become "to suck" in general, without specific reference to the female breast. Finally, in Amerind this root became the general word for "to swallow" and "throat."

Support for the Amerind hypothesis came from an unexpected quarter in 1988, a little more than a year after it was first announced. A team of geneticists led by L. L. Cavalli-Sforza of the Stanford University School of Medicine discovered that Native Americans fell neatly into three distinct groups whose boundaries essentially coincided with those of their respective language families [see "Genes, Peoples and Languages," by L. L. Cavalli-Sforza; SCIENTIFIC AMERICAN, November 1991]. This independent corroboration virtually confirms the validity of the Amerind family because the probability that the biological and linguistic classifications would coincide fortuitously is very small indeed.

Yet a third line of evidence supporting a tripartite classification of Native Americans has been developed by Christy G. Turner II of Arizona State University. A specialist in human dentition, Turner found that on the basis of their teeth, New World populations fall into the same three groups [see "Teeth and Prehistory in Asia," by Christy G. Turner II; SCIENTIFIC AMERICAN, February 1989]. Finally, in 1990 Douglas C. Wallace of the Emory University School of Medicine reported preliminary results of the analysis of mitochondrial DNA in Native American populations, and this analysis also appears to support the Amerind hypothesis.

We must hasten to add that the close correspondence of biological and linguistic classifications does not mean that genes determine the language one speaks. That depends solely on the community in which one is raised. The classifications correspond because the same processes that lead to linguistic divergence also give rise to genetic divergence. When a group of people depart from their homeland and move, say, to some distant island, they take with them both their language and their genes. From this time on, their language and their gene pool will diverge from those of the group left behind. It is for this reason that the classifications correspond so nicely.

The evidence of comparative linguistics indicates that the Americas were originally settled by three major migrations from Asia. There are, of course, many unresolved problems, such as how the Amerind family initially broke up in its spread through North and South America. But the recent discoveries at least, in part, fulfill Jefferson's hope that one day the languages of Native Americans would illuminate their relations to one another and reveal the Asian origins of the first Americans.

FURTHER READING

THE SETTLEMENT OF THE AMERICAS: A COMPARISON OF LINGUISTIC, DENTAL, AND GENETIC EVIDENCE. Joseph H. Greenberg, Christy G. Turner II and Stephen L. Zegura in *Current Anthropology,* Vol. 27, pages 477–497; 1986.

LANGUAGE IN THE AMERICAS. Joseph H. Greenberg. Stanford University Press, 1987.

THE AMERICAN INDIAN LANGUAGE CONTROVERSY. Joseph H. Greenberg in *Review of Archaeology,* Vol. 11, No. 2, pages 5–14; Fall 1990.

A GUIDE TO THE WORLD'S LANGUAGES, Vol. 1: CLASSIFICATION. Merritt Ruhlen. Stanford University Press, 1991.

EVOLUTION OF HUMAN LANGUAGES. Edited by John Hawkins and Murray Gell-Mann. Addison-Wesley, 1992.