Taxonomic controversies in the twentieth century

Merritt Ruhlen

At the beginning of the twentieth century there were a number of hotly debated taxonomic controversies in linguistics, with scholars sharply divided into two camps. One camp argued that linguistic taxonomy had already progressed as far as it possibly could, and that attempts to find relatives of Indo-European were doomed to failure for the simple reason that evidence of genetic affinity disappears after 6,000 years – the presumed age of Indo-European – and thus no evidence could possibly still persist from earlier times, even if such evidence had once existed. Similar proposals to find relatives of Basque, Algonquian, and hundreds of other languages and language families were rejected for the same reasons. Obviously, in this context, any attempt to argue for monogenesis of all extant languages was met with disbelief and hostility.

The second camp saw things quite differently. It argued that Indo-European's closest relatives were already quite obvious, as were the immediate relatives of Basque, Algonquian, and many other languages and language families considered "independent" by the first camp. And some members of the second camp – notably the Italian linguist Alfredo Trombetti – even dared to argue that existing evidence already quite strongly supported the idea of monogenesis.

What is remarkable is that these same controversies remained even more hotly debated at the end of the twentieth century and, though the participants in the debate had of course changed, the controversies themselves had often remained remarkably similar in form and content. In this paper I will examine several of these controversies and argue that it is the ideas of the second group, particularly those of the much-maligned Trombetti, that are winning the debate. The linguistic evidence today is much richer than it was in Trombetti's day, and it confirms in virtually every respect Trombetti's daring hypotheses of a century ago.

1. The Sapir-Michelson debate

I will begin with an obscure taxonomic dispute that occurred in the second decade of the twentieth century. In 1913 Edward Sapir announced a surprising discovery, namely, that two languages located side by side on the northern

California coast line – Wiyot and Yurok – were most closely related to the vast Algonquian family that extended from Montana to the eastern seaboard. In support of this relationship Sapir offered some 200 lexical and grammatical similarities. But the most potent piece of evidence was the virtual identity of the pronominal prefixes in the three groups.

Table 1. Pronominal prefixes in Wiyot, Yurok, and Algonquian.

	my	your	his	someone's
Proto-Algonquian	*ne-	*ke-	*we-	*me-
Wiyot	du(ʔ)-	khu(ʔ)-	u(ʔ)-	b-
Yurok	ʔne-	kʔe-	ʔwe-/ʔu-	me-

The task of judging this evidence fell to the leading Algonquianist of the day, Truman Michelson, who recognised that "the importance of this discovery, if valid, can hardly be overestimated" (Michelson 1914: 362). Michelson's verdict was, however, entirely negative, and he dismissed all of Sapir's evidence as "fancied lexicographical similarities", misanalysis of morphological elements, accidental resemblances, and features in Wiyot and Yurok that were thoroughly un-Algonquian. His conclusion left no doubt that he considered Sapir's discovery without merit: "Enough has been said to show the utter folly of haphazard comparisons unless we have a thorough knowledge of the morphological structure of the languages concerned" (Michelson 1914: 367). As a consequence of Michelson's opposition the Algic hypothesis (Algonquian + Wiyot + Yurok) became one of the unresolved taxonomic controversies of the twentieth century.

In the early 1950s, during his work on the classification of African languages, Joseph Greenberg examined a number of taxonomic controversies around the world, one of which was the Algic hypothesis. Upon examining the Sapir-Michelson debate and its attendant evidence, he concluded that this relationship was, in fact, "not very distant ... and, indeed,... evident on inspection" (Greenberg 1953: 283). The real mystery was why anyone thought there was a mystery. During the 1950s the relationship became accepted and the decisive proof was often attributed by American Indianists to Mary Haas's work. Haas herself made no such claim. In fact, she seconded Greenberg's opinion and merely gave additional evidence for the relationship, implying, correctly, that

it was Sapir who had proved the relationship beyond a reasonable doubt (see Ruhlen 1994a for a more detailed discussion).

There are, of course, just four possible explanations for linguistic similarities: common origin, borrowing, accident, and onomatopoeia. Which explanation best accounts for the similarity of the Algic pronominal prefixes? Clearly onomatopoeia can be ruled out since there is no intrinsic connection between any pronoun and any particular sound, despite ill-founded and unsubstantiated claims by Johanna Nichols (1992) and others. The possibility that these resemblances are accidental may also be easily ruled out. While the probability that such similarities could arise by accident is not zero, it is as close to zero as we need to get in historical linguistics. Borrowing too may be eliminated since the nearest Algonquian language (Blackfoot) lies over 600 miles to the east, on the other side of the Rocky Mountains, and there is no evidence that either Wiyot or Yurok was ever anywhere near an Algonquian language. A claim of borrowing, which has in fact never been made, would be little more than a deus ex machina. The only reasonable explanation for such pronominal similarities is common origin, as Sapir, but not Michelson, realised. In a letter to Sapir, Alfred Kroeber wrote "Michelson's review strikes me as puritanical. I have never had any doubt of the validity of your union of Wiyot and Yurok with Algonkin.... I hardly consider it worth while seriously to refute Michelson. His attitude speaks for itself as hypercritical and negative.... I regard the case in point so one-sided as to be already conclusively settled" (Golla 1984: 153). Less well known is that the Algic relationship was independently discovered by Trombetti, though later than Sapir.

In sum, the Algic "controversy" was little more than one scholar's inability to see the obvious. Scholars with a broad knowledge of languages and an understanding of taxonomy – Sapir, Trombetti, Kroeber, Greenberg – realised immediately that the Algic pronominal similarities could only reasonably be explained by common origin. Narrow specialists like Michelson assumed – or perhaps hoped – that there could be some other explanation. It is quite clear that had Michelson congratulated Sapir on his brilliant discovery, as Kroeber did, there never would have been any controversy at all. But Michelson's position of power as the leading Algonquianist of the day allowed him to initiate a pseudo-controversy that lasted over half a century. The lesson to be learned from the Sapir-Michelson debate has been aptly stated by Greenberg in a recent article: "I believe that one lesson of the Sapir-Michelson controversy is that 'controversial' is not to be equated with 'doubtful'" (Greenberg 1997: 669).

2. The Amerind controversy

The second taxonomic dispute I would like to examine is the Amerind controversy, initiated in 1987 by the publication of Greenberg's *Language in the Americas*, in which he argued that all New World languages belong to one of three groups, Eskimo-Aleut, Na-Dene, and Amerind. Of these, the first two had long been accepted as valid families; the controversy centred on Greenberg's claim that all other American Indian languages belonged to a single family, Amerind. This claim was at sharp variance with the prevailing opinion among American Indianists, who believed that what Greenberg called Amerind was really a group of over 200 unrelated families, or, to be more precise, 200 families among which there was no evidence of genetic affinity (Campbell 1997). Specialists concede that some of these families may be related, but the time depth is so great that any evidence of this fact would have long since disappeared.

Once again pronouns are among the most convincing pieces of evidence, for Greenberg provided abundant examples of a specifically Amerind pronominal pattern: *na* 'I' and *ma* 'you'. In fact precisely this pattern had been noted at the beginning of the past century by both Trombetti and Sapir. In a book published in 1905 Trombetti devoted an entire appendix to documenting this pattern in the Americas and concluded that, "from the most northern regions of the Americas the pronouns NI 'I' and M 'thou' reach all the way to the southern tip of the New World, to Tierra del Fuego. Although this sketch is far from complete, due to the insufficient materials at our disposal, it is certainly sufficient to give an idea of the broad distribution of these most ancient and essential elements" (208). In a personal letter written in 1918, Sapir wrote: "Getting down to brass tacks, how in the Hell are you going to explain general American *n-* 'I' except genetically? It's disturbing, I know, but (more) non-committal conservatism is only dodging, after all, isn't it? Great simplifications are in store for us" (quoted in Ruhlen 1994b: 87). Franz Boas, aware of the widespread *n/m* pattern, nonetheless did not explain it as due to common origin, but rather attributed it to "obscure psychological causes".

Surprisingly, even though the *n/m* pattern in the Americas was recognised long ago, Greenberg's critics, such as Lyle Campbell, have claimed that this pattern is not particularly frequent in the Americas, no more frequent, according to Campbell (1994), than the *m/t* pattern that we will see characterises language families of northern Eurasia. I have shown elsewhere (Ruhlen 1995a) that Campbell's claim is false, but his assertion pinpoints

one of the major weaknesses of contemporary linguistics: the lack of a linguistic database that would quickly disprove Campbell's assertion.

In addition to the distinctive Amerind pronominal pattern, there is a lexical item characterising the Amerind family that by itself virtually guarantees the validity of the family. Throughout North and South America there is a root *t-na* (- indicates an indeterminate vowel) with the general meaning of 'child, son, daughter'. A comparison of hundreds of such forms indicates that, in the original Amerind system, the first vowel in the root was correlated with the gender of the child. Thus, the root had three morphologically determined grades: masculine *t'ina* 'son, brother', feminine *t'una* 'daughter, sister', and indeterminate *t'ana* 'child, sibling' (Ruhlen 1994c).

No contemporary Amerind language preserves all three grades of this gender ablaut system in this particular root, but a number have retained two (e.g., Iranshe *atina* 'male relative', *atuna* 'female relative'; Tiquie *ten* 'son', *ton* 'daughter'). Elsewhere, however, a single language can preserve all three grades, for example in the Tucano numeral 'one': *nik-e* 'one (masculine)', *nik-o* 'one (feminine)', *nik-a* 'one (indeterminate)'. It is noteworthy that this numeral is the general Amerind numeral for 'one' (Ruhlen 1995b).

Languages retaining only one grade of the root are far more abundant. Examples of the masculine grade include Molala *pne: -t'in* 'my older brother', Yurok *tsin* 'young man', Mohawk *-ʔtsin* 'male, boy', Proto-Tzeltal-Tzotzil **ʔih-ts'in* 'younger brother', Cuicatec *ʔdiino* 'brother', and Yagua *deenu* 'male child'. Feminine examples are Central Sierra Miwok *tu:ne-* 'daughter', Salinan *a-t'on* 'younger sister', Tacana *-tóna* 'younger sister', and Piokobyé *a-ton-kä* 'younger sister'. Examples of the indeterminate grade are Nootka *t'an'a* 'child', Coahuilteco *t'an-pam* 'child', Proto-Uto-Aztecan **tana* 'son, daughter', Aymara *tayna* 'first-born child', and Urubu-Kaapor *taʔ+in* "child". There is no intrinsic connection between any vowel and any particular gender, and in fact in Afro-Asiatic *i* is feminine and *u* is masculine. The combination of this particular root (*t-na* 'child') with this particular gender ablaut system (*i/u/a* 'masculine/feminine/indeterminate') is a trait found only in Amerind languages. Its explanation can only be genetic.

Significantly, the Amerind gender ablaut system, in conjunction with this root, can be reconstructed on the basis of just North American languages, or just South American languages. One might infer from this fact that the Amerind population must have passed through North America rapidly enough that the entire ablaut system reached South America intact. Had South America been populated by people with languages that retained only

two or just one grade of the root, the entire system could never have been put back together. This concords well with the archaeological record in the Americas, which appears to indicate just such a rapid migration by the first Americans throughout North and South America around 10,500 years ago (Klein 1999).

3. The isolation of Indo-European

The third controversy that I would like to focus on is the supposed isolation of the Indo-European family. Throughout the twentieth century most historical linguists maintained that Indo-European had no known relatives. Nevertheless, the Indo-European family was the subject of two disputes in the twentieth century with regard to putative relatives. I will discuss them in reverse chronological order.

3.1. The discovery of Hittite

It is largely forgotten today that the discovery of Hittite led initially to a dispute over whether Hittite (and later the other Anatolian languages) was, or was not, a member of the Indo-European family. At first scholars such as Antoine Meillet argued that Hittite was too different from the other Indo-European languages to be considered a member of the family: "The interpretation [of Hittite texts] is still very hypothetical, and the assertion that Hittite is Indo-European would seem to be very risky; it has been disputed by most of those who have examined the documents" (Meillet [1914] 1965: 99). Yet only a few years later the presence of certain diagnostic Indo-European traits led Meillet and others to conclude that Hittite was in fact an Indo-European language after all: "The Indo-European character of Hittite immediately struck the first interpreters" (Meillet [1922] 1964: 38).

It is seldom realised today, when no one questions the Indo-European nature of the Anatolian languages, that Meillet's initial position was in fact correct. Hittite and the other Anatolian languages are not members of the Indo-European family as that family was understood at the time of its discovery. Part of the problem in this debate was that scholars focussed on two options: Hittite is Indo-European, or it is not Indo-European. What was overlooked was the third option: that Hittite is not Indo-European, but is related to Indo-European. It was Edgar Sturtevant who first pointed this out

in 1933 when he argued that Hittite was really a sister language to Indo-European, not a daughter language. According to Sturtevant, this higher-level family, which he called Indo-Hittite, consisted of two branches, Hittite (Anatolian) and Indo-European, as that term had been understood before the discovery of Hittite. As Sturtevant pointed out, there were numerous traits shared by Indo-European languages which were absent in Hittite, one of which was the Indo-Hittite two-gender system, which developed in the Indo-European branch into a three-gender system. Paradoxically, Sturtevant's Indo-Hittite hypothesis was generally opposed by Indo-Europeanists during the past century (Warren Cowgill was a major exception), even though – somewhat schizophrenically – they conceded its fundamental correctness: "A number of archaic features in morphology and phonology set Anatolian apart from the other branches, and indicate that it was the earliest to hive off. But Anatolian remains derivable from Proto-Indo-European; and periodic efforts to situate Anatolian as a sister language to Indo-European, with both deriving from a putative 'Indo-Hittite', have not found a following" (Watkins 1992: 209).

What happened in the twentieth century was that taxonomy and genetic relationship were confused. Because Hittite had been convincingly shown to be *related* to Indo-European, it must therefore be a member of the family and therefore the term Indo-Hittite is not needed. But what really happened was that the term Indo-European was redefined to include languages that did not fit the original definition. This then raised the question of what to call the non-Anatolian branch of the redefined Indo-European, whose name had been usurped for the higher-level taxonomic unit. One finds in the literature that Indo-European is then referred to as "Early Indo-European" and the non-Anatolian branch is referred to as "Late Indo-European". But of course these two terms are simply different names for Indo-Hittite and Indo-European. In biology, where names of taxa are not allowed to be changed, this terminological confusion would not occur.

3.2. Nostratic

The confusion of genetic relationship and taxonomy is even clearer in the case of Nostratic. The Nostratic hypothesis – that Indo-European is related to certain other families – was first advanced by Holger Pedersen in the first decade of the twentieth century. Indeed, the very definition of the family was based on relationship, not taxonomy, for Nostratic was defined as those

families that are related to Indo-European. It should be clear that a taxon cannot be defined by a relationship to some particular family. Of course one of the motivating factors for such a definition was that Indo-European was not supposed to be related to any other family, and the Nostratic hypothesis was an attempt to break this supposed barrier. One consequence of this definition was that Nostratic has included different families for different scholars. The Nostratic of Vladislav Illich-Svitych (1971–84), sometimes referred to as Classical Nostratic, included Indo-European (by definition), Afro-Asiatic, Kartvelian, Uralic, Altaic, and Dravidian. Aaron Dolgopolsky's (1984) version, developed at about the same time, also included Chukchi-Kamchatkan and Eskimo-Aleut, but omitted Dravidian. And both of these differ from the original conception of Pedersen, which contained Semitic, Finno-Ugric, Samoyed, Yukaghir, Altaic, and Eskimo-Aleut. Such differences are in part due to the improper definition of Nostratic. If, as I will argue below, all the world's language families are related, then Nostratic becomes identical with Proto-Sapiens.

In recent years the accuracy of the Nostratic hypothesis has once again become the focus of discussion in historical linguistics (Renfrew and Nettle 1999), and one often hears the question "Is the Nostratic hypothesis correct?" It is crucial, however, to distinguish two different questions in this regard. First, does the Nostratic evidence prove that Indo-European is related to other families? Second, is Nostratic – in any of its definitions – a valid taxon, that is, a set of language families more closely related to one another than to any other family? In my (and Greenberg's) opinion the answer to the first question is yes, but the answer to the second question is no. Certainly the evidence offered by the Nostraticists shows overwhelmingly that Indo-European does have relatives, as was clearly recognised by Pedersen and Trombetti early in the twentieth century. However, none of the definitions of Nostratic are valid taxa. For example, in Illich-Svitych's version Afro-Asiatic and Dravidian are included in the family, yet Chukchi-Kamchatkan and Eskimo-Aleut are not. But in fact these latter two families are clearly more closely related to Indo-European than is Afro-Asiatic or Dravidian (Greenberg 2000). Thus Illich-Svitych's version is not a valid taxon. In fact, it has recently been conceded by most Nostraticists that Afro-Asiatic is really a sister to Nostratic, not a daughter (Dolgopolsky, however, still includes Afro-Asiatic).

It is for these reasons that Greenberg's version of Indo-European's closest relatives differs from the various versions of Nostratic. Greenberg's version, which he calls Eurasiatic, includes Indo-European, Uralic-Yukaghir, Altaic,

Korean-Japanese-Ainu, Gilyak, Chukchi-Kamchatkan, and Eskimo-Aleut, as shown in Figure 1. Greenberg believes that these families do form a valid taxon. Several Nostraticists have recently come to conclusions similar to Greenberg's, and Allan Bomhard considers Eurasiatic one branch of a larger Nostratic family (Bomhard and Kerns 1994). The only differences in Sergei Starostin's view from that of Greenberg is that he would include Kartvelian in Eurasiatic, but not Ainu.

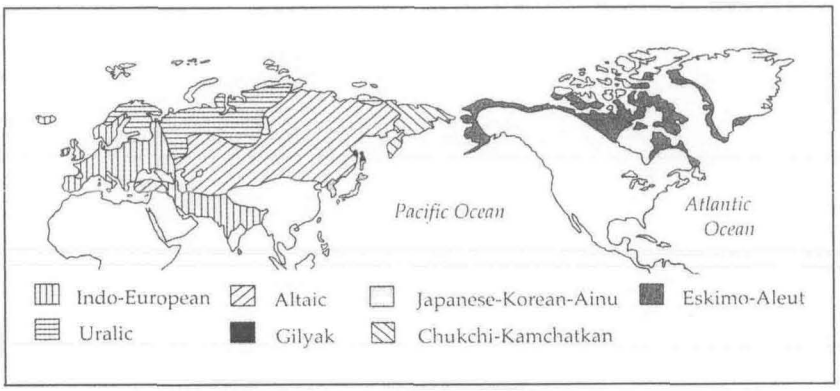

Figure 1. The Eurasiatic language family.

I will discuss Indo-European's relatives in terms of Greenberg's Eurasiatic family, though much of the evidence for Eurasiatic appeared first in the Nostratic literature (as Greenberg readily acknowledges). What is the evidence for Eurasiatic? One of the most salient pieces of evidence is the specific pronoun pattern *m/t* 'I/you', already clearly recognised by Pedersen, Trombetti, and others at the beginning of this century. Trombetti remarked somewhat acerbically in 1905 that "it is clear that in and of itself the comparison of Finno-Ugric [Uralic] *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" (Trombetti 1905: 44). Table 2 lists a few Eurasiatic cognates.

Table 2. Eurasiatic cognates.

	I, me, my	thou, thy	who?	what?	two, dual	plural
Indo-European	*mē-	*tu	*k^{w}i-	*ma		
Uralic-Yukaghir	*-m	*te	*ke	*mi	*-k	*-t
Turkic	men		*kim	*mi-	iki	-t
Mongolian	mini	*ti	ken	*ma	ikire	*-t
Tungus	mini	-ti	*xa-	*ma		-te
Korean	-ma		-ka	mai		
Japanese				mī		
Ainu				mak	-ki	-ti
Gilyak		ti	-ka		-gi	-t
Chukchi-Kamchatkan	-m	-t	k'e	mi-	-k	-ti
Eskimo-Aleut	-ma	-t	*kina	*mi	-k	-t

However, in addition to the evidence already advanced by the Nostraticists, Greenberg's book contains additional evidence that leaves no doubt whatsoever that Indo-European can hardly be considered an "isolate" in any sense. While the distinctive M/T pronominal pattern, by itself, constitutes compelling evidence, there is an additional complication in the first-person pronoun that is even more decisive. While first-person M is characteristic of every Indo-European language – and is abundantly attested in other branches of Eurasiatic – there is a peculiar subject form of this pronoun that has never been satisfactorily explained. Indo-European shows different roots for the subject and object forms of the first-person pronoun: English 'I, me', French *je*, *me*, Russian *ya*, *menya*, and so forth. The Proto-Indo-European reconstructions of these two forms are **eg(h)om* 'I' and **me* 'me'. This particular suppletive alternation has always been considered a diagnostic trait of the Indo-European family. After all, the possibility that two unrelated languages would independently invent the same suppletive alternation is unlikely in the extreme. Greenberg shows, however, that this suppletive alternation, far from being an innovation of Indo-European, is in fact a trait that Proto-Indo-European inherited from an earlier Proto-Eurasiatic language.

The clearest evidence comes from Chukchi-Kamchatkan, at the other end of Eurasia, which shows a pronominal paradigm for the first- and second-person pronouns that is virtually identical with that of Proto-Indo-European: *eɤəm* 'I', *ma* 'me'. Furthermore, Chukchi-Kamchatkan has extended the pattern to the second-person, *eɤət* 'thou', in which second-person *-t* replaces first-person *-m*. The extension of this pattern is also found in Uralic, for example, in the Hungarian object pronouns *en-gem*(*-et*) 'me' and *te-ged*(*-et*) 'thee'.

Greenberg's hypothesis regarding the origin of this suppletive alternation is that Proto-Eurasiatic **egom* 'I' was originally a periphrastic form of the first-person pronoun used for emphasis. Just as in French *c'est moi* 'it's me' can be used in place of *je* for emphasis, a typologically similar development occurred in Proto-Eurasiatic, and the morphological analysis that Greenberg gives for **egom*, **e-go-m* 'this-is-me', is identical to that of French. The Kamassian language in Uralic supports exactly this analysis: *i-gä-m* 'I am' (= 'this-is-me').

Clearly neither onomatopoeia, accident, nor borrowing can be taken seriously as an explanation of these facts. Only common origin provides a plausible explanation for such pronominal similarities. The dozens of other grammatical items, and hundreds of shared lexical items pointed out by the Nostraticists and Greenberg, only confirm what can be surmised on the basis of just two pronouns.

4. A family of isolates: Dene-Caucasian

The fourth controversy that I will discuss is the question of language isolates, of which Basque and Ket are two of the most famous. One of the more exciting developments in the past two decades has been the identification of a language family, now called Dene-Caucasian, that includes several of these supposed isolates. The six branches of this family – Basque, Caucasian, Burushaski, Ket (Yeniseian), Sino-Tibetan, and Na-Dene – are shown in Figure 2. The current conception of this family derives from the work of Starostin (1984), Sergei Nikolaev (1991), and John Bengtson (1991), though as usual there were precursors, and as usual one of the primary precursors was Trombetti, who devoted an entire monograph to the origins of the Basque language: "I connect Basque most closely with the Caucasian family. But this linguistic group is then most closely related to the Sino-Tibetan family" (Trombetti 1923: 6).

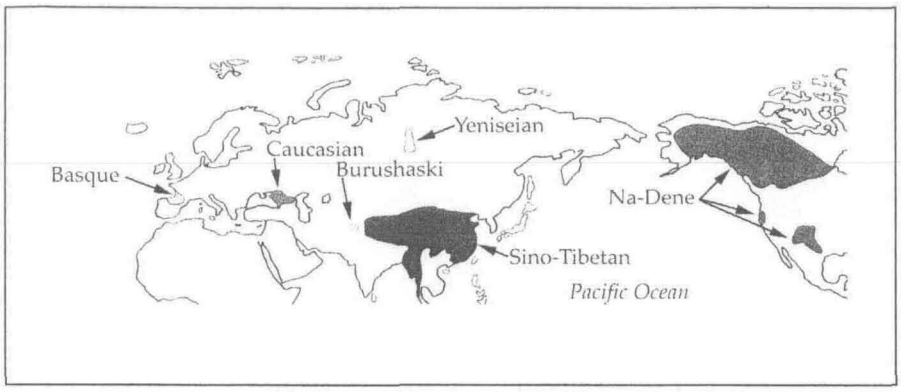

Figure 2. The Dene-Caucasian family.

Most of the recent work on Basque is the work of John Bengtson (1991) (the Nostraticists avoid languages that cannot be reconstructed). Table 3 shows six of the roots that characterise Dene-Caucasian. Note that Basque shares both the general Dene-Caucasian interrogatives, one in N and one in S. Note further that these interrogative pronouns are completely different from those of Eurasiatic, in which K, M, and Y form the basic interrogatives.

Table 3. Dene-Caucasian cognates.

Family	who?	what?	dry	day(light)	water	hungry
Basque	no-r	śe-r	agor	egun	ur	gose
Caucasian	*na	*sa	*-ɢ^{w}Vr-	*-ɢɨnV	*ħwiri	*gaši
Burushaski	ana	be-sʌ-n	qaqər	goon	hur-	
Sino-Tibetian	*naai	*su	*qaɾ̄	*k^{w}aaŋ		
Yeniseian	*ʔan-	*sV-	*qVr	*geʔn	*xur	
Na-Dene		sa	*-ɢaŋ	kuuŋ		*gas

5. Criticism

What are we to make of the criticism levelled at Greenberg and the Nostraticists in recent years and the similar criticism aimed at Trombetti and Sapir at the beginning of the twentieth century? Space permits only a brief discussion of some of the most common criticisms.

5.1. The temporal limit of the comparative method: 6,000 years

It has become an accepted dogma of contemporary historical linguistics that the Comparative Method is limited to roughly 6,000 years, before which time linguistic change has obliterated all evidence of genetic affinity. We have already seen many examples where this is clearly not true. The 6,000 year limit appears to be tied to the supposed age of the Indo-European family. In a sense it explains why Indo-European has no relatives, which, of course, we have already seen is incorrect. Furthermore, the Afro-Asiatic family must be over 10,000 years old since the language was pre-agricultural. There is really no empirical basis for the supposed 6,000 year limit on comparative linguistics; it is simply a self-imposed limitation of twentieth-century historical linguists.

5.2. Reconstruction and genetic affinity

Probably the most common criticism of Greenberg's work is that he has not reconstructed anything and therefore he has not followed the real comparative method. It has become a dogma of contemporary historical linguistics that only reconstruction can prove genetic affinity; this stricture appears in virtually every historical linguistics textbook. According to Hock, for example, only reconstruction proves genetic affinity, and Indo-European, Uralic, Dravidian, Austronesian, Bantu, and Uto-Aztecan have all been proved by successful reconstructions (Hock 1986: 567). And yet all of these families were universally accepted as valid families before anyone even thought of trying to reconstruct the protolanguage. If reconstruction proved Indo-European, as Hock claims, then *who* proved it, and *when*? When I posed this question to two historical linguists at a meeting at the University of Cambridge in 1998 – Don Ringe and Roger Lass – they replied that it was neogrammarians like Brugmann and Delbrück who had proved Indo-European. This would come as

quite a shock to these scholars, who never imagined that they were proving Indo-European. Indo-European had long been accepted as a valid family by everyone, and reconstruction was never cited as "proof" of anything. This idea is entirely an innovation of twentieth-century scholars; I have found no trace of this bizarre notion in the nineteenth century.

It is instructive to look at what the neogrammarians themselves had to say about the basis of genetic affinity, and it has nothing to do with reconstruction. Delbrück stated that "it was proved by Bopp and others that the so-called Indo-European languages were related. The proof was produced by juxtaposing words and forms of similar meaning. When one considers that in these languages the formation of the inflectional forms of the verb, noun, and pronoun agrees in essentials and likewise that an extraordinary number of inflected words agree in their lexical parts, the assumption of chance agreement must appear absurd" (Delbrück 1880: 121-122). Delbrück considered Indo-European to have been proved by the time of Bopp at the beginning of the nineteenth century, and the basis of this proof was the "juxtaposition of words and forms of similar meaning", a virtual paraphrase of Greenberg's methods.

If one were to tell a biologist that one should not believe in mammals because no one has ever reconstructed Proto-Mammal and then explained how this animal evolved through all the intermediate mammals before arriving at the current array of 4,006 species, the biologist would laugh and think it's a joke. Yet if you tell a traditional historical linguist the same thing regarding, say, Amerind, he will nod solemnly and say "of course".

The problem with modern historical linguistics is that the very nature of the comparative method has been misunderstood. Reconstruction has been confused with taxonomy. In reality the Comparative Method consists of two separate stages, Taxonomy and Historical Linguistics, and it is taxonomy that determines genetic affinity and necessarily precedes the normal concerns of historical linguistics.

It should be obvious that one cannot reconstruct a protolanguage until one has somehow identified a group of related languages, and it is the first stage of the comparative method, taxonomy, that identifies groups of related languages, just as Delbrück explained. One finds, however, that in contemporary historical linguistics the reconstruction of protolanguages – with of course regular sound correspondences – is called the Comparative Method. Taxonomy has disappeared from modern historical linguistics, and one searches in vain in any modern historical linguistics textbook for any discussion of classification.

5.3. Internal subgrouping

Another criticism that has been levelled at Greenberg's Amerind hypothesis is that he has not worked patiently backward through all the intermediate nodes, but has simply jumped to conclusions – Amerind – without even establishing the validity of the intermediate families. It should be obvious, however, that it is often far easier to discern more ancient groups than chronologically younger ones. For example, the Austronesian family, which extends from Madagascar to Easter Island and Hawaii and is thought to be about 6,000 years old, was recognised in the early eighteenth century, yet the internal subgrouping of the family remained largely unknown until recently.

Whether or not it is possible to identify intermediate nodes is a function of the rate of expansion and divergence of a population and has nothing to do with the validity of the overall family. If a population spreads rapidly through unoccupied territory – as the Amerind population seems to have done in North and South America roughly 11,000 years ago – intermediate groupings will be more difficult to detect than the overall family precisely because there was not sufficient time for the defining innovations to develop in the intermediate groupings. But the presence of the *n/m* pronoun pattern and the ubiquitous *tina/tana/tuna* example (and hundreds of other elements) throughout Amerind languages, but not elsewhere, establishes the validity of Amerind, whether or not intermediate subgroupings can be worked out. Even for Indo-European the family as a whole is quite obvious, whereas the internal subgrouping – how the dozen or so branches actually split up – is poorly understood. Except for Indo-Iranian and Balto-Slavic (and the fundamental divergence of Anatolian) little is known of this internal structure.

6. Monogenesis of extant languages

The final taxonomic controversy I will discuss is the question of the monogenesis of all extant languages. This topic is generally considered the most controversial of all, and most historical linguists regard even the possibility of ever proving monogenesis as intrinsically beyond the methods of comparative linguistics. There have, however, long been linguists who argued for monogenesis, and of course Trombetti is the best known. In fact, Trombetti's name is usually associated exclusively with the theory of monogenesis and, because of the stigma that has been attached to this idea, Trombetti's contributions to lower levels of taxonomy are generally unknown. Sapir, how-

ever, had a different opinion of Trombetti's work, as revealed in a letter to Kroeber in 1924: "There is much excellent material and good sense in Trombetti in spite of his being a frenzied monogenist. I am not so sure that his standpoint is less sound than the usual 'conservative' one" (Golla 1984: 420).

The question of monogenesis is an empirical one. Do ancient language families such as Niger-Kordofanian, Eurasiatic, Australian, and Amerind themselves share certain basic roots which would indicate a common origin? Figure 3 shows the world's major language families according to Greenberg. What would a comparison of these families show?

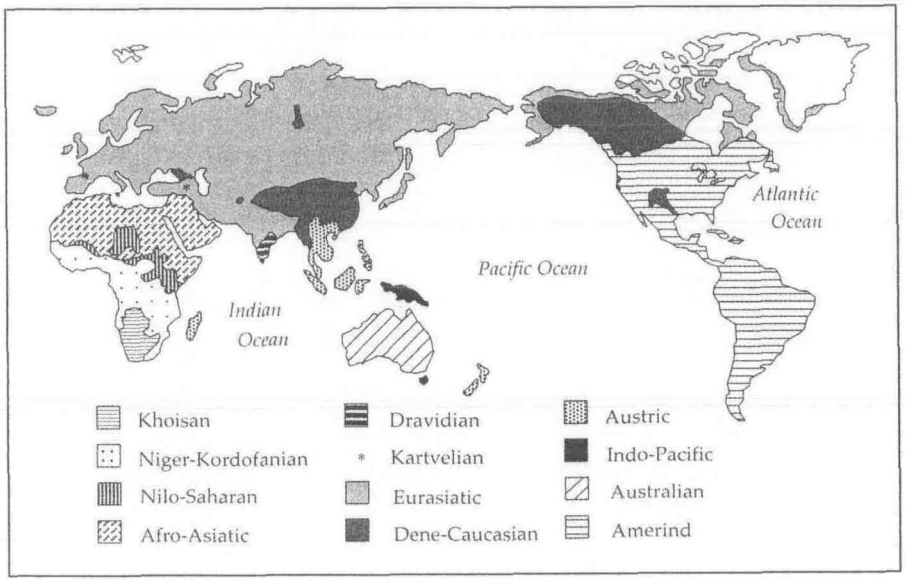

Figure 3. Language families of the world.

In the early 1990s John Bengtson and I (Bengtson and Ruhlen 1994) compared the roots that had been identified by specialists in 33 different families that included all the world's languages. We found that there were in fact a sizeable number of widespread roots and we argued, like Trombetti before us, that these similar roots could only be explained by common origin. One of the most widespread is *pal* 'two', examples of which are given in Table 4.

Table 4. pal 'two'.

Language-Family	Language	Form	Meaning
Niger-Congo	Nimbari	bala	'2'
Nilo-Saharan	Kunama	ba:re	'2'
Afro-Asiatic	Proto-Central Chadic	*-bwVr	'2'
Indo-European	Proto-Indo-European	*pol	'half'
Uralic	Proto-Uralic	*pälä	'half'
Dravidian	Proto-Dravidian	*pa:l	'part'
Austroasiatic	Proto-Austroasiatic	*ʔ(m)bar	'2'
Indo-Pacific	Kede	-po:l	'2'
Australian	Proto-Australian	*pula	'2'
Miao-Yao	Proto-Miao-Yao	*(a)war	'2'
Austronesian	Proto-Austronesian	*kə(m)-bal	'twin'
Amerind	Wintun	pa:le-t	'2'

How can such similarities be best explained? We have, of course, the same four possible explanations. Onomatopoeia can immediately be eliminated because there is no intrinsic connection between the sounds *pal* and the meaning 'two'. Borrowing can also be eliminated because intercontinental borrowing – from Africa to Australia to the Americas – could not have taken place until only a few centuries ago. We are left with a choice between common origin and accidental resemblance. But is it plausible that so many large families would have independently chosen the sounds *pal* to represent the number 'two'. Of Greenberg's 12 large families, *pal* occurs in all but Khoisan, Kartvelian, and Dene-Caucasian, and in the Amerind family *pal* is found in 11 of the 13 Amerind branches. Even for a language with just seven consonants and three vowels, there are 147 possible consonant-vowel-consonant (CVC) sequences, and *pal* is just one of them. For a language with 14 consonants and five vowels the possible CVC sequences increase to around 1,000. Clearly it is implausible that so many supposedly unrelated families should have independently chosen the same sequence of sounds to represent 'two'. The only reasonable explanation for these data is

common origin. As Sapir said: "It's disturbing, I know, but (more) non-committal conservatism is only dodging, after all, isn't it? Great simplifications are in store for us" (quoted in Ruhlen 1991: 222).

7. Conclusion

I would like to conclude by placing the linguistic evidence I have discussed here in the broader context of human evolution. Both archaeology (Klein 1999) and human genetics (Cavalli-Sforza, Menozzi, and Piazza 1994) indicate that all modern humans share a recent common origin in Africa. Anatomically modern people – who are identical with modern humans morphologically – first appear, in Africa, around 100,000 years ago. However, these people did not behave like modern humans; they behaved like Neanderthals, with both groups using essentially the same tool kit. Roughly 50,000 years ago there was a major (probably the major) transition in human evolution, as anatomically modern humans started, quite suddenly, to exhibit modern human behaviour. Artifacts that had changed little in hundreds of thousands of years, over vast geographical areas, suddenly became much more complex and began to change rapidly in both space and time. For the first time tools were made not just from stone, but also from bone, antler, tusks, shells, and other materials. It is at this time that the first clear indications of art appear, in both Australia and Europe. It was apparently also at this moment that these behaviourally modern people spread out of Africa, carrying with them the genes that attest to their recent African origin, and the language that has left traces, such as *pal* 'two', in contemporary languages down to the present day. Eventually these people spread throughout the world, replacing earlier hominids such as the Neanderthals and occupying for the first time Australia, Oceania, and the Americas. A number of scholars have maintained that this sudden and profound change in human behaviour could hardly have been accomplished without modern human language (Klein 1999). Indeed, the emergence of fully modern human language at this time is often seen as the underlying mechanism behind the swift change to behavioural modernity and the subsequent occupation of the entire world. If this is so, then the emergence of human language at this time is not only consistent with the Out-of-Africa hypothesis, it may help to explain it.