

13

The Origin of Language: Retrospective and Prospective

*Il linguaggio . . . è l'archivio più
copioso e più sicuro dell'umanità.*

—Alfredo Trombetti (1905)

The question of the origin of language—all discussion of which was banned by the Société de Linguistique de Paris in 1866—has always been one of the few linguistic problems of interest to the general public.* Perhaps because of this widespread but uncritical attention, coupled with much amateurish work of no scientific value and the general reluctance of scholars to deal with such matters, the question has remained in a muddled state, its various components often not even clearly distinguished. I will explore here but a single aspect of the question, namely, whether or not all of the world's presently extant languages share a common origin. I will be seeking neither the *locus* of that origin nor its *antiquity*.

THE PROBLEM

The striking parallels between biology and linguistics, particularly in their evolutionary aspect, have been generally recognized since at least the mid-nineteenth century. In one of his few references to language, Charles Darwin

* A preliminary version of this chapter was presented at an International Conference on Language Change and Biological Evolution, Torino, Italy, May 1988. A Russian translation was published in *Voprosy Jazykoznanija* 1 (1991): 5–19.

pointed out in 1871 that “the formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel.” Were it not for these “curious” similarities it is doubtful that biologists and taxonomic linguists would ever conceive of a joint conference; historians and mathematicians seldom confer. Yet my focus here is on an area where the biological and linguistic perspectives appear to clash rather than to complement one another. For biologists the monogenetic origin of *Homo sapiens sapiens* is now generally accepted (though supporters of “Multiregional Evolution” would dispute this point), and for them, the notion that the Indo-European peoples have no known biological relatives would be ludicrous. Yet for most linguists a common origin of all human languages is very much in doubt, and the belief that Indo-European has no known linguistic relatives is not only a safe position, but practically a merit badge for sober scholarship. In practice, if not in theory, the linguistic approach is pre-Darwinian, in the sense that dozens, or even hundreds, of linguistic taxa are treated *as if* they were historically independent developments. Linguists seldom go so far as to deny the possibility that all these taxa are ultimately related; what they deny is that there is any *linguistic evidence* for such a hypothesis.

To be sure, the monogenetic origin of *Homo sapiens sapiens* need not *necessarily* entail the monogenesis of human language; the two topics are, and should be kept, distinct, and when we find correlations between biology and linguistics, we insist that these correlations be arrived at independently. Yet there is something strange about the spectacle of hundreds of supposedly unrelated language families, when the biological differences among the people who speak the tongues of the various language families are often minuscule. Surely no one imagines that each of these hundreds of language families represents an independent creation of language. But if they are not independent developments is it plausible, or even possible, that they have all been separated from each other for so long that any trace of deeper relationships has vanished? Is it not also strange that the comparative method, which was discovered in its broad outlines over two centuries ago, largely in terms of the Indo-European family, has never been able to connect that family with any other—at least to the satisfaction of the linguistic community? This mystery is compounded by the fact that Indo-European is in no sense an archaic or poorly distinguished family.

I believe the general rejection of attempts to connect Indo-European with other families, encouraged in an earlier day by chauvinistic arrogance, has effectively blocked consideration of the question of monogenesis by acting as a dike against *all* long-range comparison. For if Indo-European—that most studied and best understood of all families—cannot be convincingly connected

with any other family, then what confidence can we have in connections proposed between even less well studied families, often with a postulated time depth many times that of Indo-European? Thus if the splendid genetic isolation of Indo-European can be maintained, the question of monogenesis is moot.

Although the Paris interdiction of the study of the origin of language expressed primarily the European disenchantment with work on this topic, at a time when Europe dominated the world, culturally and otherwise, in the United States the historical linguist William Dwight Whitney (1867: 383) was no less pessimistic: “Linguistic science is not now, and cannot ever hope to be, in condition to give an authoritative opinion respecting the unity or variety of our species.” Although this remains the conventional wisdom among most linguists to the present day, I believe it is largely a myth that scholars have passed on uncritically from generation to generation for over a century. Aside from barely concealed racism, there are multiple academic and institutional reasons why this myth has persisted so long, the two most responsible being the predominantly synchronic slant of twentieth-century linguistics and the ever-increasing grip of specialization on academia. Although historical linguistics has never altogether succumbed, and indeed has always been considered an integral part of general linguistics, there can be no doubt that during most of the present century it has taken a back seat to synchronic exploration. Moreover, even *within* historical linguistics, the increasing pace of publication and specialization has discouraged most scholars from seeking to keep abreast of developments outside of their own family or language of interest. There has been a tacit assumption that developments in, say, Bantu linguistics can be safely ignored by Romance scholars.¹

AN ALTERNATIVE VIEW

Despite this generally hostile intellectual climate, and the risk of provoking one’s colleagues to consternation or condemnation, there have always been some scholars who rejected the Paris edict and Whitney’s admonition and sought to find evidence of more comprehensive classifications of the world’s languages. Six such scholars bear mention here. In several works published during the first quarter of the twentieth century, the Italian linguist Alfredo Trombetti sought to establish the monogenesis of human language by comparing lexical and grammatical roots from languages and language families around the world. Though one can hardly deny that some of Trombetti’s

¹ In Ruhlen (1979) I sought to challenge that notion. See also the remarks of Henry Sweet (1901) quoted in Chapter 6 herein.

proposals were incorrect, many others were later adopted (or independently discovered) and elaborated by other scholars. As early as 1905 he presented a strong *prima facie* case for the monogenesis of human language.

In the New World, from roughly 1910 to 1930, the American linguist Edward Sapir made a number of sweeping proposals for the consolidation of numerous Native American language families that had been identified only decades before, during the nineteenth-century catalog stage of linguistic taxonomy. Sapir was also the first to propose genetic affinity between Na-Dene and Sino-Tibetan, a connection that has recently been revived and extended (see Chapters 1 and 4). For a full discussion of Sapir's many contributions to Amerindian classification, see Ruhlen (1987).

Morris Swadesh, a student of Sapir's, shared both his mentor's interest in Amerindian linguistics and Trombetti's passion for global exploration. Though his early work was much in the spirit of Sapir, giving etymologies for various proposed groups, in his later work he shifted to his own technique of lexicostatistics, with mixed results. During this later phase he became interested in establishing a worldwide network of linguistic relationships, and the problem of a hierarchical classification of the world's languages became less important to him. His premature death in 1967 meant that his life's work was never adequately realized or summarized.

During the early 1960's two Russian scholars, Vladislav Illich-Svitych and Aron Dolgopolsky, revived an earlier proposal of the Dane Holger Pedersen that grouped Indo-European with several other families of Eurasia and North Africa in a Nostratic phylum. Though they had at first been unaware of each other's work, their results coincided to such an extent that Nostratic theory became a single unified field, and since Illich-Svitych's tragic death in 1966 the work of the field has been carried on by Dolgopolsky and others. In classical Nostratic theory, Indo-European is one of six related subgroups, the others being Afro-Asiatic, Kartvelian, Uralic, Altaic, and Dravidian. Roughly 400 etymologies supporting the Nostratic grouping have so far been published. Recently Dolgopolsky (1984) has proposed the inclusion of Elamite, Gilyak, and Chukchi-Kamchatkan as well.

The American linguist Joseph Greenberg has probably made the greatest contributions of all to linguistic taxonomy. Beginning with his revolutionary classification of African languages in the 1950's, he undertook to investigate those regions of the world where linguistic classification was least advanced. By 1963 he had classified all African languages into four phyla (Khoisan, Niger-Kordofanian, Nilo-Saharan, and Afro-Asiatic), and it is this classification that forms the basis of all contemporary research in African linguistics. In 1971 Greenberg offered evidence for an Indo-Pacific phylum that includes the very diverse Papuan languages of New Guinea and surrounding islands,

and in 1987 he presented substantial evidence for an Amerind phylum that would include all New World languages except those belonging to the Na-Dene and Eskimo-Aleut families. He is currently at work on a book on Eurasiatic, a vast grouping that differs from Nostratic by the exclusion of Afro-Asiatic, Kartvelian, and Dravidian, and by the inclusion of Japanese, Ainu, Gilyak, Chukchi-Kamchatkan, and Eskimo-Aleut.

A CLOSER LOOK AT THE PROBLEM

Despite the pathbreaking work of these scholars, the majority of the linguistic community still adheres to the belief that Indo-European has no known linguistic relatives, and none is likely ever to be demonstrated, because—so the argument goes—beyond the time depth of Indo-European all trace of genetic affiliation has been obliterated by ceaseless phonetic and semantic erosion. This belief is so strong that even linguists in possession of evidence to the contrary will often provide an ad hoc explanation of the contradictory evidence rather than challenge the reigning tenets of comparative-historical linguistics.

The Australian language phylum is instructive in this regard. Humans have occupied Australia continuously since at least 40,000 years BP (before present), and there is no reason to think that Proto-Australian does not date from about that time (though the fact that Australia was not permanently cut off from New Guinea until about 10,000 years ago must be borne in mind). At the very least, Proto-Australian must be twice as old as Indo-European, and more likely seven or eight times as old. At this time depth, evidence of a primitive unity can no longer exist, according to the standard view of linguistic evolution, and yet the Australian phylum is universally accepted as a valid taxon.

In order to reconcile this clear contradiction R. M. W. Dixon (1980) discarded the principle of linguistic uniformitarianism and proposed that, because of their isolation, Australian languages have changed much more slowly than have languages elsewhere in the world:

Proto-Australian was probably spoken a considerable time in the past, perhaps some tens of millennia ago. It is this which makes it unlikely that it will ever be possible to demonstrate a genetic connection between Australian and any other language family. Any sister language that Proto-Australian may have left behind in South-East Asia, say, is likely to have changed out of all recognition over the intervening period, so that there would be insufficient points of similarity remaining for any connection to be recognizable. (Or it could well be that relatives of Proto-Australian have NO living descendants.) Generally, languages change at such a rate that after more than about three or four thousand years of separation genetic links are no longer recognizable. Australian languages have been relatively isolated from contact with other languages and cultures, and may well have changed at a comparatively slow

rate; but any relative that they left behind in regions that were more linguistically cosmopolitan would not have been sheltered in this way. (p. 237)

This romantic notion of Australia as a Land That Time Forgot is most certainly unsupportable, and one should not lose sight of the fact that it is based on expectations rather than evidence. In any event, Dixon is categorical in his belief that “there is absolutely no evidence for a genetic connection between Australian languages and anything outside the continent; there is not even any remote ‘possibility’ that scholars could argue about. It seems that the languages of Australia have been so long in their present location that any evidence of connection with other languages has been, through time, eroded away.” (p. 238)

Though the number of roots that have been reconstructed for Proto-Australian is rather small, some of these would appear to be cognate with roots found in other families (see Chapter 14). Consider, for instance, Proto-Australian **bungu* ‘knee,’ which in various modern languages has semantic extensions to things that bend (wave, bend in a river, hump in a snake’s body). This form is very similar in sound and meaning to the Indo-Pacific etymology for ‘knee,’ which includes forms such as Tobelo *buku*, Koianu *poku*, and Teri Kawalsch *bugu*. Traces of this root are also found in Eurasia. Ainu (*he-*) *poki(-ki)* ‘bow down’ appears to belong here, as do Proto-Indo-European **bheug(h)* ‘to bend,’ and Proto-Altaiic **bük(ä)* ‘to bend’ (including forms such as Uighur *bük* ‘to kneel,’ Yakut *bük* ‘to bend,’ Khalkha *bóx(ón)* ‘hump of a camel,’ and Evenki *buku* ‘bent, crooked’). In Africa, Proto-Bantu **bóngó* ‘knee’ is virtually identical with the Australian form in both sound and meaning, and in the West Atlantic branch of Niger-Congo we find forms such as Baga *-buŋ* ‘knee.’ One may anticipate that additional reflexes of this root will be found elsewhere in the Niger-Congo family, but the lack of any kind of Niger-Congo etymological dictionary makes this difficult to verify for the moment. Finally, this same root is well attested in the Amerind family, where we find North American forms like Chumash (*si-*)*buk* ‘elbow’ and Walapai (*mi-*)*puk* ‘knee’ and South American forms like Guamaca *buka* ‘knee, elbow’ and Iranshe *poku* ‘bow’ (n.).

This example is by no means the only genetic connection between the Australian phylum and the rest of the world’s languages. Dixon reconstructs **bula* ‘2’ for Proto-Australian, and Blake (1988) shows how this number has been used to form dual pronouns in the Pama-Nyungan subgroup: **nyuN-palV* ‘you-2’ and **pula* ‘they-2.’ Two of the extinct Tasmanian languages (considered by Dixon unrelated to Australian languages) exhibit similar forms, Southeastern *boula* ‘2’ and Southern *pooalih* ‘2.’ In the context of his Austro-Tai hypothesis Paul Benedict (1975) pointed out the similarity of the number 2 in all of the major families of Southeast Asia. Benedict reconstructs **ʔ(m)bar*

'2' for Proto-Austroasiatic (cf. Santali *bar*, Jeh *bal*, Khmu' *bār*, Old Mon *ʔbar*) and *(a)*war* '2' for Proto-Miao-Yao. He also considers Daic forms like Mak *wa* 'twin' and Austronesian forms like Javanese *kēmbar* 'twin' to be cognate with the preceding. In Africa one of the pieces of evidence that Edgar Gregersen (1972) offered in support of Congo-Saharan (his proposal for joining Niger-Kordofanian and Nilo-Saharan in a single family) was forms for the number 2 that hardly differ from those we have seen so far. In Niger-Congo we have Temne (*kə*)*bari* 'twin,' Nimbari *bala* '2,' Mano *pere* '2,' and Proto-Bantu **bàdí* '2'; Nilo-Saharan has forms such as Nubian *bar(-si)* 'twin,' Merarit *warē* '2,' and Kunama *barā* 'pair.' In Eurasia one of Illich-Svitych's Nostratic etymologies appears related to the forms discussed so far, but in these families the meaning has shifted from '2' to 'half,' 'side,' and 'part.' Specifically, Illich-Svitych (1967) connects Proto-Indo-European **pol* 'half, side' (cf. Sanskrit (*ka-*)*palam* 'half,' Albanian *palë* 'side, part, pair,' Russian *pol* 'half,') with Proto-Uralic **pälä*/**pole* 'half' (cf. Yurak Samoyed *peelee* 'half,' Hungarian *fele* 'half, one side of two,' Vogul *pääl* 'side, half,' Votyak *pal* 'side, half') and Proto-Dravidian **pāl* 'part, portion' (cf. Tamil *pāl* 'part, portion, share,' Telugu *pālu* 'share, portion,' Parji *pēla* 'portion'). Finally, cognate forms are found in Amerind languages of North and South America (cf. Wintun *palo(-l)* '2,' Wappo *p'ala* 'twins,' Huave *apool* 'snap in two,' Colorado *palu* '2,' Sabane *paʔlin* '2').

The final piece of evidence I would like to offer for the proposition that the Australian phylum is demonstrably related to the rest of the world's languages involves an interrogative whose most usual form is *mi(n)* or *ma(n)* and whose meaning is 'what?, who?,' or some other interrogative. This root has been discussed in the work of Trombetti, Illich-Svitych, and Greenberg and appears to be one of the most broadly distributed formatives in human language. For Proto-Australian Dixon reconstructs **miNH*² 'what,' with modern reflexes such as Dyirbal *minya* 'what' and Pitta-Pitta *minha* 'what.' These forms are strikingly similar to those contained in one of Greenberg's Indo-Pacific etymologies that includes forms such as Matap *mina* 'what,' Arapesh *mane* 'what,' Nyaura *məndə* 'what, thing,' Kati *man* 'something,' Biada *min* 'thing,' and Laumbe *mina* 'thing.'

In Eurasia there are a variety of forms that are in all likelihood cognate with those just mentioned. In the Austroasiatic phylum one can point to Kurku *amae* 'who,' Mon *mu* 'what,' Central Sakai *mā/mō* 'what.' Two isolated languages of the Indian subcontinent, Burushaski and Nahali, show reflexes of the interrogative under discussion. Burushaski has *men* 'who' and

² *NH* represents a correspondence between lamino-interdental *nh* and lamino-palatal *ny* that is found in the modern Australian languages.

amin 'which'; Nahali, *miŋ gay* 'where' and *miyan* 'how much.' In the Caucasus **ma* has been reconstructed as an interrogative particle in Proto-(North) Caucasian, and some languages show pronominal use of the *m* interrogative as well, e.g. Chechen *mila* 'who' and Bats *me* 'who.' For Kartvelian, G. Klimov (1964) has reconstructed **ma* 'what' and **mi-n* 'who.' Across Northern Eurasia the *m* interrogative is widely attested, its numerous occurrences having been emphasized in both Illich-Svitych's Nostratic etymologies and Greenberg's Eurasiatic etymologies. Examples include Proto-Indo-European **mo-*, a base of interrogative adverbs; Proto-Uralic **mi* 'what' (cf. Vogul *män* 'which, what,' Tavgy *ma* 'what,' Hungarian *mi* 'what, which,' Finnish *mi/mi(-kä)* 'what, which'); Proto-Turkic **mi* 'what' (cf. Chuvash *měn* 'what' and Turkish *mi*, a sentence enclitic); Mongolian *-ū* (< **mu*), a sentence interrogative, and Monguor *amu/ama* 'what'; Tungus *-ma*, an indefinitizer; Korean *muōt* 'what' and Old Korean *mai* 'why'; Ryukyuan (the language most closely related to Japanese) *mī* 'what'; Ainu *mak/makanak* 'what,' *makan* 'what kind'; Chukchi *mikin* 'who,' Kamchadal *min* 'which, what sort.'

In the New World the *m* interrogative is not found, to the best of my knowledge, in either Eskimo-Aleut or Na-Dene, but it is widespread in Amerind. North American examples include North Sahaptin *mēn/mna* 'where,' Central Sierra Miwok *minni* 'who,' San Jose *mani* 'where,' Choctaw *mana* 'when,' and Chickasaw *mano* 'where.' In South America we have Kagaba *mai* 'who,' *mani* 'where,' Paez *maneh* 'when,' Allentiac *men* 'who,' Catio *mai* 'where,' Guajajara *mon* 'who,' Maripu *manub* 'in which direction,' Cofan *mañi* 'where,' Krenje *menō* 'who,' and Botocudo *mina* 'who.'

In Africa the *m* interrogative is widely attested in Afro-Asiatic and is perhaps found in Khoisan as well. The Khoisan examples are relatively few (cf. Kxoe *má* 'who,' *má* 'which,' Naron *kama* 'if, when,' Nama *maba* 'where'), but in Afro-Asiatic it is attested in every branch of the family. Examples include Akkadian *mīn* 'what,' *mann* 'who,' Amharic *mīn* 'what,' Arabic *man/min* 'who,' Tuareg *ma* 'what,' *mi* 'who,' Saho *mā* 'what,' *mi* 'who,' Somali *máḥâ* 'what,' Oromo *máni* 'what,' Kaffa *amone* 'what,' Hausa *mè/mì* 'what,' Bata *mən* 'what,' and Logone *mini* 'who.'

In all of the above forms I would maintain that the initial *m*-portion of the root is cognate. The final *-n* portion, however, has multiple sources, one of which is locative. A *k* interrogative, which can be seen in some of the forms above, is the chief rival of the *m* interrogative in the world's languages, sometimes joining with it, sometimes supplanting it, and sometimes being replaced by it. The complex interplay of these two interrogatives (as well as a third, *j*) has been discussed by Trombetti, Illich-Svitych, and Greenberg, all of whom have called attention to the generally personal character of the *k* interrogative (who?), which contrasts with the generally nonpersonal character of the *m* in-

terrogative (what?). Finally, it may be significant that despite the enormous geographic distribution of this root—from Africa to the Americas—it appears not to occur in Niger-Kordofanian or Nilo-Saharan, a point to which I will return.

Notwithstanding the very widespread distribution of these roots, many of which have been known from at least the time of Trombetti, they have had little impact on comparative-historical linguistics, which, from a theoretical perspective, has remained devoutly Eurocentric. Indo-European, as an examination of any of the introductory textbooks will reveal, is still almost universally regarded as unrelated to any other family. The Americas have actually seen retrogression in linguistic taxonomy during this century, the number of supposedly independent families growing to over 200 by the time that Greenberg (1987) offered compelling evidence that there are in fact just three (see Chapter 6). In Asia many Altaicists have come to reject any connection among the three Altaic branches, and even the Uralic affinity of Yukaghir is sometimes denied or questioned. The reasons for the rejection of more comprehensive classifications are multiple and have been amply discussed in Greenberg (1987) and Ruhlen (1987). Rather than cover that ground once again I will merely cite a few particular criticisms that exemplify the flavor of the opposition to long-range comparison.

Traditionally, the most common criticism of long-range comparison has been that the proposed cognates were semantically and/or phonologically too disparate to be historically related. This is hardly the case with the etymologies discussed above, which are in fact criticized as being *too similar* in sound and meaning to be related at the time depths proposed. The putative cognates cited in these etymologies are sufficiently extensive and sufficiently similar in sound and meaning that their genetic affinity would not be controversial if they represented merely some group of South American Indian languages, but because they represent instead supposedly unrelated language families from every corner of the earth the actual substance of the etymologies is given less weight than are the expectations of scholars about what genetic connections are possible. What would be considered an obvious etymology at a low level of classification becomes “random noise” at higher levels—with no change in content at all. Trombetti (1905: 44) criticized this inconsistency as follows: “It is clear that in and of itself the comparison of Finno-Ugric *me* ‘I,’ *te* ‘you’ with Indo-European *me-* and *te-* [with the same meaning] is worth just as much as any comparison one might make between the corresponding pronominal forms in the Indo-European languages. The only difference is that the common origin of the Indo-European languages is accepted, while the connection between Indo-European and Finno-Ugric is denied.”

No one would deny that the member languages of low-level groups like Romance often display cognates that are very similar or even identical. What is in dispute is (1) whether supposedly independent, higher-level groupings (e.g. Indo-European, Australian, and Amerind) can share cognates that are similar in form and meaning and (2) whether a reconstructed proto-language (e.g. Proto-Indo-European, Proto-Nostratic, or Proto-Australian) can show reflexes in its extant daughter languages that are similar or identical to the reconstructed form. The answers to both of these questions depend on the rate and nature of linguistic change. As Dixon's comments indicate, many linguists believe that the rate of linguistic change is such that all trace of genetic affiliation is effaced after only several thousand years, so for him the answer to both questions is no. But if all of the supposedly independent linguistic families do derive from a common origin, then the fact that the earliest reconstructable items in the various families look alike should hardly be surprising. *Greater convergence with greater depth is what one would expect.*

As regards the second question I would point out only that, in every etymological dictionary I have examined, some of the reconstructed forms for the proto-language are similar or identical to some of the reflexes in its extant daughter languages. Pokorny (1959) reconstructs Proto-Indo-European **nepōt* 'nephew, grandson,' a form that must have existed at least 5,000 years ago. Yet this same form, with the same meaning, is preserved to this day intact in Rumanian *nepot* 'nephew, grandson.' At least in this instance Dixon's inexorable erosion seems not to have taken place. And even phyla that are much older than Indo-European show the same phenomenon. In the first etymology above, for example, Proto-Australian **bungu* 'knee' shows the reflex *bungu* 'knee' in many modern languages (e.g. Guugu Yimidhir, Yidiny, Dyirbal). Now if Proto-Australian, which in all likelihood dates from 40,000 BP or more, can be accorded reflexes in contemporary languages that are identical to the reconstructed form, on what grounds can one object to a similar phenomenon between Proto-Sapiens and modern languages, given that Proto-Sapiens could be only 20,000–30,000 years older than Proto-Australian?

Furthermore, the assumption that linguistic change has been constant and continuous since the emergence of *Homo sapiens sapiens* may be incorrect. It is well known among anthropologists, archaeologists, and even historians that cultural evolution in general appears to have developed at an ever accelerating pace as one approaches the present, and the same may well be true for linguistic evolution. Biological evolution, too, is no longer necessarily conceived of as a very long, slow process of gradual and constant change; scholars like Niles Eldredge and Stephen Jay Gould have argued instead for a more episodic character to evolution, in which little change may occur over very long periods of time (see Eldredge 1985), and recent research on catastrophes

and mass extinctions tends to support that mode. Given that we do not *know* what linguistic evolution was like during the past 100,000 years, it would seem premature to rule out any of the possibilities on a priori grounds.

Some linguists, of course, are simply unaware that other language families often have roots similar to those in the family they are interested in, and I suspect that this is the case with Dixon. Other linguists, however, are aware of such roots but choose to ignore them. One of the most cogent pieces of evidence that Greenberg (1987) offered in support of the Amerind phylum was the presence of first-person *n* and second-person *m* in all eleven branches. As noted in Chapter 12, the first- and second-person pronouns are known to be among the most stable meanings over time. Dolgopolsky (1964) found that the first-person pronoun is the most stable item, and the second-person pronoun ranked third in stability (following the number 2). It is also well known that initial nasal consonants are among the most stable sounds, and the conjunction of stable sounds with stable meanings has meant that even after 12,000 years these pronouns have been preserved in every branch of the Amerind phylum. Greenberg did not claim to be the first to notice the broad distribution of these two pronouns in North and South America. Swadesh (1954) had underscored their distribution in an article containing additional evidence for Amerind (not yet so named), and a year later Greenberg, unaware of Swadesh's article, discovered the same distribution. Greenberg observes, "That two scholars should independently make the same basic observation is an interesting sidelight in the argument for the Amerind grouping as I have defined it" (1987: 54).

Lyle Campbell, an Amerindian scholar and one of Greenberg's chief critics, sees things differently: "The widespread first-person *n* and less widespread second-person *m* markers . . . have been recognized from the beginning without significant impact on classification" (Campbell 1986: 488). Lamentably, Campbell is correct, but that such crucial evidence has been overlooked—or, worse, scorned—is not something to take pride in. Were a biologist to remark smugly, "That group of animals you keep mentioning, the ones with a backbone, has been recognized for a long time and I am not impressed," his colleagues would chuckle and move on to other business. Here we see perhaps one measure of the difference between biology and linguistics, especially as they present themselves today. Greenberg (1987) summarizes the fundamental and obvious importance of the two Amerind pronouns as follows:

It is the business of science to note non-random phenomena and to explain them. Were we to plot the occurrence of specific first- and second-person markers on a world map, we would not fail to notice a clustering of first-person *m* and second-person *t* (along with *s*) in Europe, northern Asia, and the northern part of North America as far as Greenland, with a second clustering of first-person *n* and second-person *m* covering the rest of the Americas, outside of the Eskimo-Aleut and Na-Dene areas.

In my opinion, this observation alone would suffice to lead any historically minded anthropologist to the view that there must be at least one very large stock to account for the first set and another to account for the second set. (p. 55)

AN END TO MYTHOLOGY

I have suggested here that the currently widespread beliefs, first, that Indo-European has no known relatives, and, second, that the monogenesis of language cannot be demonstrated on the basis of linguistic evidence, are both incorrect. Belief in these erroneous assertions is based largely on extralinguistic criteria and a priori assumptions, rather than on a serious survey of the world's linguistic literature. A growing, though still small, number of linguists are coming to realize that all the world's languages do share a common origin, and they are beginning to work on that basis. In the remainder of this chapter I shall discuss several implications of monogenesis for linguistic taxonomy.

First, the search for linguistic "relationships" is now over (or should be), since it no longer makes sense to ask if two languages (or two language families) are related. *Everything* is related, and the question to be investigated within or among different families is the *degree* of their relationship, not the fact of it. All taxonomic questions dissolve into one: discovering the hierarchical subgrouping of the human family on the basis of linguistic traits. Such traits may be either lexical (i.e. roots, affixes) or typological (e.g. nasal vowels, the SOV word order, an inclusive/exclusive distinction for 'we'). The use of lexical evidence to support a particular subgrouping needs no justification, since it has long been an essential technique of the comparative method. It is the *total distribution* of a root that reveals its taxonomic significance, not its mere presence in this or that family. Within that total distribution, particular developments within particular subgroups, such as Grimm's Law within Indo-European, may provide additional evidence for subgrouping.

The use of typological traits in genetic classification is more controversial, the generally accepted view being that such traits are in fact not indicative of genetic relationships. As is well known, the use of typological traits in some of the earlier work on African linguistics led to classifications that were definitely not phylogenetic, but the error of those early taxonomists lay more in their reliance on too few traits (sometimes just one) than in the traits themselves. Thus from a strictly historical perspective the use of sex gender, nasal vowels, or word order alone to classify languages would lead to absurd results. Still, such features, no less than grammar or the lexicon, are genetically transmitted as a part of language, and thus have some, if not absolute, evidentiary value. It can hardly be an accident of nature that the 700 or so Papuan languages are uniformly SOV in basic word order (with a few notable exceptions

under Austronesian influence). The fact that the Indo-European sex-gender system is not cognate with gender systems in other families does not imply that gender has not been a genetically transmitted trait during the history of Indo-European (and of course in other families where it has developed independently). In a preliminary taxonomic analysis of the world's linguistic phyla (Darlu, Ruhlen, and Cavalli-Sforza 1988), we found, using typological traits such as consonants (presence or absence of *p*, *m*, *s*, etc.), vowels (presence or absence of *i*, *e*, *a*, etc.), pronouns (presence or absence of a first-person dual inclusive pronoun), and word order (presence or absence of SOV word order in the basic declarative sentence), that often those phyla that linguists had previously connected on the basis of cognates also were immediately affiliated with each other on the basis of typological traits (e.g. Uralic and Altaic, Chukchi-Kamchatkan and Eskimo-Aleut, Na-Dene and Caucasian, Niger-Kordofanian and Nilo-Saharan, Indo-European and Afro-Asiatic). These preliminary data indicate that there is a greater genetic component in typology than has previously been assumed.

A second consequence of monogenesis is that it becomes possible, at least theoretically, to compare a phylogenetic tree of the human family based on linguistic traits with one based on biological traits. Many linguists still believe that there is little correlation between linguistic and biological traits. According to Campbell (1986: 488), "repetition of the obvious seems required: there is no deterministic connection between language and gene pools or culture." However, recent work by L. L. Cavalli-Sforza et al. (1988) shows that the correlations between biological and linguistic classifications are of a most intimate nature: "Linguistic families correspond to groups of populations with very few, easily understood overlaps, and their origin can be given a time frame. Linguistic superfamilies show remarkable correspondence . . . , indicating considerable parallelism between genetic and linguistic development."

As we saw in Chapter 1, many linguistic taxa correspond almost exactly to biological taxa, not only in the lower levels of classification (e.g. Eastern Austronesian, Altaic), but in the higher levels as well (e.g. Congo-Saharan, Austric, Nostratic/Eurasiatic, Amerind). This being the case, one is left to wonder whether the basic biological dichotomy between Sub-Saharan Africa and the rest of the world will be matched by a basic linguistic dichotomy between Congo-Saharan (for which Gregersen 1972 offered lexical and grammatical evidence) and non-Congo-Saharan. As will be shown in Chapter 14, some of the roots posited by Gregersen for Congo-Saharan are in fact even more widespread, occurring elsewhere in the world, and hence cannot be innovations within Congo-Saharan. Gregersen's etymologies thus involve a mixture of those restricted to Congo-Saharan and those shared by Congo-Saharan and other families. The interrogative etymology *min* 'what' discussed above

is interesting in this regard, for it seems to be found everywhere *but* in Congo-Saharan, and thus could be an innovation of the non-Congo-Saharan grouping, if this turns out to be a valid linguistic taxon.

A final consequence of monogenesis, should it be accepted, is that it will necessarily lead to a reappraisal of a good many family-internal explanations of various phenomena. One such example is Dixon's (1980) explanation of the origin of the Australian interrogative *miNHa* 'what':

Languages in North Queensland which do not have *miNHa* as an indefinite-interrogative form generally have a lexical item *miNHa* 'meat, (edible) animal.' It is likely that there has been semantic shift, with the generic noun *miNHa* 'animal' shifting to become an indefinite term 'something'; like other indefinites in most Australian languages this also had an interrogative sense 'what.' *miNHa* is now found with indefinite-interrogative sense over a large region centered in New South Wales (and in a few scattered languages outside this area), suggesting a pattern of areal diffusion." (p. 376)

Thus Dixon derives the interrogative *miNHa* 'what' from a phonologically identical root meaning 'meat' via an intermediate stage meaning 'something.' Even if there were not abundant evidence that the Australian interrogative *miNHa* is cognate with similar forms in many other phyla, as we saw above, I believe Dixon's family-internal explanation would still have to be rejected. The semantic shift MEAT > SOMETHING is unusual, to say the least, and quite probably unattested in the world's languages. Furthermore, the normal semantic evolution is from INTERROGATIVE to INDEFINITE, not vice versa. All in all, Dixon's family-internal explanation is most improbable, whereas the family-external explanation is simple and straightforward. Greenberg (1990, 1991) illustrates the value of a broad perspective for understanding certain phenomena within Indo-European, a theme that has been one of the major contributions of Nostratic research as well. Biological taxonomists have long understood the crucial importance of the broad perspective, and one may hope that linguistic taxonomists will soon gain a similar appreciation.

We are only beginning to understand the structure of the human population as it is reflected in biological and linguistic traits. For the highest levels of human classification, biological taxonomy seems to be in a more advanced state for the moment, but clearly both biology and linguistics have their own separate and important roles to play in unraveling the phylogeny of our species. Perhaps when both biological and linguistic taxonomy have been elaborated more confidently and in greater detail, the many parallels and similarities between the two fields will come to be viewed not as "curious" but as natural.

REFERENCES

- Benedict, Paul K. 1975. *Austro-Thai: Language and Culture*. New Haven, Conn.
- Bengtson, John D., and Merritt Ruhlen. 1992. "Global Etymologies," Chapter 14 of this volume.
- Blake, Barry J. 1988. "Redefining Pama-Nyungan: Towards the Prehistory of Australian Languages," *Yearbook of Australian Linguistics* 1.
- Campbell, Lyle. 1986. "Comment," on an article by Joseph H. Greenberg, Christy G. Turner, and Stephen L. Zegura, *Current Anthropology* 27: 488.
- Cavalli-Sforza, L. L., Alberto Piazza, Paolo Menozzi, and Joanna Mountain. 1988. "Reconstruction of Human Evolution: Bringing Together Genetic, Archeological and Linguistic Data," *Proceedings of the National Academy of Sciences* 85: 6002–6.
- Darlu, Pierre, Merritt Ruhlen, and L. L. Cavalli-Sforza. 1988. "A Taxonomic Analysis of Linguistic Families," ms.
- Darwin, Charles. 1871. *The Descent of Man*. London.
- Dixon, R. M. W. 1980. *The Languages of Australia*. Cambridge, Eng.
- Dolgopolsky, Aron B. 1964. "Gipoteza drevnejšego rodstva jazykovyx semei severnoj Eurasii s verovatnostnoj točki zrenija," *Voprosy Jazykoznanija* 2: 53–63. [English translation in Vitalij V. Shevoroshkin and Thomas L. Markey, eds., *Typology, Relationship and Time*, 1986. Ann Arbor, Mich., 27–50.]
- . 1984. "On Personal Pronouns in the Nostratic Languages," in Otto Gschwantler, Károly Rédei, and Hermann Reichert, eds., *Linguistica et Philologica*. Vienna, 65–112.
- Eldredge, Niles. 1985. *Time Frames*. New York.
- Greenberg, Joseph H. 1963. *The Languages of Africa*. Bloomington, Ind.
- . 1971. "The Indo-Pacific Hypothesis," in Thomas A. Sebeok, ed., *Current Trends in Linguistics*, Vol. 8. The Hague, 807–71.
- . 1987. *Language in the Americas*. Stanford, Calif.
- . 1990. "The Prehistory of the Indo-European Vowel System in Comparative and Typological Perspective," in Vitaly Shevoroshkin, ed., *Proto-Languages and Proto-Cultures*. Bochum, Germany, 77–136. [Russian translation in *Voprosy Jazykoznanija* (1989), No. 4: 5–31.]
- . 1991. "Some Problems of Indo-European in Historical Perspective," in Sydney M. Lamb and E. Douglas Mitchell, eds., *Sprung from Some Common Source: Investigations into the Prehistory of Languages*. Stanford, Calif., 125–40.

- . To appear. *The Eurasiatic Language Family: Indo-European and Its Closest Relatives*. Stanford, Calif.
- Gregersen, Edgar A. 1972. "Kongo-Saharan," *Journal of African Languages* 11: 69–89.
- Illich-Svitych, Vladislav. 1967. "Materily k sravnitel'nomu slovarju nostratičeskix jazykov," *Etimologija* 1965 (Moscow), 321–73. [English translation in Vitaly Shevoroshkin, ed., *Reconstructing Languages and Cultures*. Bochum, Germany, 125–76.]
- . 1971–84. *Opyt sravnenija nostratičeskix jazykov*, 3 vols. Moscow.
- Klimov, G. A. 1964. *Etimologičeskij slovar' kartvel'skix jazykov*. Moscow.
- Pokorny, Julius. 1959. *Indogermanisches Etymologisches Wörterbuch*. Bern.
- Ruhlen, Merritt. 1979. "On the Origin and Evolution of French Nasal Vowels," *Romance Philology* 32: 321–35.
- . 1987. *A Guide to the World's Languages*, Vol. 1: Classification. Stanford, Calif.
- . 1991. "Proisxoždenije jazyka: retrospektiva i perspektiva," *Voprosy Jazykoznanija* 1: 5–19. [Russian translation of this chapter.]
- Swadesh, Morris. 1954. "Perspectives and Problems of Amerindian Comparative Linguistics," *Word* 10: 306–32.
- Sweet, Henry. 1901. *The History of Language*. London.
- Trombetti, Alfredo. 1905. *L'unità d'origine del linguaggio*. Bologna.
- Whitney, William Dwight. 1867. *Language and the Study of Language*. New York.