Genetic Linguistics

Essays on Theory and Method

Joseph H. Greenberg

Edited and Introduced by William Croft

Genetic Linguistics

This page intentionally left blank

Genetic Linguistics

Essays on Theory and Method

JOSEPH H. GREENBERG

Edited with an introduction and bibliography by
WILLIAM CROFT

OXFORD
UNIVERSITY PRESS

Contents

Preface

The idea for *Genetic Linguistics* arose in discussions between myself and Joseph H. Greenberg a few months before his death. I discussed the theme of the book with Greenberg, and we came up with an initial choice of essays. However, this process was not completed before Greenberg's death, and I take full responsibility for the final selection of essays and the organization of the volume.

Genetic Linguistics is not intended to be a complete collection of essays by Greenberg on the topic of genetic linguistics; this would be too large for a single volume (see the bibliography for a complete listing of Greenberg's essays on genetic linguistics). The choice of essays in *Genetic Linguistics* was governed chiefly by the following criteria: the essays should include discussion of theoretical and methodological issues in genetic linguistics; overlap in content between essays should be minimized; and essays published in the recent collection *On Language* (edited by Keith Denning and Suzanne Kemmer, 1990) would be excluded. In fact, there is no section devoted to essays on genetic linguistics in *On Language* (although some important short essays and reviews on African historical linguistics are included). *Genetic Linguistics* can be seen as a complement to *On Language*.

The final selection of essays largely conforms to these criteria. The most important exception is 'The Indo-Pacific hypothesis' (§12). This essay represents a proposed language family that Greenberg never followed up with further research. In order to make this essay more accessible to contemporary research, a comparison between Greenberg's classification and the Wurm classification, as modified by Ethnologue (13th edition), is included as an appendix to the essay. This comparison was prepared by Timothy Usher; I thank him for undertaking this task.

Undoubtedly the most important methodological essays on genetic linguistics by Greenberg are 'Historical linguistics and unwritten languages' (1953, reprinted in *Language, Culture and Communication*, out of print); 'Genetic relationship among languages' and 'The problem of linguistic sub-groupings' from *Essays in Linguistics* (1957; out of print); and 'The principles of genetic linguistic classification' from *Language in the Americas* (1987). In addition, the short essays on African and American Indian languages from the 1960s (§§4–5) cover important methodological issues.

The remaining essays in the collection are from the period following the publication of *Language in the Americas* in 1987. The publication of this book aroused a storm of controversy which has not abated. As a result, Greenberg wrote a number of essays defending his methods in general and the Amerind hypothesis in particular, as well as reviews of books by others on topics related to his own work. In many of these essays, Greenberg responded more directly to criticisms of his method, as well as discussing the methods of his critics. I have selected those essays from the later period which I believe have raised interesting points not presented in the earlier essays.

Of course, despite the best efforts at selection, there will be a certain degree of repetition across the essays in the collection. I have included all of the essays as written, except for minor stylistic and typographical corrections and the addition of cross-references to other essays in this collection [indicated in square brackets]. Removing the repetitive passages would destroy the unity of the essays involved. Also, the passages subject to repetition are significant in that they presumably represent the points that Greenberg considered most important (or most contentious), and in some cases, subtle shifts in explication indicate a certain evolution of Greenberg's ideas on the points made in the repeated passage.

The intensity of the controversy around Greenberg's methods and results calls for a certain amount of introduction to the issues. I have attempted to summarize Greenberg's method, the criticisms of the method, and Greenberg's responses to the criticisms (where available) in the introduction to this volume. I have also included a bibliography of literature on Greenberg's methodology and classifications. Interested readers are encouraged to examine the literature, in particular the specific articles, monographs, and reviews that Greenberg responds to in his essays.

I would like to thank a number of people and organizations for their support in this project, none of whom should be blamed for any shortcomings in the final result. Four anonymous reviewers made useful recommendations, including suggestions for the choice of essays to include; some of these suggestions were taken up, though I take responsibility for the final selection. Alan Kaye, John Rawlings, and Merritt Ruhlen assisted me in obtaining offprints of several of the essays included in this collection or considered for it. I would also like to thank the Max Planck Institute for Evolutionary Anthropology and the Center for Advanced Study in the Behavioral Sciences; much of this volume was prepared while a visiting Fellow to those two institutions. Chris Ehret, Paul Newman, Merritt Ruhlen, and Allan Taylor provided references for the bibliography. John Rawlings of Stanford University Library assisted me in the bibliographic search.

This page intentionally left blank

Editor's Introduction

Joseph H. Greenberg was the most prolific researcher to date on genetic classification of human languages, proposing large-scale classifications of languages in Africa, the Americas, Eurasia, and Oceania—in other words, the entire world—over a period spanning half a century. Greenberg began his classification research in the late 1940s, developing a classification of African languages over a period of some fifteen years. In the 1950s he began research in classifying languages in Oceania, the Americas, and Eurasia. However, publication of this work took considerably longer. The Indo-Pacific hypothesis, a proposed preliminary classification of the non-Austronesian languages of Papua New Guinea and neighboring islands, was published in 1971; a classification of the native American languages was published in 1987; and a classification of many of the languages of northern Eurasia was published in two volumes, one appearing in 2000 and the other posthumously in 2002. At the time of his death in 2001 he was planning to investigate a very large hypothesized family of languages in the southern hemisphere.

While the classification of African languages has received wide acceptance, Greenberg's hypotheses for other parts of the world have not. His Indo-Pacific hypothesis has been largely ignored. His classification of the languages of the Americas, in particular the Amerind hypothesis, which joins all of the languages of the Americas apart from the Na-Dene family (whose genetic unity is also questioned, at least for the affiliation of Haida) and Eskimo-Aleut, on the other hand, aroused a massive storm of controversy that has not abated. This controversy also surrounds his Eurasiatic hypothesis, which joins a range of language families in northern Eurasia from Indo-European in the west to Eskimo-Aleut in the east. Much of the controversy over Greenberg's classification rests on his methods as much as on his application of the methods to particular sets of languages.

The purpose of this collection is to bring together the papers by Greenberg describing his methods of language classification, and his principal responses to criticisms of the methods (Parts I–III). The collection concludes with four brief essays on the relationship between genetic linguistics and human prehistory in Part IV, a topic of particular interest to anthropologists and archaeologists. Most of these essays, particularly the original essays from the 1950s and 1960s setting out the methods, are out of print; and many of the more recent responses, postdating the publication of *Language in the*

Americas, are difficult to access (one, written in 1990, has not yet been published).

The controversy over Greenberg's genetic classifications raises important theoretical and methodological issues in genetic linguistics. Greenberg was one of the most influential linguists of the twentieth century in many areas of linguistics. Although his methods and thus his results in other parts of the world have been fundamentally challenged, his African classification has held up remarkably well. For this reason, it is useful to bring together the original presentation of the methods and their justification. The later essays discuss many fundamental questions in genetic linguistics raised by Greenberg's critics. This collection will also clarify some misconceptions about his methods found in the literature. In order to put Greenberg's work in its intellectual and historical context, the reader should be aware of the criticisms that have been directed towards Greenberg's methods (and therefore his results). In this introduction, I will briefly outline Greenberg's methods, the major critical responses to the methods, and Greenberg's reply to these responses (where available).

Greenberg's methods of language classification

Greenberg's methods are outlined in a series of publications in the 1950s and early 1960s, making up Part I of this collection (see also Greenberg 1948). Greenberg's presentation of his methods was determined in large part by the context of his research at the time, namely the classifications of African languages current in the 1940s. Thus, the major principles as he presented them at that time were formed in reaction to previous efforts by scholars to classify African languages. The major principles Greenberg proposed are:

(P1) *The universality of the mechanisms of linguistic change* (see §1 in this volume). On the basis of reconstructions of unwritten language families (e.g. Central Algonkian, Austronesian, and Bantu), Greenberg argues that language change is no different in unwritten languages than in written languages, and therefore the same methods of classification should apply. This view is now uncontroversial, and forms the underpinnings of theories of universals of language change, which were not prevalent in 1953.

(P2) *The exclusion of nonlinguistic evidence in language classification* (§§1, 5; Greenberg 1948: 27–8). Early twentieth-century classifications of African languages were influenced by theories of race and by a disproportionate significance attached to widely spoken languages

of 'civilized' peoples. Greenberg argued that all such considerations were irrelevant to linguistic classification. This view is also widely accepted among linguists today. Once a linguistic classification has been established, however, it may shed light on human prehistory (§§1, 6, 17–20; see also Greenberg 1957*a*).

(P3) *The exclusion of typological criteria in language classification* (§§1, 2, 4, 5, 8; Greenberg 1948: 24–6). In the early papers, Greenberg describes this as the exclusion of similarities based on form only or meaning only. Examples of form-only criteria are the existence of certain phonemes or phoneme classes, the use of prefixes, and word order patterns. Examples of meaning-only criteria are the existence of gender distinctions in nouns or person inflection in verbs. This view is also widely accepted among linguists today. The separation of typological criteria has stimulated two active areas of research. The study of the distribution, origin, and explanation of typological traits themselves is the topic of typology, a field which Greenberg can be said to have launched in the 1960s (Greenberg 1966/1990; Greenberg, Ferguson, and Moravcsik 1978). The study of the historical process of the areal diffusion of typological traits through language contact (§1; Greenberg 1957*a*, 1960/1990) is now actively engaged in by many linguists.

The strict application of these principles leaves form–meaning pairings as the sole evidence to be used to establish a genetic classification of languages. These are valuable criteria because they are arbitrary and largely independent of each other (§§ 1, 2, 6). With respect to form–meaning pairings, Greenberg enunciates one further principle which is widely accepted:

(P4) *The use of both grammatical and lexical form-meaning pairings*, that is, both vocabulary and grammatical inflections and derivations (implicit in §§1, 2, and explicit in §§4, 5, 6). Greenberg argues that in almost all cases, the classifications supported by lexical and grammatical form–meaning pairings coincide. This does not include abstract or schematic grammatical patterns, which are typological and can cross-cut genetic classifications (§1).

All of the aforementioned principles are now widely accepted, even by Greenberg's critics (see for example Campbell and Mithun 1979*a*: 18, 51). The next steps in Greenberg's method are the ones that generate controversy on theoretical grounds today. In the 1953 paper (§1), which gives Greenberg's approach in the most detail, its antecedent to this method is called 'collateral

comparison'. However, in an early version of the African classification (Greenberg 1954: 406–8; see §5), and parenthetically in the 1957 paper(§2), it is named 'mass comparison'. In 1987 Greenberg replaced the name with 'multilateral comparison' (§6). In the remainder of this section, multilateral comparison is briefly outlined, based on the essays in Part I of this book. The following section will discuss the major criticisms of multilateral comparison and Greenberg's responses based on the essays in Parts II-III of this book.

Multilateral comparison is essentially the inferral of a classification from similarities in grammatical and lexical form–meaning pairings across some or all of the languages under examination. 'Implausible' similarities are excluded (§1). Greenberg judges sounds to be similar based on known sound changes (§4). Plausible similarities in meaning are described as identity of translation into some other language (e.g. English) or single-step, widely attested semantic shifts such as 'sun' > 'day'; greater semantic latitude is discouraged (§2).

The next problem is to identify the source of similarities in form–meaning pairings in the languages involved. Greenberg describes four possible sources, two nonhistorical and two historical (he attributes them to Pott 1855, 1884; Greenberg 2001: 133). The two nonhistorical sources are accident (chance), which Greenberg calls 'convergence' (§1), and (sound) symbolism (§1). The two historical sources are borrowing and common origin (§1), the last being the goal of a genetic classification of languages.

Greenberg argues that sound symbolism is restricted to certain vocabulary items such as 'mother' and 'father'. Greenberg does not exclude these meanings from comparison, but states that any anomalous results are likely to be corrected by the evidence from other form–meaning pairings, and that the meanings likely to be sound-symbolic can be given less weight (§1).

In order to rule out chance comparison, Greenberg describes three methods (§§1, 2). The first two are quantitative. One is to calculate the likelihood of a combination of sounds to occur based on the phonemic structure of the language. Greenberg objects that this method does not exclude sound symbolism and does not account for the relative frequency of phonemes. He suggests that the latter factor can be corrected for, but that 'it would be difficult to carry out' (§1). The other quantitative method, which Greenberg considers 'more desirable', is to calculate the likelihood of matches between form–meaning pairings occurring randomly across the forms that actually arise in the languages being compared. Greenberg also states that 'this is a very tedious procedure' (§1). We will return to these methods at the end of this introduction.

The third method is to examine a sample of pairs of languages that are 'admittedly unrelated' and determine what percentage of form–meaning similarities exist. Greenberg did so and found a level of accidental form–meaning similarities of 4 per cent between languages with unlike phonemic systems and 7 per cent between languages with like phonemic systems (§1). He proposes that resemblances in a proportion above 20 per cent should be taken as not due to chance (§§1, 2), though he suggests that even 8 per cent is unlikely to be accidental (§1).

However, Greenberg then argues that although the sheer number of similarities are relevant, weighting of particular types of resemblances can lead to a relatively small number of such similarities being decisive (§1). The weighting factors Greenberg uses to increase (or decrease) the value of similarity for historical relationship—in particular, of common origin—are given in H1–H10.

- (H1) *Sound symbolism* (§1). Form–meaning similarities with meanings that are likely to be sound-symbolic are assigned less weight.
- (H2) *Length* (§§1, 11). The longer the form in a form–meaning similarity is, the more weight is assigned.
- (H3) *Allomorphic alternation* (§§1, 2, 6, 11, 13). If an allomorphic alternation is found in more than one language, especially an irregular or—even more significantly—a suppletive alternation, then a historical connection, indeed common origin, is extremely likely. Greenberg suggests that even one such irregular alternation will suffice to establish a historical connection (§1).
- (H4) *Rare morphological process* (§1). If a form–meaning similarity is found in a rare morphological process, a greater weight is assigned. For example, Greenberg considers the occurrence of similar grammatical infixes to be a significant criterion in his argument for the genetic unity of Austroasiatic.
- (H5) *Same morpheme combination* (§2). If one form–meaning pairing is found combined with another form–meaning pairing in two (or more) languages, greater weight is assigned to the form–meaning similarities. Greenberg gives the example of words resembling *to*, meaning 'ear', repeatedly combined with a class affix resembling *ku* in Niger-Congo languages.
- (H6) *Type of morpheme* (§§1, 2). Form–meaning similarity between grammatical morphemes is assigned greater weight than similarity between 'fundamental' vocabulary items, which in turn is assigned greater weight than similarity between 'cultural' vocabulary items.

Greenberg includes pronominal morphemes as grammatical morphemes (§1).

Greenberg notes that there is a trade-off between this criterion and the length and number of form–meaning similarities: grammatical morphemes have greater weight, but (core) vocabulary items are longer and more numerous (§§1, 2). Classification can therefore sometimes be successful even if one has only word lists (§2).

(H7) *Recurrent sound correspondence* (§§1, 6, 10, 11; Greenberg 1957b: 368–9). If there is a sound correspondence recurring across different form–meaning similarities, then these similarities are assigned greater weight.

(H8) *Number of languages sharing the form–meaning similarity* (§§1, 2, 6). The likelihood of form–meaning similarity being due to chance decreases exponentially with the increase in the number of languages sharing the form–meaning similarity. Greenberg suggests four languages sharing a similarity as a sufficient number to establish a historical relationship, specifically one of common origin if the morpheme is grammatical or (core) lexical (§1).

(H9) *Cumulative effect of including closely related languages* (§§1, 2, 6, 11). If a form–meaning similarity exists between language A and language B which also exists with other languages that have been established as closely related to B, then a greater weight is assigned to the similarity between A and B. The principle behind this factor is that form–meaning pairings shared among many languages in a group are more likely to be common retentions from the protolanguage and therefore more indicative of a historical relation to another language (or language group). For this reason, 'Languages should never be compared in isolation if closer relatives are at hand' (§1). Greenberg describes this process as 'collateral comparison' (§1).

(H10) *Convergence of reconstructed forms* (§1). This factor is closely related to H9. If one provides a tentative reconstruction of a form–meaning pairing for one set of historically related languages, and that reconstruction is more similar to the form in another language (or the tentatively reconstructed protoform of another group of languages), then a greater weight is assigned to the form–meaning similarities among the languages in question.

While factors H1–H7 operate in the comparison of a pair of languages as well as a larger number of languages, factors H8–H10 only operate in the

comparison of a larger number of languages, that is, multilateral comparison proper. Greenberg develops the argument in H9, found in §1, into a more general argument for multilateral comparison in §2. Comparisons between single languages without also examining their closest relations can be misleading (§2, discussion of data in Table 1). Simply stating that two languages, or a set of languages, are related is not useful from a cultural-historical perspective; a classification, however, would tell us much about human history and prehistory (§2; see below). The reasoning for multilateral comparison given in §2 is quoted here in full:

As a heuristic principle, the swiftest and surest method of bringing into play many of the considerations discussed here is the compilation of comparative fundamental ['core'] vocabularies of all the languages of an extended area. This accomplishes a number of purposes simultaneously. It involves the aspect of language least subject to borrowing outside grammatical elements. The forms are generally of fair length. Semantic straightforwardness is attained by using the translation equivalent of the same term in English or whatever language is used as the language of translation. The tendency of similar forms to appear in a number of languages, as well as the plausibility of descent from a common original, can easily be noted. The presence of recurrent phonetic correspondences can be seen without great difficulty. If, as is often the case, word lists or dictionaries include noun plurals or other morphological facts, even details of morphological combinations and alternations can be taken into account. Most important of all, perhaps, is that, where more than one family is represented, as is always the case when the languages examined are from an extensive area, the contrast between the relatively numerous and qualitatively superior resemblances among related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances.

For this reason in particular, Greenberg examines as many languages as possible in the area under study (§4, §8; he suggests examining at least a whole continent in Greenberg 1990c: 8). He notes that striking morphological irregularities are particularly good clues in examining an otherwise unclassified group of languages (§2; see H3 above). Finally, Greenberg notes that a random choice of a subset of languages in an area to compare is highly unlikely to lead to a valid genetic group, because the number of possible ways to classify a set of languages increases astronomically (§2; footnote 4 gives the mathematical formula).

At first, Greenberg did not apply the method of multilateral comparison on a very large scale. His original classification of African languages contained sixteen families, later reduced to twelve (Greenberg 1955). In his classification

of the languages of South America first presented in 1956 (published in 1960 as §4), Greenberg began with forty words in forty languages, chosen to include several members of groups that had already been established (following H9). He then increased the list to 300 words (where obtainable) and made tentative reconstructions based on recurrent correspondences. He then added further languages, assigning them to existing groups in accordance with correspondences to the tentative reconstructions. New groups were begun where necessary, and occasionally larger (deeper) families were identified (listed in the appendix to §4). This work encompassed around 250 languages. Likewise, in an unpublished report submitted in 1958, Greenberg proposed fourteen families of non-Austronesian languages in Papua New Guinea and neighboring islands (§12).

Not long afterward, Greenberg applied multilateral comparison to a much larger scale: 'one day, probably in early 1959, as I put my foot on the pavement to cross Amsterdam Avenue on my way to Columbia, an idea flashed before me. Why shouldn't I just look at all of my then twelve families in Africa together? Nothing changes methodologically just because the groupings are more distant from each other' (Greenberg 1996: 147). After that point, Greenberg began using large-scale multilateral comparisons. He eventually established four families for Africa (Greenberg 1963), examining grammatical elements and around 400 lexical entries (where available; the average was about 200 words) for about 800 African languages (Greenberg 1972/1990: 447–9). For Oceania, he examined some 350 lexical entries plus grammatical comparisons for about 800 non-Austronesian languages, with the same number of items for fifty neighboring Austronesian languages for comparison, and concluded that the fourteen families he had identified two years earlier formed a single large group, Indo-Pacific (excluding the Austronesian and Australian families; §12). In the Americas, Greenberg had already concluded by 1956 that there was one large Amerind family apart from Eskimo-Aleut and Na-Dene (as had Lamb 1959 and Swadesh 1960); but by the 1987 book, he had examined up to 400 lexical entries plus grammatical comparisons for around 1600 native American languages.

If accidental resemblances are ruled out by the factors described above, then one is left with two historical explanations, borrowing and common origin. One must therefore attempt to distinguish between these two possibilities. Greenberg argues that although individual cases can be doubtful, borrowing of a large mass of form–meaning pairings can always be identified (§§1, 2). He presents the following criteria for distinguishing borrowing from common ancestry.

(B1) *Borrowing scale* (§§1, 2). Borrowing of cultural vocabulary is more likely than borrowing of core vocabulary, which in turn is more likely than borrowing of grammatical forms (inflectional and derivational morphemes and pronominal elements).

(B2) *Degree of similarity* (§§1, 2). Borrowed forms will look too similar to the forms in the probable donor language, compared to formal similarities of the borrowing language to other languages indicating common origin.

(B3) *Single source* (§§1, 2, 6, 8; Greenberg 1957*a*: 69). Form–meaning similarities due to borrowing will be to a single language or at most two to three languages—the probable donor language(s)—and not to other languages that are closely related to the probable donor language(s) (§2; Greenberg 1960/1990: 423). Massive borrowing that is multilateral is improbable (§2).

(B4) *Semantic clustering* (§2; Greenberg 1957*a*: 69). Form–meaning similarities clustered in a particular semantic domain of cultural vocabulary are more likely to be due to borrowing.

(B5) *Special sound correspondences* (§§1, 6, 8; Greenberg 1960/1990: 421–2). The form–meaning similarities due to borrowing may have their own special sound correspondences not found in other form–meaning similarities that the borrowing language has with other languages (i.e. the borrowing language's true close relatives).

(B6) *Grammatical analyzability* (Greenberg 1960/1990: 422). A form that is grammatically analyzable in one language but not in the other language that shares the form–meaning similarity is likely to have been borrowed into the latter language from the former language.

Greenberg argues that factors (B1)–(B6), combined with (H7) and (H9) above, can be used to distinguish borrowings from cognates even in relatively closely related languages (Greenberg 1957*a*: 69–70).

The method described in Greenberg's earliest writings on genetic classification apparently remained unchanged in its basic form for the rest of his career, other than the expansion of its scope after 1959 described above.

The following essay in Part I (§3) describes problems of subgrouping. This essay is not summarized here as most of the points Greenberg makes are not controversial, since the genetic unity of the family to be subgrouped is assumed. Greenberg does raise two issues in §3 that are relevant to the classification question. Since occurrence in two or more subgroups is a widely used criterion for common ancestry (if borrowing is ruled out; §§3, 13 [citing Karl Brugmann]), such a form may either be indicative of the genetic unity of

the family as a whole, or evidence that the subgroups in which it is attested themselves form an intermediate level group below the level of the family as a whole. Greenberg proposes that the best solution is to examine more distantly related languages, where such exist: if the form in question is broadly attested in the still higher grouping, then it is not likely to represent an innovation defining an intermediate-level group (§3). He writes, 'The entire problem then remains, in a sense, one of subgrouping, but on a wider scale' (§3). Greenberg develops this argument in later writings in reply to his critics, as will be seen in the next section.

The second issue relevant to overall genetic classification in Greenberg's discussion of subgrouping is that the ease of identifying (sub)groups is dependent on the time between splits in the family tree: if the time from the first split to a second split in an ancestral group is short relative to the time between the second split and the present, then the subgroups defined by the first split will be very difficult to identify (§3). This observation is behind remarks Greenberg makes in later essays, for example that Amerind is easier to identify than its immediate subgroups (§11).

Criticism of multilateral comparison and Greenberg's response

Greenberg's most important discussion of the method of genetic classification in his later years, the first chapter of *Language in the Americas* (1987; §6) postdates the original explication of his method by thirty years. There is relatively little overlap in the content of this chapter with previous work, however. This is due both to the changes in historical linguistics over that time and the state of genetic classification of languages in the Americas in 1987.

When Greenberg began classifying the languages of Africa in the middle of the century, the existing classifications were heavily based on external (e.g. racial) and typological traits. Greenberg's methodological analysis was based on a rejection of those traits in favor of similarities in form–meaning pairings, as described in the preceding section. Although there were criticisms of Greenberg's method of mass comparison at the time (see Sapir 1987: 663; Newman 1995), most—though not all—Africanists agreed with Greenberg's rejection of nonlinguistic and typological features. Many prominent Africanists endorsed Greenberg's classification, even if some were skeptical of mass comparison (for the reception of Greenberg's African classification, see the bibliography at the end of this book).

The situation in the Americas was quite different. Although only lower-level classifications had been put forward in South America before

Greenberg's 1960 paper (§4), in North and Central America much genetic classification research had taken place in the early part of the century, leading to the relatively high-level classification made by Sapir (1929), and even to suggestions that most or all of the native American languages belonged to a single large family (Radin 1919 is one of the first; see also Greenberg 1990*a*; for the history of classification in the Americas, see Campbell and Mithun 1979*a*: 4–37; Greenberg 1987: 38–43; Ruhlen 1987: 205–27). However, by the time that Greenberg published *Language in the Americas*, there had been a significant shift away from higher-level genetic classification (see the aforementioned references). In fact, in the conference leading to the publication of Campbell and Mithun (1979*b*), most of the higher-level groups proposed by Sapir were rejected by the contributors to the volume. Instead, Campbell and Mithun advocate starting with lower-level groups established using the comparative method and then proceeding to higher-level groupings by further use of the comparative method (Campbell and Mithun 1979*a*: 37, 55; an early exponent of this approach is Haas 1958: 259). As a consequence, most Americanists criticized Greenberg's methods, and unlike the reception of his African classification, no Americanist who criticized Greenberg's methods endorsed his results (e.g. Rankin 1992: 325; a cautious exception is Munro 1994 on Greenberg's Yuki-Gulf sub-subgroup of Amerind). Very few Americanists endorsed Greenberg's classification of the languages of the Americas; those who did (e.g. Hymes 1987; Fox 1986; Lamb 1987) largely accepted his method (for the reception of Greenberg's classification of the languages of the Americas, see the bibliography at the end of this book).

Since Greenberg was aware that *Language in the Americas* went in the opposite direction of this change in Americanist genetic linguistic research, he devoted its first chapter (§6) to the notion of classification and the comparative method, which had received relatively brief discussion in the earlier essays. The remaining papers in Part II of this collection largely elaborate on certain topics covered in §6. In this section, I briefly summarize the methodological criticisms and Greenberg's responses.

The basic criticism of Greenberg's method of multilateral comparison is that (*a*) only application of the comparative method can 'prove' or 'establish' a genetic relationship and (*b*) the comparative method should be applied to lower-level groups and then applied to progressively higher-level (i.e. larger) groups. (These two criticisms are logically independent of one another: one could apply the comparative method at a large scale to begin with; or one could apply multilateral comparison to a smaller range of languages first, and then proceed to a larger range of languages.)

The comparative method is widely used and partly for that reason is given a variety of definitions in textbooks and scholarly works on the subject. Although there are a number of significant differences (see e.g. Ross and Durie 1996: 7; Lohr 1998: 7–16), certain basic features recur in published definitions. The first two quotations below describe the comparative method narrowly construed, while the last quotation represents a broad construal:

the comparative method . . . is a procedure for postulating reconstructed protoforms on the basis of the attested evidence of the descendant languages which have already been shown to be related. It is based on the principle that sets of recurring phoneme correspondences between two related languages continue blocks of positional allophones from the parent language.   (Baldi 1990: 3)

Systematic comparison yields sets of regularly corresponding forms from which an antecedent form can often be deduced and its place in the protolinguistic system determined. In practice this has nearly always involved beginning with cognate basic vocabulary, extraction of recurring sound correspondences, and reconstruction of a proto-phonological system and partial lexicon.   (Rankin 2003: 183)

The comparative method (in its strict sense) can be summarised as a set of instructions:

1. Determine on the strength of diagnostic evidence that a set of languages are genetically related, that is, that they constitute a 'family';
2. Collect putative cognate sets for the family (both morphological paradigms and lexical items);
3. Work out the sound correspondences from the cognate sets, putting 'irregular' cognate sets on one side;
4. Reconstruct the protolanguage of the family as follows:
   a. Reconstruct the protophonology from the sound correspondences worked out in (3), using conventional wisdom regarding the directions of sound changes.
   b. Reconstruct protomorphemes (both morphological paradigms and lexical items) from the cognate sets collected in (2), using the protophonology reconstructed in (4a).   (Ross and Durie 1996: 7; further steps involve subgrouping, producing the family tree, and constructing an etymological dictionary)

Essentially, the first two quotations describe steps 3–4 (and beyond) in Ross and Durie's formulation while Ross and Durie also describe their steps 1–2 as part of the comparative method, rather than being presupposed by it (see also Nichols 1996: 48; Rankin 2003: 209, n. 13). In fact, Baldi states as a precondition for the comparative method, 'A significant percentage of

cognates in core vocabulary areas must be demonstrated in order to establish genetic affinity between languages' (Baldi 1990: 2), and Rankin adds in a footnote to the quoted passage, 'Here I refer only to reconstruction. Grammatical correspondences have often been the feature that first established genetic relationship' (Rankin 2003: 208, n. 1).

Greenberg is criticized for not following steps 1–4, in particular steps 3–4, in the comparative method. His published evidence for the genetic classification of various language families does not include lists of sound correspondences and reconstructed forms (§§6, 13; Baldi 1990: 12; Ross and Durie 1996: 9). Yet the vast majority of historical linguists consider the successful achievement of steps 1–4 as necessary for the proof of the existence of a linguistic family:

Clearly, the most credible distant genetic proposals will be supported by both systematic sound correspondences and grammatical correspondences, anchored systematically in the grammars of the compared languages.   (Campbell and Mithun 1979*a*: 56; see also Campbell 1988: 599, n. 4)

Most historical linguists believe that the ultimate proof of genetic relationship lies in reconstruction.   (Hock and Joseph 1996: 466)

. . . a genetic linguistic relationship is first assumed, or hypothesized, by inspection or whatever. At that point must begin the careful and above all systematic comparison, which will lead, if the hypothesis or supposition of genetic relationship is correct, to the reconstruction of the linguistic history of the language concerned . . . It is the history which is, de facto, the proof of the genetic relation.   (Watkins 1990: 292)

The reconstruction of proto-languages on the basis of the Comparative Method does, of course, presuppose that the languages compared are related, and indeed, successful reconstruction provides a demonstration of such relationship.   (Fox 1995: 217)

In particular, it is argued that multilateral comparison, in particular lexical comparison, without following the comparative method, does not distinguish similarities due to common origin from similarities due to chance, sound symbolism, or borrowing (see above and §1); the more specific criticisms regarding the latter issues are described further below.

Greenberg's responses to these criticisms are found in the essays in Parts II-III of this collection (most of the issues were also touched upon in the 1957 essay (§2)). Greenberg argues that his method is not opposed to the comparative method as described in the quotations above: 'My criticisms are not directed at the comparative method as such. There is no other way of doing comparative linguistics' (§11; see also Greenberg's reply to Voegelin in Tax

1953: 60). Instead, he argues that multilateral comparison is the best method for achieving steps 1–2 in Ross and Durie's description quoted above, or that it forms the preconditions to applying the comparative method in Baldi's and Rankin's definitions quoted above, namely establishing genetic relations among languages and identifying a set of putative cognates (§§2, 6, 8, 11). That is, establishing genetic relationships among languages is a logically prior step to the comparative method in the narrow sense, namely identifying sound correspondences and reconstructing the protolanguage, i.e. steps 3–4 in Ross and Durie's description. (Greenberg, like Ross and Durie, includes steps 1–2 as part of the comparative method; see §§11, 14.) Greenberg argues elsewhere that steps 1–2 represent an earlier phase in the history of linguistics, in that several of the well-established language families were identified as genetic units before the development of the comparative method (§§6, 8, 11), and that multilateral (that is, not pairwise) comparison was employed (§4; Greenberg 1993: 84 and references cited therein). Greenberg also argues that in fact he seeks recurrent sound correspondences and performs tentative reconstructions where possible in the process of classifying languages using multilateral comparison, weighting such supporting evidence more heavily than simple similarity (§2 (see H7, H10 above), §8). He describes this in §8 as 'the protocomparative method'.

Greenberg offers several reasons why he believes that multilateral comparison is the best method for steps 1–2 in the comparative method. The first reasons were given above in the discussion of multilateral comparison from §2: the problem of identifying which languages or language groups to compare is combinatorially explosive, therefore some method is required to ensure that a valid taxon is identified (§§6, 7, 8). Greenberg draws on the terms from phylogenetic classification in biological systematics. The cladistic model of biological taxonomy is used to construct a phylogeny, that is, a classification that reflects the actual historical fissioning events for populations of organisms (or languages). A taxon is a group defined by common historical descent. Here Greenberg develops the comments on genetic grouping and subgrouping described at the end of the preceding section of this introduction into a general approach to genetic classification of languages. The essential problem is classification, that is, determining the relative relatedness of languages (or species in biology); simply saying two languages or language groups are related is not very informative or useful (§§6, 7, 8, 11). Greenberg argues here as in §2 that the best method for identifying a valid taxon is to look at as wide a range of languages in a geographical area as possible, and beyond it if necessary. (However, Rankin (1992: 326–8) argues that Greenberg's use of similarity of form–meaning pairings to identify

groups is more like biological phenetics than cladistics; Greenberg did not reply directly to Rankin, but see §3.)

Greenberg argues that the method of multilateral comparison provides as strong a 'proof' of genetic relationship as one can obtain in an empirical field (§8). That is, he disagrees with his critics that identifying sound correspondences and reconstruction of the protolanguage (steps 3–4 of the comparative method) are necessary to 'prove' a genetic classification; steps 1–2 may suffice, where enough probative data are present (§§6, 8). Greenberg argues that the verification of any hypothesis in an empirical science is a probabilistic one; what matters is the likelihood that similarities, and degrees of similarity, in form–meaning pairings among languages is not due to chance, sound symbolism, or borrowing. Greenberg argues that this probability can be established through multilateral comparison. We will return to this claim in more detail below.

Greenberg also argues that steps 3–4 in the comparative method are not completely deterministic, that is, establishing sound correspondences and reconstructions does not provide infallible proof of a genetic classification (§§6, 8, 9, 11). Greenberg distinguishes between the use of steps 3 and 4 for proving relationships. Step 3 is the requirement of regular sound correspondences. Greenberg argues that actual sound correspondences, though often recurrent, are rarely perfectly regular, and so strict adherence to step 3 would lead to skepticism and rejection of even widely accepted families (§§6, 13). Step 4 is reconstruction. Greenberg argues that even strict followers of the comparative method recognize diachronically valid reasons for irregularities in sound correspondences to arise. But given those reasons, a comparativist faced with an irregular sound correspondence has many different options in reconstructing the language that can be taken—and in well-established families have been taken—in order to explain the irregularity while preserving the principle of the regularity of sound change (§6):

(R1) Give additional conditioning factors for the sound change resulting in the correspondence.

(R2) Attribute the irregular correspondence to different protophonemes in the protolanguage.

(R3) Attribute the irregular correspondence to analogy (Greenberg describes various subtypes): a morphophonemic alternation in the protolanguage was analogized in different directions in the daughter languages.

(R4) Attribute the irregular correspondence to word boundary effects (word sandhi).

(R5) Hypothesize that the protolanguage had two variants, and different variants have been selected in the daughter languages, leading to the irregular sound correspondence. Greenberg elaborates on the phenomenon of 'protolinguistic variation' in §10.

(R6) Propose the occurrence of one or more of various sporadic processes (assimilations, dissimilations, metatheses, etc.), which were recognized by the Neogrammarians, that lead to irregular sound correspondences.

(R7) Reconstruct variant forms of the same word.

(R8) Attribute the irregular correspondence to dialect mixture in the protolanguage (see also §10).

(R9) Reject the etymology; that is, do not consider the words to be cognate.

(R10) Accept the irregular correspondence, leaving the possibility that it will be explained in the future.

Greenberg argues that all of these are legitimate approaches to an irregular sound correspondence in particular circumstances (§6). However, he argues from this that the availability of this range of approaches means that great latitude is allowed in reconstruction, leading to quite different results depending on the theoretical principles one adopts for the protolanguage (§§6, 11; Greenberg 1993: 81; §9 discusses this problem with respect to three proposed reconstructions for proto-Afroasiatic). Greenberg also argues that the comparative method itself involves not just regular sound correspondences but the appropriate application of R1–R10 (§§6, 9). Finally, Greenberg argues that sound correspondences may not differentiate borrowings from common origin, since borrowings also display recurrent sound correspondences (§6; see B4 above; see also Campbell and Mithun 1979*a*: 55).

Greenberg's critique of steps 3–4 in the comparative method for establishing a genetic classification is of course independent of his defense of multilateral comparison as steps 1–2 in the process. Much of the criticism of multilateral comparison rests on its value in achieving steps 1–2. Greenberg argues that the evidence for genetic classification is probabilistic; therefore, the similarities he observes must be demonstrated to be more likely to be due to common origin than chance, sound symbolism, or borrowing. Greenberg's critics argue that Greenberg has not satisfactorily ruled out chance, sound symbolism, and borrowing as possible reasons for the similarities in form–meaning pairs that he has identified. The arguments and Greenberg's responses (where available) are summarized below.

1. *Chance.* Much attention has been devoted to the question of whether the observed similarities are due to chance. The major criticisms can be

summarized under the following headings (references cited are selective; see also the bibliography at the end of this book):

(C1) *Number of languages.* The greater the number of languages examined, the more likely resemblances will be due to chance (Campbell 1988: 596, 603; Chafe 1987: 652–3).

(C2) *Loss of similar forms.* After a certain time depth, loss of vocabulary or grammatical morphemes will have proceeded at a such a rate that the remaining cognates will not be distinguishable from chance (Ross and Durie 1996: 9; Callaghan 1990: 16).

(C3) *Genuine cognates.* Genuine cognates may change beyond recognition without a careful reconstruction; without such reconstructions, we cannot identify the cognates that would support the classification (Ringe 1992: 16, 18, n. 25; Hock and Joseph 1996: 502, Thomason 1990).

(C4) *Multiple etyma with the same meaning* (Ringe 2002: 415–16). Greenberg postulates multiple grammatical etyma with the same meaning (e.g. three first person forms in Eurasiatic); having multiple etyma increases the likelihood that similarities are due to chance.

(C5) *Selection of language(s).* One could select another language or language group at random, and identify as many similarities as Greenberg does for his proposed families (Campbell 1988: 602–3, 606–8, using Finnish compared to Amerind grammatical and lexical elements).

(C6) *Errors in the data.* The data used by Greenberg contains so many errors of phonological form and morphological analysis, including errors due to the poor quality of early sources, that the remaning similarities will not exceed chance (Chafe 1987; Goddard 1987; Campbell 1988; Poser 1992; Berman 1992; Kimball 1992).

(C7) *Phonetic latitude.* Phonetic similarities rather than regular sound correspondences form the basis of comparison of form in Greenberg's method, and these mere similarities may be due to chance (Campbell 1988: 599; Bateman *et al.* 1990: 5). Also, others criticize Greenberg for latitude in phonetic similarities allowed (Chafe 1987: 652; Goddard 1987: 667; but contrast Goodman 1970: 120, n. 3; Gregersen 1977: 84).

(C8) *Morphological latitude.* Greenberg is also criticized for matching phonologically similar person markers that are prefixed in one language and suffixed in another (Rankin 1992: 339), and for matching forms by segmenting a noun prefix that was not synchronically justified (Poser 1992: 216–17).

(C9) *Semantic latitude.* The latitude allowed for similarity in meaning is so great that the resemblances must be due to chance (Campbell 1988: 600; but contrast Goodman 1970: 120, n. 3).

Greenberg offers responses to most of the criticisms in C1–C9 in the essays included in Parts II and III, often with reference to the criteria from his earlier work, listed in H1–H10 above.

(C1) Greenberg replies that while it is true that the likelihood of pairwise similarities increases with the number of languages, the probability of n-wise similarities raises the probability by the n-1 power (§13; see H8). In particular, adding a further language which is in fact related to a putative genetic group will convert at least some n-way matches to $n+1$-way matches, further strengthening the case for the group.

(C2) Greenberg argues that adding further languages to the comparison reduces the likelihood of loss of cognate forms (§6; see H8–H9). Also, it is more likely to be true that some vocabulary is lost at a lower average rate than other vocabulary, and under an inhomogeneous rate of loss, the likelihood of loss of cognate forms is considerably lowered (ibid.). In Appendix A of Greenberg (1987), included in this collection as part of §6, Greenberg provides formulas and sample calculations using the retention rate of 80 per cent / millennium (Swadesh 1955) and the inhomogeneous retention distribution suggested by Joos (1964).

(C3) Greenberg argues that phonologically dissimilar genuine cognates, while they exist, can only be identified after the genetic unity of a family is established on the basis of a large number of putative cognates that are phonologically similar (§11), and that examining closely related languages will allow one to identify reconstructible sounds that may be more similar (Greenberg 1993: 84–7).

(C4) This criticism was published after Greenberg's death; there is therefore no response to it by him.

(C5) Greenberg argues that selecting a language at random for comparison is not multilateral comparison (§6); that is, he agrees that selecting a language at random is an invalid approach to language classification. Multilateral comparison examines all of the languages in a broad area, including languages that are already established to be closely related (see H9).

(C6) Regarding errors in the data, Greenberg argues that in some cases, the error is sufficiently minor that the hypothesis of resemblance between forms is not affected (§14). Early sources with poor transcriptions may nevertheless be sufficiently precise for classification based on form–meaning similarities (§14). He concedes that there will be at least some errors,

and that the errors identified in articles criticizing the Amerind hypothesis may indeed be errors (§§13, 14). However, Greenberg argues that what matters is the number of the similarities that survive critical analysis: 'the number of errors would have to be vast indeed to refute a classification simply by pointing to its errors' (§14). He also argues that an error in one entry in an etymology does not in itself refute that etymology for the family as a whole (§14).

(C7) Greenberg's response to the criticism that he uses only phonetic similarity and not correspondences is described above in the more general discussion of the relationship of the steps of the comparative method to genetic classification: namely, that similarity, not sound correspondence, can be the basis for establishing a genetic classification. I have not been able to find any direct response in Greenberg's writings to the criticism of latitude in phonetic similarity. This criticism is the opposite of the criticism in C3; Greenberg has himself argued against phonetic latitude (Greenberg 1957*b*: 365, and his response to C3 summarized above).

(C8) Regarding the comparison of person prefixes in one language to suffixes in another, Greenberg argues that pronominal elements are often variable in position even in a single language, and that such variation is accepted in etymological dictionaries for widely accepted families (§14). Regarding synchronically unjustified morphological analyses, Greenberg replies that in historical studies, internal reconstruction of affixes such as a noun marker is justifiable and practiced in accepted language classifications (§14).

(C9) In his review of *Language in the Americas*, Campbell lists several semantic equations in Amerind etymologies which he argues represent too great a semantic latitude (Campbell 1988: 600). Greenberg argues that on the whole his Amerind etymologies show a narrow semantic range comparable to that found in any etymological dictionary (§§13, 14). Greenberg argues that the apparently greater semantic latitude found in the first semantic equation on Campbell's list, 'excrement/night/grass', is due to leaving out the central meaning of 'black, dark in color' and that the semantic extensions are specific to particular subgroups of Amerind (§13).

2. *Sound symbolism*. A general criticism made of Greenberg's classifications is that Greenberg does not address the issue of sound symbolism. Sound symbolism, e.g. in terms of infant sucking reflexes, is argued to be behind the widespread patterns of personal pronouns which Greenberg has argued is strong evidence for Amerind, i.e. *n* first person, *m* second person (Goddard and Campbell 1994: 198, citing Goddard 1986: 202, n. 5; see also Rankin 1992:

339). Sound symbolism has been proposed to lie behind the similarities in personal pronouns in Altaic, a subgroup that Greenberg accepts as part of Eurasiatic, i.e. the alternation between first person *bi* nominative/*min* oblique, and other pronominal forms (Doerfer 1985, cited in §15). Nichols argues that all patterns of similarity among personal pronouns, including those already mentioned and also Greenberg's proposed pronouns for Eurasiatic, i.e. *m* first person, *t* second person, are onomatopoeic in the same way as 'mama' and 'papa' terms (Nichols 1992: 261–2). Greenberg is also criticized for the inclusion of other possibly onomatopoeic forms in the comparisons (Goddard 1987: 657; Campbell 1988: 600–1).

Regarding the hypothesized sound symbolism of the personal pronouns of Amerind, Altaic, and Eurasiatic, Greenberg argues that there is no 'plausible support' for particular sounds to be sound symbolic for first and second person (§15), and that different sounds are systematically used for first and second person in different parts of the world (§§14, 17); he considers these patterns to be evidence supporting the relevant families. On the criticism of the inclusion of words that may be onomatopoeic in his etymological dictionary of Amerind, Greenberg argues that standard etymological dictionaries include items that are sound symbolic (§13; see also H1).

3. *Borrowing.* Finally, Greenberg is criticized for not taking into account borrowed terms among the similarities he has claimed (e.g. Campbell 1988: 599; but contrast Schachter 1971: 34–5). Greenberg is said to rely on lexical evidence exclusively (Haas in Tax 1953: 55; Campbell 1988: 596, 2003: 264; Baldi 1990: 12; Rankin 1992: 329, 1998: 26; Dixon 1997: 5; Golla 2000: 62); but lexical evidence is less reliable than grammatical evidence, chiefly because words are more likely to have been borrowed. It is also argued in some cases that large-scale similarities are due to massive borrowing (for the proposed Amerind personal pronouns *n* and *m*, see Bright 1984: 15, 25; for the Altaic pronouns and vocabulary in general, see Clauson 1969 and Doerfer 1985, cited in §15).

Greenberg responds that he has always included grammatical as well as lexical evidence (§§13, 14; Greenberg in Tax 1953: 60–1; see also §§1, 2, 4, 5, 6, 12, and Greenberg 1955, 1963, 1987, 2000; note that Greenberg includes pronouns as grammatical evidence). He also argues that a valid genetic grouping of languages—i.e. not a chance convergence or a contact relationship—will display both morphological and lexical similarities (§5; Greenberg in Tax 1953: 61). In response to the issue of mass borrowings, Greenberg repeats his argument from §1 that using the criteria summarized in B1–B6 above, a mass of borrowings can be differentiated from common origin (§§2, 6, 11, 13). (However, Goodman (1970: 121) suggests that in some cases, terms borrowed

in unrelated parent languages will have reflexes in multiple daughter languages of both parent languages, which would weaken the utility of Greenberg's criterion B3.) Greenberg also argues that the borrowing of first and second person pronouns in particular is an extremely rare event (§§11, 13, 15). Finally, Greenberg argues that borrowing can only be identified in combination with a proposed genetic classification (first enunciated in §1; see also Greenberg 1992: 148); and that a claim of borrowing must be supported by a plausible analysis of who borrowed what from whom (§2).

More generally, Greenberg argues that one must evaluate the hypothesis of common origin versus chance resemblance, sound symbolism, or borrowing probabilistically; absolute certainty in genetic classification cannot be achieved (§§8, 11, 14). Greenberg argues that one cannot disprove a hypothesis of common origin by identifying individual instances of accidental resemblance, sound symbolism, or borrowing between a pair of languages (§§11, 14). Instead one must examine the overall pattern of resemblances between languages and use it to justify their attribution to one or another of the four possible explanations for similarity (§11, 14). He states that 'an etymological dictionary is not meant as a "proof" of relationship' (§13; see also Greenberg 1993: 82–3). Instead he directs the reader to a comparative word list of European languages used to illustrate the method of multilateral comparison (§§2, 6); and to his notebooks of comparative word lists and grammatical morphology to justify his classification of the languages of the Americas (Greenberg 1987: ix; §14).

Conclusion

In this introduction I have attempted to summarize Greenberg's method of genetic linguistic classification, as explicated in his early papers, the major criticisms of his method, and his response to the criticisms. Of course, this brief introduction cannot do justice to the details of Greenberg's method or the criticisms that have been directed against it. I have not evaluated the arguments by Greenberg or his critics here; a proper discussion of the issues would require a monograph in itself. However, I will conclude with a brief comment regarding application of the quantitative methods that Greenberg discussed in his 1953 paper (§1).

The first method that Greenberg discussed is the likelihood of a meaning having a particular phonological form in a language. Some applications of this method have been made to genetic linguistic problems, but they suffer from a number of statistical defects (see Kessler 2001: 32–3 and references

cited therein). The second method suggested by Greenberg, and preferred by him at the time, is the likelihood of the actual matches between items compared across languages relative to random matches (i.e. without regard to meaning). This method has begun to be explored by statistically minded linguists. This method can be applied to the evaluation of form–meaning similarities; examples include Justeson and Stephens (1980), Baxter and Manaster Ramer (2000), and variants of the 'shift-test' (Oswalt 1970, 1998; Lohr 1998); it can even be applied to recurrent sound correspondences, though this is a much more complex procedure (Kessler 2001).

It may be that analyzing such problems mathematically will resolve some issues in the debate (see for example §§1, 2, 3, and Appendices A (§6. 1) and B to Greenberg 1987). Of course, the complications that arise in applying such statistical methods to linguistic questions are manifold. These complications are both mathematical/computational and linguistic, the latter being represented by precisely the issues raised by Greenberg and his critics (Kessler 2001 discusses some of these problems, but many others remain).

On the other hand, Greenberg argues that the strongest support for a genetic classification is its fruitfulness in generating further results—solving historical puzzles in existing language families, identifying further etymologies, and leading to comparative reconstructions (§§4, 11, 14)—and in casting light on cultural-historical questions, of the sort discussed in the essays in Part IV of this collection (§§17–20; see also Greenberg 1957a, 1972/1990).

References

Baldi, Philip. 1990. 'Introduction: the comparative method', in Philip Baldi; (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin and New York: Mouton de Gruyter, 1–13.

Bateman, Richard, Goddard, Ives, O'Grady, Richard, Funk, V. A., Mooi, Rich, Kress, W. John, and Cannell, Peter. 1990. 'Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language', *Current Anthropology* 31: 1–13.

Baxter, William H. and Manaster Ramer, Alexis. 2000. 'Beyond lumping and splitting: probabilistic issues in historical linguistics', in Colin Renfrew, April M. S. McMahon, and R. Lawrence Trask (eds.), *Time Depth in Historical Linguistics*. Cambridge: McDonald Institute for Archaeological Research, 167–88.

Bender, M. Lionel. 1996. 'Nilo-Saharan 95'. *Afrikanistische Arbeitspapiere* 45: 1–25.

Berman, Howard. 1992. 'A comment on the Yurok and Kalapuya data in Greenberg's *Language in the Americas*', *International Journal of American Linguistics* 58: 230–48.

Bright, William. 1984. *American Indian Linguistics and Literature*. Berlin: Mouton.

Callaghan, Catherine A. 1990. 'Comment on Bateman *et al.* 1990'. *Current Anthropology* 31: 15–16.

Campbell, Lyle. 1988. 'Review of *Language in the Americas*', *Language* 64: 591–615.

Campbell, Lyle and Mithun, Marianne. 1979*a*. 'Introduction', in Campbell and Mithun (eds.) 1979*b*: 3–69.

Campbell, Lyle and Mithun, Marianne (eds.). 1979*b*. *The Languages of Native America: Historical and Comparative Assessment*. Austin: University of Texas Press.

Chafe, Wallace. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 652–3.

Clauson, G. 1969. 'Leksikostaticheskaja otsenka altajsukoj teorii', *Voprosy Jazykoznanija* 5: 22–51.

Cole, Desmond T. 1960. 'African linguistic studies, 1943–1960', *African Studies* 19: 219–29.

Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Fox, Anthony. 1995. *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press.

Goddard, Ives. 1986. 'Sapir's comparative method', in V. Cowan, M. K. Foster, and K. Koerner (eds.), *New Perspectives in Language, Culture, and Personality*. Amsterdam: Benjamins, 191–209.

——1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 656–7.

Goddard, Ives and Campbell, Lyle. 1994. 'The history and classification of American Indian languages: what are the implications for the peopling of the Americas?' in Robson Bonnichsen and D. Gentry Steele (eds.), *Method and Theory for Investigating the Peopling of the Americas*. Corvallis, Oregon: Center for the Study of the First Americans, 189–207.

Golla, Victor. 2000. 'Language families of North America', in Colin Renfrew (ed.), *America Past, America Present: Genes and Languages in the Americas and Beyond*. Cambridge: McDonald Institute for Archaeological Research, 59–72.

Goodman, Morris. 1970. 'Some questions on the classification of African languages', *International Journal of American Linguistics* 36: 117–22.

Greenberg, Joseph H. 1948. 'The classification of African languages', *American Anthropologist* 50: 24–30.

——1954. 'Studies in African linguistic classification: VIII. Further remarks on method; revisions and corrections', *Southwestern Journal of Anthropology* 10: 405–15.

——1955. *Studies in African Linguistic Classification*. New Haven, Connecticut: Compass Press.

——1957*a*. 'Language, diffusion and migration', in Joseph H. Greenberg, *Essays in Linguistics*. Chicago: University of Chicago Press, 66–74. Reprinted in 1971,

Language, Culture and Communication: Essays by Joseph H. Greenberg, selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 93–105.

——1957*b*. 'Nilotic, "Nilo-Hamitic" and Hamito-Semitic: a reply', *Africa* 27: 364–78.

——1958. 'Report on the Classification of the non–Austronesian languages of the Pacific', MS.

——1960/1990. 'Linguistic evidence for the influence of the Kanuri on the Hausa', *Journal of African History* 1: 205–12. Reprinted in Greenberg 1990*b*, 419–27.

——1963. *The Languages of Africa. International Journal of American Linguistics* 29:1 (Part II). (Publication of the Indiana University Research Center in Anthropology, Folklore and Linguistics, 25.) Bloomington: Indiana University and the Hague: Mouton.

——1966/1990. 'Some universals of grammar with particular reference to the order of meaningful elements', in Joseph H. Greenberg (ed.), *Universals of Grammar*, 2nd edn. Cambridge, Mass: MIT Press, 73–113. Reprinted in Greenberg 1990*b*, 40–70.

——1972/1990. 'Linguistic evidence regarding Bantu origins', *Journal of African History* 13: 189–216. Reprinted in Greenberg 1990*b*, 446–75.

——1987. *Language in the Americas*. Stanford: Stanford University Press.

——1990*a*. 'Address to NACAL 18, Emory University, March 1990', *California Linguistic Newsletter* 20/3: 8–10.

——1990*b*. *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer. Stanford: Stanford University Press.

——1990*c*. 'The American Indian language controversy', *Review of Archaeology* 11/2: 5–14.

——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: Stanford University Press, 127–40.

——1992. 'Preliminaries to a systematic comparison between biological and linguistic evolution', in John A. Hawkins and Murray Gell-Mann (eds.), *The Evolution of Human Languages*. Redwood City, CA: Addison Wesley, 139–58.

——1993. 'Observations concerning Ringe's *Calculating the Factor of Chance in Language Comparison*', *Proceedings of the American Philosophical Society* 137: 79–90.

——1996. 'The genesis of multilateral comparison', *Mother Tongue* 2: 145–8.

——2000. *Indo-European and its Closest Relatives: The Eurasiatic Language Family, vol. 1: Grammar*. Stanford: Stanford University Press.

——2001. 'The methods and purposes of linguistic genetic classification', *Language and Linguistics* 2: 111–36.

——2002. *Indo-European and its Closest Relatives: The Eurasiatic Language Family, vol. 2: Dictionary*. Stanford: Stanford University Press.

Greenberg, Joseph H., Ferguson, Charles A., and Moravcsik, Edith A. (eds.). 1978. *Universals of Human Language* (4 vols.). Stanford: Stanford University Press.

Gregersen, Edgar A. 1977. *Language in Africa*. New York: Gordon and Breach.

Haas, Mary R. 1958. 'A new linguistic relationship in North America-Algonquian and the Gulf languages', *Southwestern Journal of Anthropology* 14: 231–64.

Hock, Hans Henrich and Joseph, Brain D. 1996. *Language History, Language Change and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.

Justeson, John S. and Stephens, Laurence D. 1980. 'Change cognation: a probabilistic model and decision procedure for historical inference', in Elizabeth Closs Traugott, Rebecca LaBrum, and Susan Shepherd (eds.), *Papers from the Fourth International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 37–46.

Kessler, Brent. 2001. *The Significance of Word Lists*. Stanford: Center for the Study of Language and Information.

Kimball, Geoffrey. 1992. 'A critique of Muskogean, "Gulf", and Yukian materials in *Language in the Americas*', *International Journal of American Linguistics* 58: 447–501.

Lamb, Sydney. 1959. 'Some proposals for linguistic taxonomy', *Anthropological Linguistics* 1/2: 33–49.

Lohr, Marisa. 1998. 'Methods for the genetic classification of languages', Ph.D. dissertation, University of Cambridge.

Newman, Paul. 1995. *On Being Right: Greenberg's African Linguistic Classification and the Methodological Principles which Underlie it*. Bloomington, Indiana: Institute for the Study of Nigerian Languages and Cultures, African Studies Program, Indiana University.

Nichols, Johanna. 1996. 'The comparative method as heuristic', in Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: Oxford University press, 39–89.

Oswalt, Robert L. 1970. 'The detection of remote linguistic relationships', *Computer Studies in the Humanities and Verbal Behavior* 3: 117–29.

—— 1998. 'A probabilistic evaluation of North Eurasiatic Nostratic', in Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the Evidence*. Amsterdam: John Benjamins, 199–216.

Poser, William J. 1992. 'The Salinan and Yurumanguí data in *Language in the Americas*', *International Journal of American Linguistics* 58: 202–29.

Pott, Augustus Friedrich. 1855. 'Max Müller und die Kennzeichen der Sprachwissenschaft', *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 9: 405–64.

—— 1884. 'Einleitung in die allgemeine Sprachwissenschaft', *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1: 1–68.

Radin, Paul. 1919. 'The genetic relationship of North American Indian languages', *University of California Publications in American Archeology and Ethnology* 14: 489–502.

Rankin, Robert L. 1992. 'Review of *Language in the Americas* by Joseph H. Greenberg', *International Journal of American Linguistics* 58: 324–51.

—— 1998. 'Grammatical evidence for genetic relationship and the Macro-Siouan problem', in Marie–Lucie Tarpent (ed.), *Papers from the 21st Annual Meeting of the Atlantic Provinces Linguistic Association/Actes du Colloque Annuel del'Association de Linguistique des Provinces Atlantiques (PAMAPLA).* Atlantic Provinces Linguistic Association, 20–44.

—— 2003. 'The comparative method', in Brian D. Joseph and Richard D. Janda (eds.), *Handbook of Historical Linguistics.* Oxford: Blackwell, 183–212.

Ringe, Donald A., Jr. 1992. *On Calculating the Factor of Chance in Language Comparison.* (Transactions of the American Philosophical Society, vol. 82, part 1.) Philadelphia: American Philosophical Society.

—— 2002. 'Review of Joseph H. Greenberg, *Indo–European and its Closest Relatives*', *Journal of Linguistics* 38: 415–20.

Ross, Malcolm and Durie, Mark. 1996. 'Introduction', in Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* Oxford: Oxford University Press, 3–38.

Ruhlen, Merritt. 1987. *A Guide to the World's Languages, vol. 1: Classification.* Stanford: Stanford University Press.

Sapir, Edward. 1929. 'Central and North American languages', *Encyclopedia Britannica* (14th edn.), 5: 138–41.

Sapir, J. David. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 663–4.

Schachter, Paul. 1971. 'The present state of African linguistics', in Thomas A. Sebeok (ed.), *Current Trends in Linguistics, vol. VII: Linguistics in sub-Saharan Africa.* The Hague: Mouton, 30–44.

Swadesh, Morris. 1955. 'Towards greater accuracy in lexicostatistical dating', *International Journal of American Linguistics* 21: 121–37.

Swadesh, Morris. 1960. 'On interhemisphere linguistic connections', in Stanley Diamond (ed.), *Culture in History: Essays in Honor of Paul Radin.* New York: Columbia University Press, 894–924.

Thomason, Sarah G. 1990. 'Hypothesis generation vs. hypothesis testing: a comparison of Greenberg's classifications in Africa and in the Americas'. Paper presented at the Conference on Language and Prehistory in the Americas, Boulder, Colorado, 1990.

Watkins, Calvert. 1990. 'Etymologies, equations and comparanda: types and values and criteria for judgement', in Philip Baldi (ed.), *Linguistic Change and Reconstruction Methodology.* Berlin: Mouton de Gruyter, 289–304.

Part I

Classification, grouping, and subgrouping

This page intentionally left blank

1

Historical linguistics and unwritten languages*

1.1 Historical linguistics and descriptive linguistics

Unlike some other aspects of anthropology affected by the functionalist attack on history, the validity and fruitfulness of the historic approach in linguistics has never been seriously questioned. The objections which have been raised to certain assumptions of classical Indo-European comparative linguistics, such as the existence of sound laws without exceptions or the overliteral interpretation of the family-tree metaphor of language relationship, have not involved any fundamental doubt as to the legitimacy and value of historical reconstruction as such; at the most, they have, in the case of the Italian group of neo-linguists,[1] suggested specific alternative reconstructions of certain Proto-Indo-European forms.

The possibility of the application of traditional Indo-European methods to 'primitive' (i.e. unwritten) languages has been deprecated by some Indo-Europeanists (Vendryes 1925). It is evident that, while in principle the same procedures are appropriate, the absence of direct documentation for earlier historic periods is a distinct methodological handicap. The last decades, however, have seen the successful employment of classical reconstruction methods in a number of areas, including Central Algonkian by L. Bloomfield, Bantu by C. Meinhof, and Malayo-Polynesian by O. Dempwolff. It should be borne in mind that in all these cases we have rather closely related forms of speech, so that the task involved is more comparable to the reconstruction of Proto-Germanic or Proto-Slavic than that of Proto-Indo-European. These

---

* *Anthropology Today*, ed. Alfred L. Kroeber, Chicago: University of Chicago Press, 1953, 265–86.

[1] The reconstructions of the neo-linguistic school are not generally accepted by other scholars. For an exposition of neo-linguistic method, see G. Bonfante (1945). For a hostile critique see Robert Hall, Jr. (1946). It should perhaps be added that the approach of L. Hjelmslev in Denmark seems to exclude diachronic problems in principle but that this remains hardly more than a theoretic model.

attempts do furnish an important demonstration of the universal scope of those mechanisms of linguistic change which were already known to function in the more restricted area of the traditionally studied Indo-European, Finno-Ugric, and Semitic stocks (Hockett 1948).

Much more serious than skepticism regarding the possibility of linguistic reconstruction in the absence of early written records is the widely held opinion, which will be discussed in a later section of this paper, that remote relationships or even those of the order existing within the Indo-European family cannot be established for primitive languages because of the far-reaching influence which one language can exercise on another even in fundamental traits of grammatical structure. It is even claimed that the genetic question here loses its meaning, in that one language can go back to several distinct origins and cannot therefore be said to belong to one family more than to another (Boas 1920). It is worth observing that even in these cases the value of historic investigation is not denied as providing evidence of specific contacts, even though, it is held, the genetic question cannot be resolved. Thus Uhlenbeck, who, in his later writing, takes the view of genetic connections just mentioned, has lavished much time and effort on an attempt to show resemblances between the Uralic languages and Eskimo which require a historical explanation, while avoiding commitment as to the nature of the historic relationship involved.

While historic linguistics thus continues as a legitimate and major area of linguistic endeavor, it is undeniable that, with the rise of structural schools in European and American linguistics, the center of interest has shifted in the recent period from the historical problems which dominated linguistic science in the nineteenth century to those of synchronic description. The present preoccupation with descriptive formulations, which appears to be the linguistic analogue of the rise of functionalism, can contribute much that is valuable to diachronic studies. Most obviously, perhaps, any advance in descriptive techniques, by improving the quality of the data which constitute the basis of historical investigation, can furnish material for hypotheses of wider historical connections and likewise increase the precision of reconstruction for those already established. Another factor of great significance is the influence of the fundamental approach to language which all structuralists share, whatever their other divergences, namely, the concept of languages as a system of functional units. In its diachronic aspect this provides us with a view of change as related to a system and at least partially explainable in terms of its internal functioning through time. In the realm of sound patterns, some of these implications have been realized for some time. Thus Trubetskoy, as well as others, has distinguished between those sound

changes which affect the sound structure of the language and those which leave it unchanged (Jakobson 1931). This clearly parallels the synchronic distinction between phonetic and phonemic sound differences. Under the influence of this manner of thinking, sound change in language is more and more considered in terms of the shifts and realignment it produces in the sound structure of language rather than as a haphazard set of isolated changes, as in the traditional handbooks of historical linguistics.[2] The more rigorous formulation of alternations in the phonemic shape of morphemes (morphophonemics) has also borne fruit in Hoenigswald's exposition of the bearing of such data on internal reconstruction, that is, the reconstruction of certain aspects of the former states of a given language without resort to either related languages or historical records (Hoenigswald 1950). Although historical linguists had in effect used this method without formulation, the emphasis on rigorous formulation of assumptions is, on the whole, beneficial in an area, such as historical reconstruction, in which it has so largely been lacking.

Although there is thus no fundamental opposition between the historical and descriptive approaches to language, the focusing of attention on synchronic problems in the recent historic period, combined with the traditional concentration of linguistic forces in the areas of a few major Eurasiatic speech families, has led to comparative neglect of the basic problems of historical research in unwritten languages.

1.2 The establishment of linguistic relationship

The fundamental achievement of nineteenth-century science in linguistics, as in certain other areas, notably biology, was to replace the traditional static interpretation of similarities in terms of fortuitous coincidence among species as kinds, all of which were created at the same time and could vary only within fixed and narrow limits, with a dynamic historic interpretation of similarities as reflecting specific historical interrelationships of varying degrees of remoteness. Taxonomy, the science of classification, thus was no longer the attempt to find essential features connecting certain things more closely than others as part of a divine plan but rather based itself on the selection of those criteria which reflected actual historic relationships. In the language of biology, it was the search for homologies rather than mere analogies. In spite of the fruitfulness of the Indo-European hypothesis and the further successes of similar hypotheses in establishing the Finno-Ugric,

---

[2] Examples are the recent studies of Grimm's laws and other changes in Germanic by Twaddell and others, and various studies by Martinet of sound shifts (e.g. 1950).

Semitic, and other families, the assumptions on the bases of which these first victories of linguistics as a science were obtained were never clearly formulated, and the extension of these methods to other areas of the world has suffered from the beginning from a lack of clarity regarding the criteria of genetic relationship, resulting, in almost every major area, in a welter of conflicting classifications and even in widespread doubt as to the feasibility of any interpretation of linguistic similarities in terms of historical connections. Yet assumptions which have been the very foundation on which the edifice of modern linguistics has been reared and which have helped give it a rigorousness of method and precision of result which are admittedly superior to those dealing with any other phase of human cultural behavior should not be lightly abandoned unless, of course, the data actually demand it. In what follows, an attempt is made to formulate the principles in accordance with which similarities in language can be given a historical interpretation. It is hoped that this will furnish the guiding principles on the basis of which problems in the subsequent sections referring to specific areas can receive a reasonable solution.

The fundamental assumption concerning language on the basis of which historical interpretation of linguistic similarities becomes possible seems to have been first explicitly formulated by the great Swiss linguist, Ferdinand de Saussure, in his *Cours de linguistique générale*, although its relevance for historical problems is not there stated. According to de Saussure, language is a system of signs having two aspects, the *significant* and the *signifié*, equivalent, in the terminology of Bloomfield and of American linguists, to 'form' and 'meaning', respectively. Moreover, the relationship between these two aspects of the linguistic sign is essentially arbitrary. Given any particular meaning, there is no inherent necessity for any particular set of sounds to designate it in preference to any other. Although first stated in this manner by de Saussure, this assumption actually underlies the nineteenth-century hypotheses of linguistic relationships and represents essentially the solution accepted by all modern linguists of the controversy descending from the Greeks concerning the naturalness versus the conventionality of language. Given the arbitrariness of the relationship between form and meaning, resemblances between two languages significantly greater than chance must receive a historical explanation, whether of common origin or of borrowing.

This statement regarding the arbitrariness of the sign does need some qualification, in that there is a slight tendency for certain sounds or sound combinations to be connected more frequently with certain meanings than might be expected on a purely chance basis. Conspicuous instances are the nursery words for 'mother' and 'father' and onomatopoeias for certain species

of animals. This is generally recognized as only a slight derogation from the principle of the arbitrariness of the sign, since the sound can never be predicted from the meaning; and, since such instances are relatively a minor factor from the point of view of frequency of occurrence, they will add slightly to the percentage of resemblances to be expected beyond those merely the result of chance between any two unrelated languages; but they are not adequate for the explanation of wholesale resemblances between two particular languages, such as French or Italian. Moreover, the few resemblances which rest on this factor can be allowed for by assigning them less weight in judging instances of possible historical connections between languages. This factor making for specific resemblances between languages will hereafter be called, somewhat inappropriately, 'symbolism', in accordance with the terminology employed by psychologists.

Given any specific resemblance both in form and in meaning between two languages, there are four possible classes of explanations. Of these four, two—chance and symbolism—do not involve historic relationship, in contrast to the remaining pair—genetic relationship and borrowing. These four sources of similarity have parallels in nonlinguistic aspects of culture. Genetic relationship corresponds to internal evolution, borrowing to diffusion, chance to convergence through limited possibilities (as in art designs), and symbolism to convergence through similarity of function.

Up to this point resemblances in form between two languages unaccompanied by similarity of meaning and those of meaning not bound to similarity of form have not been considered. I believe that such resemblances must be resolutely excluded as irrelevant for the determination of genetic relationship. They practically always arise through convergence or borrowing. Form without function (e.g. the mere presence of tonal systems or vowel harmony in two languages) or function without form (e.g. the presence of gender morphemes in two languages expressed by different formal means) is often employed as relevant for the determination of relationship, sometimes as the sole criterion, as in Meinhof's definition of Hamitic, or in conjunction with other criteria. The preference for agreements involving meaning without accompanying sound resemblances is sometimes based on metaphysical preconceptions regarding the superiority of form over matter (Kroeber 1913).

Resemblance in meaning only is frequently the result of convergence through limited possibilities. Important and universal aspects of human experience, such as the category of number or a system of classification based on sex or animacy in the noun or one of tense or aspect in the verb, tend to appear independently in the most remote areas of the world and can never be

employed as evidence for a historical connection. That the dual number occurs in Yana (California), ancient Greek, and Polynesian is obviously an instance of convergent development. Sometimes semantic similarity without similarity in the formal means of expression is present in contiguous languages of similar or diverse genetic connection. In these cases we have the linguistic analogue of Kroeber's concept of 'stimulus diffusion'—indeed, a remarkably clear-cut instance of this process. Languages spoken by people in constant culture contact forming a culture area tend to share many such semantic traits through the mechanism of diffusion. This process may be carried to the point where it is possible to translate almost literally from one language to another. However, since it is precisely the semantic aspect of language which tends to reflect changes in the cultural situation and since such semantic resemblances cover continuous geographical areas, these resemblances are clearly secondary, however far-reaching they may be in extent. Beyond the inherent probabilities, there is much empirical evidence in areas from which documented history exists. Those traits which various Balkan languages share in common and which are one of the marks of the Balkans as a cultural area are largely semantic, involving a difference in the phonemic content employed as the mode of expression. Thus Rumanian, Serbian, and Greek express the future by 'to wish' followed by an infinitive, but in Rumanian we have (first person sing.) *voiu* + V, in Serbian *ću* + V, and in Greek *tha* + V. These are all known to be historically relatively recent and not a result of the more remote Indo-European genetic connections which all of them share. Roughly similar arguments hold for resemblances of form without meaning. There are limited possibilities for phonemic systems. For example, such historically unconnected languages as Hausa in West Africa, classical Latin, and the Penutian Yokuts share a five-vowel system with two significant degrees of length (*a, a:, e, e:, i, i:, o, o:, u, u:*). Some resemblances in form without function are the result of the influence of one language on another, e.g. the clicks of Zulu which have been borrowed from the Khoisan languages. Normally, when related languages have been separated for a fairly long period, we expect, and find, considerable differences both in their sound systems and in their semantic aspects resulting from differential drift and the diversity of the cultural circumstances under which their speakers have lived. Too great similarities in such matters are suspect.

Since, as has been seen, resemblances in form without meaning and meaning without form are normally explainable by hypotheses other than genetic relationship, their presence does not indicate, nor their absence refute, it. Hence they may be left out of consideration as irrelevant for this particular problem.

The evidence relevant to the determination of genetic relationship then becomes the extent and nature of meaning–form resemblances in meaningful elements, normally the minimal element, the morpheme. Lexical resemblance between languages then refers to resemblances in root morphemes, and grammatical resemblances refer to derivational and inflectional morphemes. The two basic methodological problems become the exclusion of convergence and symbolism, on the basis of significantly more than chance resemblance leading to a hypothesis of some kind of historical connection, and among these the segregation of those cases in which borrowing is an adequate explanation of the more-than-chance resemblances from those instances in which this is inadequate and genetic relationship must be posited.

The first approach to the problem of more-than-chance resemblances is quantitative. We may ask how many resemblances may be expected between any two languages which are not genetically related and have not borrowed from each other or from a mutual source. Several approaches seem possible. One would involve the calculation for each of the two languages of the expected number of chance resemblances on the basis of its phonemic structure and allowed phonemic sequences arranged in terms of what may be called 'resemblance classes', based on a resolution as to what phonemes are to be considered similar to others for the purposes of the comparison. To such a procedure there are several objections. It does not eliminate the factor of symbolism, and it does not take into account the relative frequencies of the phonemes in each language. If, for example, in comparing two particular languages, it were agreed that the labials would all be treated as resembling one another and the dentals likewise and if, in both languages, dentals were five times as frequent as labials, the possibility of chance resemblance would be much greater than if they were equal. This objection could, of course, be met in principle by a weighting in terms of frequency, but in actual practice it would be difficult to carry out.

A more desirable procedure would be the following. Let us suppose that we have a list of one thousand morphemes matched for meaning in the two languages. In language A the first morpheme is *kan*, 'one'. Instead of calculating the abstract probability of a form resembling *kan* sufficiently to be considered similar, let us actually compare *kan* in form with all the thousand items on the other list. Let us likewise compare the meaning 'one' with all the meanings on the other list. The chance probability of the existence of a form resembling *kan*, 'one', in both form and meaning in list B will then be the product of form resemblances and meaning resemblances divided by 1,000, the total number of items. We should then do this for each morpheme in

list A and total the probabilities. As can be seen, this is a very tedious procedure. Moreover, it will not include resemblances due to symbolism.

A much more practical method, which takes into account both chance and symbolism, is simply to take a number of languages which are admittedly unrelated and ascertain the number of resemblances actually found. The difficulty here is that results will vary with the phonetic structure of the languages. A number of such counts indicates that approximately 4 per cent is the modal value, employing a very generous interpretation of what constitutes similarity. Where, however, the two languages are similar in the phonemic structure of their morphemes, the degree of resemblance can become significantly larger. For example, between Thai and Jur, a Nilotic language, which have very similar phonemic structures, it reaches 7 per cent. It can be safely asserted that a resemblance of 20 per cent in vocabulary always requires a historical explanation and that, unless similarity of phonetic structure leads to the expectation of a high degree of chance similarity, even 8 per cent is well beyond what can be expected without the intervention of historical factors. This factor of the similarity or difference of the phonemic structure of morphemes is so important that in doubtful cases a simplified version of the second test, that of matching lists, should probably be applied. We might compare a particular form in list B with all those in list A from the phonemic point of view only, allowing merely one meaning, that of its partner in list A, presumably the nearest semantic equivalent. We then compare with the expected frequency of resemblances (which is, of course, smaller than by the first method) only those cases of resemblances on the list in which the two forms are matched as nearest semantic equivalents. Thus, if as our first matching pair we had A *nem*, B *kan*, 'one', and later in the list A *ken*, B *sa*, 'only', the resemblance between A *ken*, 'only', and B *kan*, 'one', would be disregarded as not occurring in a matching pair.

In actual fact, however, this test can probably be dispensed with, since the mere quantity of resemblances in the form and meaning of morphemes is not the decisive factor in more doubtful cases. There are additional considerations based on the weightings to be accorded to individual items and the further fact that isolated languages are seldom found. The bringing-in of closely related languages on each side introduces new factors of the highest importance, which should lead to a definite decision.

Other things being equal, the evidential value of a resemblance in form and meaning between elements in two languages is proportional to the length of the item. A comparison such as A, -*k*; B, -*k*, 'in', is, from this point of view at least, less significant than such a resemblance as A, *pegadu*; B, *fikato*, 'nose'. More important is the following consideration. The unit of comparison is the

morpheme with its variant allomorphs, if these exist. If the two languages agree in these variations, and particularly if the variants are rather different in phonemic content, we have not only the probability that such-and-such a sequence of phonemes will occur in a particular meaning but the additional factor that it will be accompanied by certain variations in certain combinations. Agreement in such arbitrary morphophonemic variations, particularly if suppletive, i.e. involving no phonemic resemblance between the variants, is of a totally different order of probability than the agreement in a nonvarying morpheme or one in which the languages do not exhibit the same variation. Even one instance of this is hardly possible without historical connection of some kind, and, since, moreover, it is hardly likely to be borrowed, it virtually guarantees genetic relationship. We may illustrate from English and German. The morpheme with the main alternant *hæv*, 'have', in English resembles the German chief allomorph *ha:b*, 'have', both in form and in meaning. In English, *hæv* alternates with *hæ*-before-*z* of the third person singular present (*hæz*, 'has'). In German, correspondingly, *ha:b* has an alternant *ha-* in a similar environment, before -*t*, indicating third person singular present, to form *ha-t*, 'has'. Likewise, English *gud*, 'good', has the alternant *be-* before -*tər*, 'comparative' and -*st*, 'superlative'. Similarly, German *gu:t*, 'good', has the alternant *be-* before -*sər*, 'comparative', and -*st*, 'superlative'. The probability of all this being chance, particularly the latter, which is suppletive, is infinitesimal. Since it is precisely such arbitrary variations, 'irregularities' in nontechnical language, which are subject to analogical pressure, they tend to be erased in one or the other language, even if some instances existed in the parent-languages. Where they exist, however, they are precious indications of a real historical connection.

More generally applicable are considerations arising from the fact that the comparison is only in rare instances between two isolated languages. The problem as to whether the resemblances between two languages are merely the result of chance plus symbolism can then be tested by a number of additional methods. Let us say that, as is frequently the case, one or more other languages or language groups resemble the two languages in question but in the same indecisive way, that is, that this third or fourth language is not conspicuously closer to one than to the other of the two languages with which we have been first concerned. The following fundamental probability consideration applies. The likelihood of finding a resemblance both in form and in meaning simultaneously in three languages is the square of its probability in two languages. In general, the original probability must be raised to the $n - 1$ power where a total of n languages is involved, just as the probability of throwing a six once on a die is 1/6, but twice is $(1/6)^2$ or 1/36.

Similarly, if each of three languages shows a resemblance of 8 per cent to the other, which might in extreme cases be the result of mere chance, the expectation of the three languages all agreeing in some instance of resemblance in form and meaning will be $(8/100)^2$ or 64/10,000. In 1,000 comparisons, agreement among all three languages should occur only 6.4 times, that is, it will occur in 0.0064, or less than 1 per cent, of the comparisons. Hence a number of instances of such threefold agreements is highly significant. If four or more languages which are about equally distant from one another agree in a number of instances, a historical connection must be assumed, and if this agreement involves fundamental vocabulary or morphemes with a grammatical function, genetic explanation is the only tenable explanation.

This may be illustrated from the Afroasiatic (Hamito-Semitic) family of languages consisting of five languages or language groups—Egyptian, Berber, Semitic, Chad (Hausa and others), and Cushite. The forms involved are guaranteed as ancestral in each group by the requirement of earliest attestation, as in the requirement for Egyptian that it occur in the Pyramid Texts, our oldest document, or of appearance in at least two genetic subgroups (as in the case of Chad and Cushite), so that, in effect, we are comparing five languages. Allowing again the very high total of 8 per cent of chance resemblance between any two of the languages, the expected number of occurrences of morphemes similar in form and meaning in all five groups simultaneously becomes $(8/100)^4$ or 2,816/100,000,000. Assuming that about 1,000 forms are being compared from each language, this leads to the expectation of 2,816/100,000 of a morpheme. That is, if one compared a series of five unrelated languages at random, employing 1,000 words in each case, the operation would lead to a single successful case in approximately thirty-five such sets of comparisons. As a matter of fact, eleven morphemes are found in the case of Hamito-Semitic instead of the expected 1/35. There is only an infinitesimal probability that this could be the result of pure chance. In this case, the morphemes involved include such examples as *-t*, fem. sing., and *-ka*, second person singular masculine possessive. Genetic relationship, of which there are many other indications, seems the only possible explanation here.

Languages should never be compared in isolation if closer relatives are at hand. For the tendency of those particular forms in a language which resemble another language or group of languages to reappear with considerable frequency in more closely related forms of speech is a valuable index of the existence of a real historical connection. The statistical considerations involved may be illustrated once more from the Hamito-Semitic family. The question whether Hausa is indeed related to Egyptian, Semitic, Berber, and

the Chad language has always been treated through isolated comparisons between Hausa and the other groups, while the existence of more than seventy languages of the Chad group which show a close and obvious relation to Hausa has been ignored.

A comparison of basic vocabulary between Hausa and Bedauye, a contemporary language of the Cushite branch of Hamito-Semitic, shows 10 per cent agreement in vocabulary. It is clear that Hausa will have lost certain Proto-Hamito-Semitic words retained by Bedauye, and vice versa. The percentage of retained vocabulary is expressed by a simple mathematical relation, the square root of the proportion of resemblances. The proportion of Hausa vocabulary which is of Proto-Hamito-Semitic origin should therefore be $\sqrt{10/100}$ or approximately 32/100. If we now take another Chad language belonging to a different subgroup than Hausa, namely, Musgu, the percentage of resemblance to Hausa is 20 per cent. Applying the same reasoning, the percentage of Hausa vocabulary retained from the time of separation from Musgu, that is, from the Proto-Chad period, is $\sqrt{20/100}$, or approximately 45/100. If, then, we take forms found in Hausa which resemble Egyptian, Berber, Semitic, or Cushite and because of the existence of a true genetic relationship these forms actually derive from Proto-Hamito-Semitic, they must also be Proto-Chad. Since Hausa has lost its forms since the Proto-Chad period independently of Musgu, which belongs to another subbranch, a true Proto-Hamito-Semitic form in Hausa should reappear by chance in Musgu 32/100 ÷ 45/100 of the time, that is 32/45. On the other hand, if Hausa is not related to the other Hamito-Semitic languages, the apparent resemblances to them are accidental, and these words should reappear in Musgu no more frequently than any other, that is, 20 per cent of the time, 9/45 rather than 32/45. An actual count shows that, of thirty morphemes in Hausa which resemble those of branches other than Chad, twenty-two occur in Musgu. This is 22/30 or 33/45, remarkably close to the expected 32/45. On the other hand, of 116 forms which show no resemblances to those of other Hamito-Semitic branches, only fourteen occur in Musgu.

Beyond the frequency of resemblances and their distribution in other languages of the same group, the form which the resemblances take is likewise of importance. If the resemblances are actually the result of historical relationship, even cursory reconstruction should show greater resemblance in most cases between the reconstructed forms than between those of two isolated languages. If the resemblances are all convergences, on the whole, reconstruction should increase the difference of the forms. This can be done in a tentative manner as the comparison proceeds and without necessarily

involving the full apparatus of formal historical reconstruction, which is often not feasible with poor material or where the relationship is fairly remote and no written records are available. If, for example, we compared present-day Hindustani and English, we would be struck by a number of resemblances in basic vocabulary, including numerals, but the hypothesis of chance convergence would certainly appear as a plausible alternative. Even without going beyond contemporary Germanic languages, on the one hand, and Indo-Iranian languages, on the other, reconstruction would show a strong tendency to convergence of forms as we went backward in time, suggesting a real historical connection. Thus English *tuwÞ* resembles Hindustani *dā:t* only slightly. On the Germanic side comparison with High German *tsa:n* already suggests a nasal consonant corresponding to the nasalization of the Hindustani vowel. Conjecture of a possible **tanÞ* or the like as a source of the English and German form is confirmed by the Dutch *tand*. On the other hand, comparison of Hindustani with other Aryan languages of India suggests that the Hindustani nasalized and long vowel results from a former short vowel and nasal consonant, as in Kashmiri and Sindhi *dand*. Reconstruction has thus brought the forms closer together.

Last, and very important, a degree of consistency in the sound correspondences is a strong indication of historical connection. Thus, reverting to the English-Hindustani comparison, the presence of *t* in English *tuw*, 'two', *ten*, 'ten', and *tuwÞ*, 'tooth', corresponding to Hindustani *d* in *dō*, *das*, and *dā:t*, respectively, is a strong indication of real historical relationship.

Assuming that such a relationship has been established, there still remains the problem of whether the resemblances in question can be explained by borrowing. While in particular instances the question of borrowing may be doubtful, I believe it is always possible to tell whether or not a mass of resemblances between two languages is the result of borrowing. The most important consideration is the *a priori* expectation and historical documentation of the thesis that borrowing in culture words is far more frequent than in fundamental vocabulary and that derivational, inflectional, pronominal morphemes and alternating allomorphs are subject to borrowing least frequently of all.

The oft repeated maxim of the superiority of grammatical over vocabulary evidence for relationship owes what validity it has to this relative impermeability of derivational and inflectional morphemes to borrowing. On the other hand, such elements are shorter, hence more often subject to convergence, and usually few in number, so that in themselves they are sometimes insufficient to lead to a decision. Lexical items are, it is true, more subject to borrowing, but their greater phonemic body and number give them certain compensatory advantages. While it cannot be said, *a priori*, that any single

item might not on occasion be borrowed, fundamental vocabulary seems to be proof against mass borrowing. Swadesh, in a recent discussion of the problem of borrowing versus genetic explanations, presents quantitative evidence for the relative impermeability of fundamental vocabulary in several instances where the history of the language is known (Swadesh 1951).

The presence of fundamental vocabulary resemblances well beyond chance expectation, not accompanied by resemblances in cultural vocabulary, is thus a sure indication of genetic relationship. This is a frequent, indeed normal, situation where a relationship is of a fairly remote order. Pronoun, body parts, etc., will agree while terms like 'pot', 'ax', 'maize', will disagree. The assumption of borrowing here runs contrary to common sense and documented historic facts. A people so strongly influenced by another that they borrow terms like 'I', 'one', 'head', 'blood', will surely have borrowed cultural terms also. Where the mass of resemblances is the result of borrowing, a definite source will appear. The forms will be too similar in view of the historical remoteness of the assumed relationship. Moreover, if, as is usual, the donor language is not isolated, the fact that the resemblances all point to one particular language in the family, usually a geographically adjacent one, will also be diagnostic. Thus the Romance loan words in English are almost all close to French, in addition to hardly penetrating the basic vocabulary of English. If English were really a Romance language, it would show roughly equal similarities to all the Romance languages. The absence of sound correspondences is not a sufficient criterion, since, where loans are numerous, they often show such correspondence. However, the presence of a special set of correspondences will be an important aid in distinguishing loans in doubtful instances. Thus French loan words in English show regular correspondences, such as Fr. *š* = Eng. *č* or Fr. *ã* = Eng. *æn* (*šãs* : *čæns; šãt* : *čænt; šeːz* : *čejr*, etc.).

Genetic relationship among languages is, in logical terminology, transitive. By a 'transitive' relation is meant a relation such that, if it holds between A and B and between A and C, it must also hold between B and C. If our criteria are correct and languages do have single lines of origin, we should never be led by their application to a situation in which A appears to be related both to B and to C, but B and C themselves cannot be shown to be related. If this were so, A would consist equally of two diverse components, that is, would be a mixed language of elements of B and C. This situation is sometimes said to exist, and even on a mass scale. Africa is perhaps most frequently mentioned in this connection. Thus Boas (1929) writes: '. . . a large number of mixed languages occur in Africa. His [Lepsius's] conclusions are largely corroborated by more recent investigation of the Sudanese languages.'

Close investigation shows that, of the hundreds of languages in Africa (800 is the conventional estimate), there is only one language concerning which the problem of genetic affiliation could conceivably lead to two disparate classifications, the Mbugu language of Tanganyika. Even here the answer is clear that, in spite of the borrowing of Bantu prefixes and a large amount of vocabulary, mostly nonfundamental, the language belongs to the Cushite branch of Hamito-Semitic. The pronouns, verb forms, and almost all the fundamental vocabulary are Cushitic. The conventional African classification based on purely formal criteria, such as tone, combined with purely semantic ones, such as gender, had no connection with historical reality, and the necessarily contradictory results which followed led to the assumption of widespread mixture. If, as was done, we define a Sudanese language as monosyllabic, tonal, and genderless, and a Hamitic language as polysyllabic, toneless, and having sex gender, a polysyllabic, tonal language with sex gender (like Masai) will have to be interpreted as the result of a mixture of Sudanic and Hamitic elements.

The last full-scale treatment of this subject is Meillet's, which was followed by the counterarguments of Schuchardt, Boas, and others and a discussion of these objections by Meillet (1914). The present discussion is in fundamental agreement with Meillet in asserting that the genetic question always has a meaning and is susceptible of an unambiguous answer. Meillet differentiates between concrete grammatical resemblances involving both form and meaning and those involving meaning only without form, but only in passing. Similarly, he mentions rather casually the fact that fundamental vocabulary is not commonly borrowed, but does not exploit this insight. The advantages gained by collateral comparison with additional closely related languages, and the statistical significance of coincidences in three or more languages are not considered. The result is an unnecessarily skeptical attitude toward the possibilities of establishing genetic classification where there are no early written documents or where the grammatical apparatus is slight or non-existent (e.g. Southeast Asia).

The objections of Schuchardt and Boas are in large part taken into account in the present analysis by the distinction between resemblances based on form and meaning which result from contact with other linguistic systems and those involving form only or meaning only. It would perhaps be desirable to distinguish these by the terms 'borrowing' and 'influence', respectively. Justice is then done to Boas's insistence that diffusion is prominently operative in linguistic as in other cultural phenomena, by setting no limit to influence, which in the case of creole languages reaches its peak, while maintaining, in accordance with all the available evidence, that there are

definite bounds to borrowing, since it tends to cluster in nonfundamental vocabulary and makes only rare and sporadic inroads into basic vocabulary and inflectional and derivational morphemes. What is commonly said about the grammatical effects of one language on another refers almost entirely to influence, not borrowing, in the sense of the terms as employed here.

In other words, the effects of one language upon another are extremely widespread, fundamental, and important. What is maintained here is merely that the results are of a kind that can be distinguished from those caused by genetic relationship. Nor is it asserted that the genetic affiliation of a language is the sole important historic fact concerning it. The effects of borrowing and influence, being more recent chronologically and giving specific insights into the nature of the contacts involved, may frequently be of greater significance to the ethnologist and culture historian than the factor of more remote genetic affiliation.

These two types of historical connections between languages are carefully distinguished by Trubetskoy. A group of languages which have affected one another by influence and borrowing and form a group analogous to a culture area is termed a *Sprachbund*, while a group of genetically linked languages is termed *Sprachfamilie*. They become genera of the larger species, *Sprachgruppe*, taking in all types of historical connections between languages (Trubetskoy 1928).

The common habit of confusing these two situations by the use of the term 'mixed language', as though a language were a mechanical aggregate of a number of components which enter into it the same way but merely in different proportions that English is, say, 48 per cent Germanic, 43 per cent French, 4 per cent Arabic, and 0.03 per cent Aztec (because of 'tomato', 'metate', etc.) is a gross oversimplification and fails to distinguish the different origin and function of the Germanic as opposed to the Romance-Latin and other components in English.

From what has been said, it should be evident that the establishment of genetic relationships among languages is no mere *jeu d'esprit*. It is the indispensable preliminary to a determination of the causes of resemblances between languages by leaving borrowing as the only remaining source where more than chance resemblance does not lead to a hypothesis of relationship. Where such a relationship is present, it provides the basis for separation of autonomous from foreign elements through reconstruction of the ancestral language. Without such reconstruction, an understanding of the process of change in language undergoes a severe limitation to those few areas of the globe in which documented materials concerning the earlier forms of languages exist.

1.3 Selected regional sketches

1.3.1 *Africa*

The attempt to reduce the number of language families in Africa at all costs, leading to overambitious syntheses combined with a disregard of concrete resemblances in form and meaning between elements of language in favor of typological criteria, such as the presence of tone, noun classes, sex gender, monosyllabic roots, etc., has characterized African linguistic classification from the earliest systematic attempts (Lepsius, F. Müller, etc.) onward.

The dominant classification in England and the United States has been a kind of synthesis, varying in details with different writers, based chiefly on the investigations of Westermann on the Sudanic languages and Meinhof on the Hamitic languages. Clear statements of the basis of this classification can be found in Werner (1915) and in Tucker (1940), as well as elsewhere. According to this view, there are three great indigenous language families in Africa— Sudanic, Bantu, and Hamitic, with Semitic as a separate but late intrusion and Bushman as possibly related to Sudanic. A disputed point has been the status of Hottentot, which most assign to Hamitic with Meinhof but which some classify with Bushman to form a Khoisan family, while others leave it independent or at any rate unclassified. Each of the three main families has its basic characteristics. Thus Sudanic is monosyllabic, tonal, lacks stress, grammatical gender, and all inflection, and places the genitive before the possessed noun. Hamitic, at the opposite extreme, is defined as polysyllabic, possessing Ablaut variation, having grammatical gender and inflection, lacking tone, and placing the genitive after the noun. In addition, it possesses the characteristic of polarity, which can best be illustrated by an example. The Somali language uses the same formative for the singular of the masculine and the plural of the feminine, while another element marks simultaneously the singular of the feminine and the plural of the masculine. Meinhof often expressed the opinion that the Bantu languages, which are assigned characteristics almost midway between the Sudanic and Hamitic families, were the result of a mixture of the two or, as he once expressed it, 'had a Hamitic father and Sudanic mother' (Meinhof 1912).

It is admitted that few languages exhibit the traits of one of these families in full purity. Deviations from the ideal pattern are attributed to influences of one family on the other. It is held that such intimate fusions may result that the choice of the fundamental component can in certain cases be made only by an arbitrary decision. Such mixed groups of languages are the Semi-Bantu, formed from Sudanic and Bantu; Nilo-Hamitic, a fusion of Sudanic with

Hamitic; and, in the view of many, Hottentot, with a Sudanic-like Bushman element and a Hamitic element.

It is clear that by applying such criteria, which have no reference to the concrete relations between the form and the meaning of specific linguistic signs, Chinese is a Sudanic language and Old French is Hamitic. The latter, indeed, possesses a very striking bit of polarity in the use of *-s* to indicate the nominative singular and plural accusative of the noun as opposed to a zero suffix indicating the accusative singular and nominative plural (e.g. *murs* : *mur* = *mur* : *murs*). In addition, it possesses gender, Ablaut, and all the other stated characteristics of Hamitic speech. On the other hand, we are led to a crowning absurdity, in that forms of speech that are probably mutually intelligible can be classified as genetically distinct. Thus Meinhof, in classifying the languages of Kordofan, west of the Upper Nile, paid no attention to any other factor than the existence or absence of class prefixes in the noun. Three of these languages—Tegele, Tagoy, and Tumele—are similar, probably to the point of mutual intelligibility. Meinhof (1915–19) states: 'A comparison of vocabulary shows that the numerals [*sc.* of Tegele] completely agree with those of Tumele. Moreover they are for the most part identical with the Tagoy numerals. Besides, a number of word stems and some verb forms of Tegele are identical with Tagoy and Tumele. But the grammatical structure of the noun indicates that Tegele is a Sudanic language because noun classification is absent while Tagoy and Tumele have clear noun classes. Apparently there has been a mixture of two diverse elements.'

The other classification which has enjoyed currency is that of A. Drexel, adopted with a few modifications by Schmidt and by Kiekers in their respective volumes on the languages of the world. The Drexel classification embodies an attempt to demonstrate *Sprachenkreise* in Africa parallel to the *Kulturkreise* of the Graebner-Schmidt culture-historical school. This involves such violence to linguistic facts as the separation of the closely knit Mandingo group of languages into two unrelated families and the assumption of special Fulani-Malayo-Polynesian and Kanuri-Sumerian connections. There is no clear statement of the method employed in arriving at such conclusions.

The recent Greenberg (1949–50) classification concentrates on specific criteria which are relevant for actual historical relationship. The large heterogeneous Sudanic group, to which Westermann, in his more recent writings, denied genetic unity is split into a number of major and some minor stocks. The most important of those, Westermann's West Sudanic, shows a genetic relationship to Bantu, as evidenced by a mass of vocabulary resemblances,

agreement in noun-class affixes, and phonetic correspondences, including those relating to tone, to which Westermann himself had drawn attention and to which he had even attributed a genetic significance, without, however, modifying his general scheme of language families to take account of it. The Semi-Bantu languages show a special resemblance to the Bantu languages simply because they belong to the same subgroup of languages in the larger family, to which the name 'Niger-Congo' is applied. Since these Semi-Bantu languages do not possess common features as against Bantu, the Bantu languages must be classified as merely one of over twenty subgroups within that one of the fifteen branches of the vast Niger-Congo family which includes both Bantu and 'Semi-Bantu' languages.

Other major independent families formerly classified as Sudanic are Central Saharan, Central Sudanic, and Eastern Sudanic. This latter family includes the so-called 'Nilo-Hamitic' languages, along with the closely related Nilotic languages in a single subfamily.

Hottentot is treated along with the central Bushman languages as a single subgroup within the Khoisan languages, the other branches being Northern Bushman and Southern Bushman. The Khoisan languages, in turn, are related to Sandawe and Hatsa in East Africa to form a single Click family. Of Meinhof's various proposed extensions of Hamitic, Fulani is assigned to the westernmost subfamily of Niger-Congo; the 'Nilo-Hamitic' languages (Masai, Nandi, etc.) are classed as Eastern Sudanic; and Hottentot belongs to the Click family. Hausa, along with numerous other languages of the Chad family, is put, along with the traditionally Hamitic Berber, Cushite, and Ancient Egyptian and with Semitic, into the Hamito-Semitic family, for which the name 'Afroasiatic' is proposed, since there is no linguistic justification for granting Semitic a special status. The term 'Hamitic,' which has been the basis of much pseudo-historical and pseudo-physical reconstruction in Africa, is thus abandoned as not designating a valid linguistic entity. The Afroasiatic family thus consists of five coordinate branches: (1) Berber, (2) Egyptian, (3) Semitic, (4) Cushite, and (5) Chad.

The Greenberg classification assumes a total of sixteen independent families in Africa. There is some possibility of a reduction in this total. The hypotheses of a Kunama-Eastern Sudanic and a Songhai-Niger-Congo relationship, in particular, are worth investigating.

Westermann has indicated his adherence to this new classification in all essentials and is expected to espouse it in a forthcoming article in the journal *Africa*.[3]

---

[3] Personal communication.

1.3.2 *Oceania*

There is general agreement on the existence of only two extensive groups of related languages in Oceania—the Malayo-Polynesian and the Australian. The remaining families are the Tasmanian and a whole series of unrelated language families in New Guinea and neighboring islands, to which the cover-name 'Papuan' is applied, with the general understanding that there is no proof or even likelihood that these languages form a single stock. Regarding Malayo-Polynesian, there is general consensus concerning which languages are to be included in the family, and the historical work of reconstruction of the ancestral Malayo-Polynesian and other languages will be considered in the following section on 'Southeast Asia'.

For the other large group, the Australian languages, although the existence of widespread relationships within the continent is asserted by all investigators, there is lack of unanimity regarding the number of families, some maintaining the unity of Australian languages and others denying it.

The linguists of the period before W. Schmidt's important work were acquainted almost exclusively with the languages of the large group which covers all the south and much of the north of the continent and ignored or were unaware of certain languages of the extreme northwestern and north-central parts of Australia which differ considerably from the great mass of Australian languages. These observers, therefore, assumed the unity of all Australian languages and were concerned chiefly with hypotheses of outside connections, with Africa, with India (Dravidian), or, in the case of Trombetti, with an Australian-Papuan-Andamanese group. This latter attempt, like all the others, proved abortive in this instance, if for no other reason than that the Papuan member is no linguistic unit of any sort (Ray 1907).

It was Schmidt (1913, 1914, 1917, 1918) who laid the foundations of a more careful study of the problem in a series of articles in *Anthropos*, later republished as *Die Gliederung der australischen Sprachen* (1919). Schmidt distinguishes two main families of Australian languages: the southern, which covers approximately the southern two-thirds of the continent, and a northern. He explicitly denies the existence of a genetic relationship between these two groups. Unlike the southern family, which constitutes a true genetic unity, the northern, according to Schmidt, is not a family at all but consists of numerous diverse, unrelated forms of speech. In the light of clear statements to this effect, it is difficult to know what is meant in a historical sense by Schmidt's threefold division of these northern languages into those whose words end in consonants as well as vowels, those whose words end in vowels only, and those whose words end in vowels and liquids but not in other

consonants. This last group occupies, according to Schmidt, an intermediate position between the other two, probably through a process of language mixture. This threefold division of the northern languages, as well as the separation into a northern and a southern family, seems strongly motivated by an attempt at correlation with the *Kulturkreise* established in this area by the ethnological school of which Schmidt is a leading exponent. Kroeber (1924), in a review of Schmidt's work, criticized this division on the ground of obvious fundamental vocabulary resemblances between the northern and southern languages. He followed this up with a study of the distribution of common vocabulary items, which showed a sublime disregard in their distribution for the fundamental east-west dividing line which Schmidt had drawn across the Australian continent.

In a series of articles in *Oceania* (1939–40, 1941–3), Capell made substantial contributions to our knowledge of the languages of the northwestern and north-central parts of the continent and also revealed the surprising fact that many of these languages had noun-prefix classes resembling those of the Bantu languages in Africa in their general functioning but, one should hasten to add, without specific resemblances to them in form and meaning. Capell asserts the fundamental unity of all Australian languages. He divides them into suffixing languages, roughly equivalent to Schmidt's southern family, and prefixing languages, corresponding to Schmidt's northern division. The criterion employed is existence of verb suffixes or prefixes to form tenses and moods and to indicate pronominal reference. It is admitted that the northern languages are, to some extent, suffixing also. Within the northern group we have, again, a threefold division on principles different from those of Schmidt. Groups with multiple noun classes, two classes, and no classes are distinguished. Capell admits, in effect, that this is not a genetic analysis. It leads, as he himself points out, to an inevitable cul-de-sac similar to that of Meinhof in Africa, cited above. We are confronted with a pair of languages— Nungali and Djämindjung—which are almost identical except that Nungali has noun classes and Djämindjung has none. A similar pair is Maung and Iwaidja. Concerning these latter, Capell observes: 'It is safe to say, however, that had Iwaidja multiple classification, it would hardly be more than a dialect of Maung' (Capell 1939–40: 420).

The solution suggested here is a simple one, if one keeps in mind a primary canon of classification, one so obvious that it would hardly seem to need statement, yet is frequently disregarded in practice. Languages should be classified on linguistic evidence alone. Among the irrelevancies to be excluded is the extent of the area in which the language is found and the number of

speakers. There is no reason to expect that families of genetically equal rank should necessarily occupy territories approximately equal in extent. Germanic and Tokharian are coordinate branches of Indo-European, but a greater contrast in territory and population could hardly be imagined. Germanic covers substantial portions of four continents and numbers hundreds of millions of speakers. Tokharian has no speakers at all, since it is extinct.

The extent of fundamental vocabulary resemblance, including pronouns, among all languages in Australia and the specific similarities in the noun prefixes which connect many north Australian languages provide sufficient evidence of a single Australian family. This family has numerous subgroups, certainly at least forty, of which the large southern subgroup is just one which has spread over most of the continent (including the Murngin languages in northeast Arnhem Land and the languages of the western Torres Straits Islands). The ancestral Australian language had noun classes, and the southern subgroup has, like some of the northern languages (the prefixing, classless languages of Capell's classification), lost these classes. It still maintains a survival, however, in the distinction of a masculine and a feminine singular pronoun found in certain southern languages in which the afformatives employed resemble those of the masculine and feminine singular classes among the class languages.

1.3.3 *Southeast Asia*

There are sharp differences of opinion regarding linguistic relationships in this area. The following are the outstanding problems: (1) the validity of Schmidt's hypothesis of an Austroasiatic family consisting of Mon-Khmer, Munda, and other languages; (2) the validity of Schmidt's Austric hypothesis connecting Austroasiatic in turn with Malayo-Polynesian; (3) the affiliations of Thai and Annamite, connected by some with Chinese in one subbranch of the Sino-Tibetan family, while others place Thai with Kadai and Indonesian (Benedict) and Annamite with Austroasiatic (Schmidt and others); (4) the linguistic position of the Man (Miao-Yao) and Min-Hsia dialects spoken by aboriginal populations in China.

Accepting certain earlier suggestions and adding some of his own, Schmidt (1906) has proposed that the following groups of languages are related to one another in his Austroasiatic stock: (1) Mon-Khmer, (2) the Palaung-Wa languages of the middle Salween, (3) Semang-Sakai, (4) Khasi, (5) Nicobarese, (6) the Munda group, (7) Annamite-Muong, (8) the Cham group. If we except Cham, which most writers consider Malayo-Polynesian,

a conclusion which can hardly be doubted, then all these languages share numerous resemblances in fundamental vocabulary, extending to pronouns. Moreover, excepting Annamite, which has shed all its morphological processes, there are certain important derivational morphemes whose rather uncommon formal nature (infixes), combined with their basic functions in the grammar, absolutely excludes chance and makes borrowing a completely improbable explanation. I do not see how such coincidences as an infixed -*m* in the Mon of Burma and the languages of the geographically remote Nicobar Islands, both with agentive meaning, to mention only one of a number of such instances, can be the result of anything but genetic relationship.

Maspero has sought to demonstrate a close connection between Annamite and Thai, which he considers to be Sino-Tibetan. This case rests chiefly on the irrelevant argument from form only—the monosyllabism and tonicity of Annamite, in which it resembles Thai and Chinese. The extensive lexical resemblances to Thai, which hardly touch basic vocabulary, must be looked upon as mostly borrowing with some convergence. On the other hand, the mass of fundamental vocabulary points clearly in the direction of the Austroasiatic languages, and I do not see how any hypothesis of borrowing can explain it. If borrowed, the source is not evident, since Annamite now resembles one, now another, of the Austroasiatic languages. It often shows an independent development from a hypothetical reconstruction which can hardly be the result of anything but internal development from the ancestral Austroasiatic form. Thus Annamite *mōt*, 'one', makes sense as an independent contraction from **moyat*, found in this form only in the distant Mundari language of India. The language geographically nearest to Annamite Khmer has *muy*, presumably < *moy* with loss of final -*at*. Santali, the chief Munda language, has *mit* < **miyat* < **moyat*. The absence of the modest morphological apparatus of other Austroasiatic languages in Annamite cannot be used as an argument for any other relationship. The ancient maxim *ex nihilo nihil fit* may be appropriately applied in this instance.

Schmidt's further hypothesis of the relationship of Austroasiatic to the Malayo-Polynesian languages is of a far more doubtful nature. Most of the numerous etymologies proposed by Schmidt are either semantically or phonetically improbable or not attested from a sufficient variety of languages in one family or the other. Even with these eliminated, there remains a considerable number of plausible, or at least possible, etymologies, but very few of these are basic. Both language families employ prefixes and infixes, and the latter mechanism is certainly not very common. However, concrete

resemblances in form and meaning of these elements which can reasonably be attributed to the parent-language of both groups are very few. Only *pa-*, causative, seems certain. In view of this, the Austric hypothesis cannot be accepted on present evidence. It needs to be reworked, using Dempwolff and Dyen's reconstructed Malayo-Polynesian forms, as well as taking into account the Thai and Kadai languages, which, as we shall see, are related to Malayo-Polynesian.

The traditional theory regarding Thai is that it forms, along with Chinese, the Sinitic branch of Sino-Tibetan. Benedict has proposed the relationship of Thai to the Kadai group, in which he includes certain languages of northern Indo-China, southern continental China, and the Li dialects of the island of Hainan. He has further posited the relationship of this Thai-Kadai family to Malayo-Polynesian (Benedict 1942). Of the relation of Thai to the Kadai languages, which in the case of the Li dialects is particularly close, there can be no reasonable doubt. At the least, the traditional theory would have to be revised to include the Kadai languages, along with Thai, in Sinitic. I believe, however, that the connection of Thai with Chinese and Sino-Tibetan must be abandoned altogether and that Benedict's thesis is essentially correct. Thai resemblances to Chinese are clearly borrowings. They include the numerals from 3 on and a number of other words which are certainly the result of cultural contact. Thai is otherwise so aberrant that it must be at least another independent branch of Sino-Tibetan. Yet, when resemblances are found, the forms are always like Chinese—altogether too like Chinese, one should add. Applying a test suggested earlier, it is found that those words in Thai which resemble Malayo-Polynesian tend to reappear in the Kadai languages, while those which are like Chinese do so only rarely. The proportion of fundamental vocabulary resemblances between Thai-Kadai and Malayo-Polynesian runs to quite a high number, far beyond chance and hardly explainable by borrowing, in view of the geographical distances involved.

I believe that Benedict's thesis needs restatement in some details of grouping, where, as so often happens, he has been led astray by nonlinguistic considerations, in this case the importance of Thai as a culture language. Thai shows special resemblance to the Li dialects of such far-reaching importance that Benedict's twofold division of Kadai into Laqua-Li and Lati-Kelao must be emended to put Thai along with Li in the first subgroup. In addition, the language of the Mohammedan population of Hainan does not belong, interestingly enough, with the Li dialects of the rest of the island but forms a third subdivision alongside the continental Lati-Kelao. The emended picture is shown in the accompanying diagram.

```
                    ┌─────────────────────────────┐
                    │                             │
                  Kadai                   Malayo-Polynesian
          ┌─────────┴─────────┐      ┌───────┬───────┐
        Laqua              │        Lati    Kelao  Mohammedans
                  ┌────────┴────────┐                of Hainan
                  Li              Thai
```
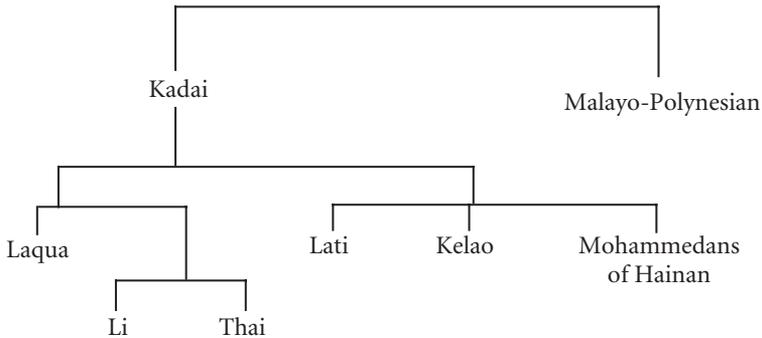
The Miao-Yao dialects of China have variously been called 'Mon-Khmer' (i.e. Austroasiatic), 'Sino-Tibetan', or 'independent'. There seems no good reason to classify them as other than a separate branch of Sino-Tibetan, no more divergent than, say, the Karen languages of Burma. The evidence cannot be summarized here. The Min-Hsia language has been variously called a 'Sino-Tibetan' or 'Austroasiatic' language with a Chinese overlay. It likewise seems to be Sino-Tibetan. When the obvious Chinese borrowings are accounted for, the language still appears to show a special affinity to Chinese in fundamentals, so that it should probably be included in the Sinitic subbranch.

The question is here raised concerning the status of the Nehari language of India, classed by Grierson as Munda. It has been strongly influenced by Kurku, a neighboring Munda language; but, when allowance is made for this, the fundamental vocabulary and morphology of the language do not resemble those of any other family in the area. It may therefore be the only language of an independent stock. More material is needed to decide this question.

In summary, the language families of Southeast Asia are probably the following: (1) Sino-Tibetan, (2) Austroasiatic, (3) Kadai-Malayo-Polynesian, (4) Andaman Islands, (5) Nehari(?)

1.3.4 *America north of Mexico*

The present discussion is restricted to a few remarks of somewhat impressionistic character because of my lack of acquaintance with the linguistic data from this area. However, even cursory investigation of the celebrated 'disputed' cases, such as Athabaskan-Tlingit-Haida and Algonkin-Wiyot-Yurok, indicate that these relationships are not very distant ones and, indeed, are evident on inspection. Even the much larger Macro-Penutian grouping seems well within the bounds of what can be accepted without more elaborate investigation and marshaling of supporting evidence. The difference between Oregon and California Penutian is comparable to that between any two

of the subdivisions of the Eastern Sudanic family in Africa. The status of Algonkin-Mosan and Hokan-Siouan and the position of Zuñí (which Sapir himself entered in the Azteco-Tanoan family with a query) strike me as the most doubtful points of Sapir's sixfold classification. The existence of a Gulf group, as set forth recently by Haas, with a membership of Tunican, Natchez, Muskoghean, and Timucua appears certain, as does the relationship of the Coahuiltecan languages both to the Gulf group and to the California Hokan in a single complex. Likewise, as Sapir pointed out, Yuki is probably no more than a somewhat divergent California Hokan language. The connection of Siouan-Yuchi and Iroquois-Caddoan with these languages is possible but far from immediately evident. Within Algonkin-Mosan, Salish-Chemakuan-Wakashan seems certain, as does Algonkin-Beothuk-Wiyot-Yurok (Beothuk may well be an Algonkin language). On the other hand, the relation of these two groups to each other and to Kutenai requires further investigation. Within the Azteco-Tanoan group it is clear that Kiowa is close to Tanoan and that Kiowa-Tanoan is related to Uto-Aztecan, as demonstrated by Trager and Whorf. The position of Zuñí, as noted above, is very doubtful.

1.4 Language and historical reconstruction

Ethnologists are rightly interested in comparative linguistic work, not so much for its own sake as for the light it sheds on other aspects of culture history. The basis for any discussion of this subject is inevitably the classic treatment of Sapir in his *Time Perspective in Aboriginal American Culture.* In spite of the brevity of this discussion, it is astonishingly complete, and there is little one would want to add to it, in spite of the lapse of time. The single most significant comment that might be made is that it serves as an essentially adequate basis for work in this field but that relatively little has been done toward the actual application of its principles. The problems involved are some of the most difficult in scientific cooperation and not easily solved. On the one hand, linguistic evidence is peculiarly suited to misapplication by ethnologists, who sometimes tend to use it mechanically and without at least an elementary understanding of the linguistic method involved. On the other hand, the linguist is often not greatly interested in problems of culture history, and the recent trend toward concentration in descriptive problems of linguistic structure draws him still further from the ordinary preoccupations of archeologists and historically oriented ethnologists. Perhaps the ultimate solution is an intermediate science, ethnolinguistics, which will treat the very important interstitial problems, both synchronic

and historical, which lie between the recognized fields of ethnology and linguistics.

The most important and promising recent development in this area is the possibility of establishing at least an approximate chronology for linguistic events in place of the relative time relations of classical historical linguistics. This method, known as 'glottochronology' and developed chiefly by Swadesh and Lees, works on the assumption that rate of change in basic vocabulary is relatively constant. A chronological time scale is provided by comparisons of vocabulary from different time periods of the same languages in areas with recorded history. The results thus far indicate an average of ca. 81 per cent retention of basic vocabulary in one millennium. Thus, by comparing two related languages for which no earlier recorded material is available, the percentage of basic vocabulary differences will allow for an approximation of the date of separation of the two forms of speech.

By combining with this a rigorous application of Sapir's insight regarding the probable center of origin of a linguistic group, on the basis of a center of gravity calculated from the distribution of genetic subgroups, an instrument of historical reconstruction surpassing any previous use of linguistic data for these purposes becomes possible.

The center-of-gravity method may be briefly described as follows. Within each of the genetic subgroups of a linguistic family, the center of distribution is selected. If the subgroup is itself divided into clear dialect areas, the central point of each dialect area is calculated and the position of all is averaged to obtain the probable center of dispersal of the subgroup. The centers of the various subgroups are then averaged to obtain the most probable point of origin for the entire family. A correction in order to minimize the influence of single aberrant groups may be made by calculating a corrected center of gravity from the one reached by the above method. The distance of the center of each subfamily is calculated from the center of gravity of the whole family. Then those subgroups which are most distant are weighted least, by multiplying the center of position of each subgroup by the reciprocal of the ratio of its distance to that of the most distant subgroup, and thus calculating a corrected value. Such results, mechanically arrived at, should, of course, be evaluated in terms of geographical and other collateral knowledge.

1.5 Goals, methods, and prospects

The goals and methods of comparative linguistics, particularly as applied to the field of primitive languages, are clear and generally agreed upon. The aims of this branch of science might be phrased in terms of the establishment of all

possible genetic relationships between languages, the detection of all borrowings and the direction they have taken, and the maximal reconstruction of the ancestral languages which have given rise to the present languages. This is of value not only for its own sake and because these results can be employed toward general historical reconstruction but also because it gives us our basic knowledge of historic change in language under diverse circumstances. It is not until considerable data have been amassed in this field and a considerable variety of historical development in different areas has been traced that questions regarding overall change from one morphological or phonological type to another, leading to general laws of linguistic change, can ever be possible.

Problems of method, also, are in the main agreed upon. These resolve themselves into two main types: those pertaining to the determination of relationship and those concerning reconstruction. The latter problems are less controversial, and, in the United States at least, there is general agreement on the employment of what are essentially the procedures of classical Indo-European linguistics. The problems of establishing genetic relationships beyond the most self-evident ones, such as those of Powell in North America, admittedly involve more differences of opinion both in Europe and in America. The abandonment of concrete criteria in favor of meaning without form or form without meaning and the abandonment of the traditional view regarding genetic relationship in some parts of the world in favor of the apparent profundity of analyses in terms of superposed strata have led only to increasing confusion and conflicting analyses, as they inevitably must. Moreover, only on the basis of clearly defined families established through specific form–meaning resemblances can reconstruction be attempted and with it the possibility of the study of historic process in language.

The greatest single obstacle to the rapid future growth of the field does not lie, however, in any conflict regarding aims or methods. It is rather the lack of trained people in sufficient number to provide the descriptive data for a vast number of languages, some of them near extinction. The top-heavy concentration of linguistic scientists in the area of a very small number of language families of Eurasia and the extreme paucity of fully trained workers in such large areas as South America and Oceania are a grave handicap to future development of this field, as well as of linguistics as a whole. At the last meeting of the Linguistic Society of America, approximately 90 per cent of the papers presented on specific languages concerned a single language family, Indo-European.

The absence of effective liaison even between anthropological linguists and other branches of anthropology and its nonexistence in the case of other

linguists, while an understandable consequence of the contemporary trend toward specialization, are likewise dangerous. Unless these situations are met and to some degree overcome, comparative linguistics must fall far short of the inherent possibilities afforded by the transparency of its material and the sophistication of its method of making a unique and significant contribution to the science of anthropology as a whole.

References

Benedict, P. 1942. 'Thai, Kadai and Indonesian: a new alignment in Southwestern Asia', *American Anthropologist* 44: 576–601.

Boas, F. 1920. 'The classification of American languages', *American Anthropologist* 22: 367–76.

—— 1929. 'Classification of American Indian languages', *Language* 5: 1–7.

Bonfante, G. 1945, 'On reconstruction and linguistic method', *Word* 1: 83–94, 132–61.

Capell, A. 1939–40. 'The classification of languages in North and Northwest Australia', *Oceania* 10: 241–72, 404–33.

—— 1941–3. 'Languages of Arnhem Land, North Australia', *Oceania* 12: 364–92; 13: 24–50.

Greenberg, J. H. 1949–50. 'Studies in African linguistic classification', *Southwestern Journal of Anthropology* 5: 79–100, 309–17; 6: 47–63, 143–60, 223–37, 388–98.

Hall, Robert, Jr. 1946. 'Bartoli's "Neolinguistica." ', *Language* 22: 273–83.

Hoenigswald, H. 1950. 'The principal step in comparative grammar', *Language* 26: 357–64.

Hockett, C. 1948. 'Implications of Bloomfield's Algonkin studies', *Language* 24: 117–31.

Jakobson, R. 1931. 'Principes de phonologie historique', *Travaux du Cercle Linguistique de Prague* 4: 247–67.

Kroeber, A. L. 1913. 'The determination of linguistic relationship', *Anthropos* 8: 389–401.

—— 1924. 'Relationship of the Australian Languages', *Journal and Proceedings of the Royal Society of New South Wales* 57: 101–17.

Martinet, A. 1950. 'Some problems of Italic consonantism', *Word* 6: 26–41.

Meillet, A. 1914. 'Le problème de la parenté des langues', *Scientia* 15, No. 35: 3.

Meinhof, Carl. 1912. *Die Sprachen der Hamiten.* Hamburg: L. Friederichsen.

—— 1915–19. 'Sprachstudien im ägyptischen Sudan', *Zeitschrift für Kolonialsprachen* 6: 161–205; 7: 36–133, 212–50, 326–35; 8: 46–74, 110–39, 170–96, 257–67; 9: 43–64, 89–117, 167–204, 226–55.

Ray, S. 1907. *Linguistics.* (Reports of the Cambridge Anthropological Expedition to Torres Straits, Vol. III.) Cambridge: Cambridge University Press.

Schmidt, W. 1906. *Die Mon-Khmer Völker.* Braunschweig: F. Vieweg & Sohn.

—— 1913. 'Die Gliederung der australischen Sprachen', *Anthropos* 9: 526–54.

—— 1914. Ibid. *Anthropos* 9: 930–1018.

——1917. Ibid. *Anthropos* 12: 437–39.

——1918. Ibid. *Anthropos* 13: 747–817.

——1919. *Die Gliederung der australischen Sprachen*. Vienna: Mechitharisten-Buchdruckerei.

Swadesh, M. 1951. 'Diffusional cumulation and archaic residue as historical explanation', *Southwestern Journal of Anthropology* 7: 1–21.

Trubetskoy, N. 1928. *Actes du premier Congrès International de Linguistes à La Haye*. Leiden: Sijthoff.

Tucker, A. N. 1940. *The Eastern Sudanic Languages*. London: Oxford University Press.

Vendryes, J. 1925. *Language: A Linguistic Introduction to History*, trans. Paul Radin. New York: A. Knopf.

Werner, A. 1915. *The Language Families of Africa*. London: Society for Promoting Christian Knowledge.

This page intentionally left blank

2

Genetic relationship among languages*

The establishment of valid hypotheses concerning genetic relationships among languages is a necessary preliminary to the systematic reconstruction of their historical development. The appropriate techniques cannot be applied to languages chosen at random but only if preliminary investigation has already indicated the likelihood of the success of such an enterprise. Correct hypotheses of relationship are also of very real significance to the archaeologist, the physical anthropologist, the ethnologist, and the culture historian, even in those instances in which systematic linguistic reconstruction has not yet begun and may, indeed, in our present state of descriptive knowledge be of only limited feasibility. The considerations advanced in this chapter are intended as a realistic analysis of the factors involved in the formulation of reliable hypotheses of such relationships. It should be possible, through clarification of the assumptions involved, to resolve the conflicting classifications found in certain areas. It is likewise hoped that a sufficient basis will be presented so that the nonspecialist can intelligently evaluate alternative classifications through an independent examination of the linguistic evidence itself.

Hypotheses concerning genetic relationship among languages are established by comparing languages. But languages are complex wholes which exhibit many facets, and the question which inevitably arises at the outset is one of relevance. Are all aspects of language equally germane for comparison? A language contains a set of meaningful forms (morphemes), themselves composed of meaningless sound types (phonemes) and entering into various combinations in accordance with the rules of its grammar. The meaningful forms (morphemes) may themselves be roots, in which case they are normally assigned to the lexicon, or nonroots (affixes) with derivational or

inflectional grammatical function, in which case their description is part of the grammar. In either instance they involve both sound and meaning.

It is clear that, in principle, the connection between sound and meaning is arbitrary, in the sense that any meaning can be represented by any combination of sounds. A dog may as easily be called *Hund, cane, sabaka*, or *kalb* and, in fact, is—in German, Italian, Russian, and Arabic, respectively. Moreover, the thousands of meaningful forms of any language are basically independent. Except for the occasional avoidance of homonyms, which involves an exceedingly small limitation, the principle holds in general that, just because you call a dog a *dog*, it does not mean you have to call a cat a *cat*. It is unlikely, however, that you will call it a *dog*. From these two principles of the arbitrariness of the sound–meaning connection and the independence of meaningful forms, it follows that resemblances beyond chance in both form and meaning require a historical explanation, whether through borrowing or through common origin.

By 'lexical resemblance' will be meant similarity in sound and meaning of root morphemes—e.g. English '*hænd* and German '*hant*, both meaning 'hand'. By 'grammatical resemblance' will be meant similarity of both sound and meaning in nonroot morphemes, e.g. English -*ər* and German -*ər*, both indicating the comparative of adjectives. Both lexical and grammatical resemblances thus defined are relevant as evidence for historical relationships. On the other hand, similarity in meaning not accompanied by similarity in sound or similarity in sound without corresponding similarity in meaning may be considered of negligible value. Thus the presence of sex gender expressed by morphemes without phonetic resemblance or the existence of tonal systems without specific form–meaning similarities in the forms employing tones should be excluded as arguments for historical connection.

The order of meaningful elements may be considered a formal characteristic, like sound. In syntactic constructions only two possibilities usually occur in the arrangement of forms, A either preceding or following B, as contrasted with the numerous possibilities of sound combinations. Hence arguments based on word order are of minor significance. This is all the more so because the kinds of constructional meaning which may be significant are necessarily small, e.g. dependent genitive or actor–action. Historically unconnected occurrences of such resemblances are therefore extremely likely and heavily documented.

The order and meaning of morpheme classes within complex words in certain cases offer far greater combinational possibilities. The meaning possibilities involved are more numerous than for syntactic construction, though less than for sound–meaning resemblances. For example, within the verb

complex we may have such meaning categories as pronominal subject, direction of action relative to speaker, tense, transitivity, or nontransitivity, etc. This method can be used with real effect only in polysynthetic languages, those with complex internal word structure. Moreover, lack of agreement in such matters is not significant where sufficient specific sound–meaning agreements in morphemes are found. For example, the verb structures of Russian and Hindustani are quite different; once the periphrastic construction based on the participles became established, the whole elaborate inherited inflectional mechanism of the Indic verb was eliminated at one stroke. Even where such agreements are found among polysynthetic languages, it would seem to provide merely confirmation, however welcome, of results also attainable by the more generally applicable method of morpheme comparison.

Granted that sound–meaning similarities of morphemes weigh most significantly in determining historic relationships, it is evident that not all such resemblances need stem from historic factors. Thus Didinga, a language of the Anglo-Egyptian Sudan, has *badh* in the meaning 'bad' and *man* means 'man' in Korean. Moreover, although, as stated previously, the connection between sound and meaning is arbitrary, that is, unpredictable, there does exist in certain instances a well-marked tendency for greater than chance association between certain sounds and meanings. Examples are the nursery words for 'mother' and 'father' and onomatopoetic terms. This factor will increase slightly the number of sound–meaning resemblances between any two languages. If we call this source of resemblance 'symbolism', then there are four classes of causes for sound–meaning resemblances, two of which—chance and symbolism—are nonhistoric, while the remaining two—genetic relationship and borrowing—involve historic processes.[1]

The two basic methodological processes then become the elimination of chance and symbolism leading to hypotheses of historic connections and the segregation of those instances in which borrowing is an adequate explanation from those on which genetic relationship must be posited.

The most straightforward method of eliminating chance would be the calculation of the expected number of chance resemblances between two languages, taking into account their respective phonemic structures. In practice, this proves extremely difficult, and no satisfactory technique for its accomplishment has yet been devised. Moreover, it requires, in addition to consideration of the possibilities of phonemic combination, a frequency

---

[1] Further complex causes, involving certain combinations of these four fundamental causes, will be discussed in the following chapter on 'The problems of linguistic subgrouping' [§3-Ed.].

weighting of phonemes. If both languages show, as is normal, considerable variation in the frequency of the various phonemes and if similar phonemes are among the most frequent in each language, the overall expectation of chance coincidences is increased. More practicable would be a percentage count of resemblances among large numbers of pairs of presumably unrelated languages. This would also have the advantage of taking into account resemblances due to symbolism also. Where the percentage of resemblance between languages is very high, say 20 per cent or more, some historic factor, whether borrowing or genetic relationship, must be assumed. Where the proportion of similarities is significantly lower, a consideration of the qualitative characteristics of the sound–meaning resemblances found and the broadening of the basis of comparison to other languages, usually numerous, which show resemblances to the pair being considered (mass comparison) bring into play factors of the highest significance which should always insure a decisive answer. These factors quite overshadow the mere percentage of resemblances. In many instances this, if small, may be approximately the same between several pairs of languages, yet in some cases there will be certainty of historic relationship beyond any reasonable doubt and in some others no compelling reason to accept such an explanation.

Qualitatively, not all sound–meaning similarities are of equal value as evidence for a historical connection. For example, the longer a form, the less likely does it become that chance is an explanation. From this point of view, 'intər'našənəl in language A and intərnatsjo'nal in language B is far more likely the result of historic factors than are -*k,* 'locative', in language A and -*g,* 'locative', in language B.

The natural unit of interlingual comparison is the morpheme with its alternant morphs.[2] The presence of similar morph alternants in similar environments is of very great significance as an indication of historical connection, normally genetic relationship. This is particularly so if the alternation is irregular, especially if suppletive, that is, entirely different. The English morpheme with alternants *gud-, bet-, be-,* with the morph alternant *bet-* occurring before -*ər,* 'comparative', and the alternant *be-* before -*st,*

---

[2] In addition to the morpheme as the fundamental unit of interlanguage comparison, resemblances below the morphemic level involving units that might be called 'submorphs' are also the result of genetic relationship. The following example will help to indicate what is meant. In certain languages of the Adamawa group of Niger-Congo languages, as a survival of a noun suffix -*ma* used with names of indefinitely divisible substances, e.g. water, fat, a far larger than chance number of nouns with meanings of this sort end in -*ma* or -*m,* depending on the languages. However, there is no class system involving plural formation or adjective or noun agreement, as in other Nieger-Congo languages. The methodology of Greenberg (1957) would not lead to the analysis of this element as a morpheme in a synchronic description of these languages. Such cases are marginal and too infrequent to be anything but a reinforcement of what is provable by normal morpheme comparison.

'superlative', corresponds in form and conditions of alternation with German *gu:t-*, *bes-*, *be-*, with *bes-* occurring before *-ər*, 'comparative', and *be-* before *-st*, 'superlative'. We have here not only the probability that a similar form is found in the meaning 'good' but that it shows similar and highly arbitrary alternations before the representatives of the comparative and superlative morphemes. The likelihood that all this is the result of chance is truly infinitesimal.

Similar rules of combinability, even without alternations in form, are also of considerable significance. In Niger-Congo languages, not only are forms similar to *to* in the meaning 'ear' found widely, but they are also found in construction with the same classificational affix *ku*.

Such indications of historical connection founded on morphological irregularities of form and combinability may not always be found. Many languages of isolating or of highly regular structure will have few or no morph alternants. Even where originally present, they are subject to constant analogical pressure toward replacement by regular alternations. Hence their chance of survival in related languages is not great. Where they are found, however, they are precious indexes of historical relationships.

Another factor bearing on the value of particular resemblances is semantic plausibility. This is greatest where the meanings are similar enough to have been given as translation equivalents for the same term in some third, usually European, language or for translation equivalents in two other languages. Semantic plausibility likewise attaches to comparisons involving single-step, widely attested shifts in meaning, e.g. 'moon' and 'month'. The more intermediate semantic steps allowed, the larger the chance of obtaining form–meaning similarity, some of which may indeed stem from historical connection. But the greater the methodological latitude permitted, the less plausible is each individual comparison.

Considerations derived from the extension of comparison beyond the pair of languages initially considered are of fundamental importance. The problem as to whether the resemblances between two languages are merely the result of chance plus symbolism can then be subjected to further and decisive tests. Let us say that, as is usually the case, one or more other languages or language groups resemble the two languages in question. The following fundamental probability considerations apply. The likelihood of finding a resemblance in sound and meaning in three languages is the square of its probability in two languages. In general, the probability for a single language must be raised to the ($n - 1$th) power for n languages. Thus if five languages each showed a total of 8 per cent sound–meaning resemblance to one another, on a chance basis one would expect $(0.08)^4$ or 0.00004096 resemblances in all

five languages. This is approximately 1/25,000. In other words, were one to compare sets of one thousand forms from all five languages, one would have to do this twenty-five times before a single instance of a resemblance in all five languages would occur. Even recurrence in three languages would be rare on a chance basis, 0.0064, that is, less than 1 per cent. Hence the presence of a fair number of recurrent sound–meaning resemblances in three, four, or more languages is a certain indication of historical connection.

Finally, there are considerations based on the phonetic form. The presence of recurrent, i.e. non-unique correspondences, adds greatly to the value of the comparison. In this area, also, mass comparison is of significance. Are the forms found in a number of languages such as to suggest that they are changed forms of a common original? The bringing-in of closely related languages on each side will then show tentative reconstructions converging as we go back in time. This procedure is not possible where only two languages are being compared.

Assuming that the factors just cited lead to the establishment of a historical connection, there still remains the problem of whether the resemblances in question can be explained by borrowing. While in particular and infrequent instances the question of borrowing may be doubtful, it is always possible to tell whether a mass of resemblances between two languages is the result of borrowing. A basic consideration is the *a priori* expectation and the historical documentation of the thesis that borrowing in culture words is far more frequent than in fundamental vocabulary and that derivational, inflectional, and pronominal morphemes and morph alternations are the least subject of all to borrowing. While it cannot be said that any single item might not on occasion be borrowed, fundamental vocabulary is proof against mass borrowing. The presence of fundamental vocabulary resemblances and resemblances in items with grammatical function, particularly if recurrent through a number of languages, is a sure indication of genetic relationship. Where a mass of resemblances is due to borrowing, they will tend to appear in cultural vocabulary and to cluster in certain semantic areas which reflect the cultural nature of the contact, and the resemblances will point toward one or, at most, two or three languages as donors. The forms will be too similar to those found in these particular languages, considering the great differences in other respects and the consequent historic remoteness of the relationship, if it really existed. Thus the Romance loanwords in English are almost all close to the French forms, in addition to hardly penetrating the basic vocabulary of English. Were English really a Romance language, it would show roughly equal similarities to all the Romance languages.

The presence of recurrent sound correspondences is not in itself sufficient to exclude borrowing as an explanation. Where loans are numerous, they often show such correspondences; thus French loanwords in English often show Fr. *š* = Eng. *č*, Fr. *ã* = Eng. *æn* (*šãs* : *čæns*; *šãt* : *čænt*; *še:z* : *čejr*, etc.).

All these principles are well illustrated from Thai, whose resemblances to Chinese are the result of borrowing rather than genetic relationship, as is being realized more and more. Most of the resemblances usually cited between Thai and Sino-Tibetan languages, such as the existence of a tonal system, involve sound only or meaning only and are therefore irrelevant. The specific resemblances found with Sino-Tibetan languages always occur in forms found in Chinese, usually to the exclusion of other Sino-Tibetan languages. The specific form, even when found elsewhere, is always very close to Chinese. Moreover, the resemblances cluster in a few semantic spheres, the numerals from 2 to 10 and a few names of metals and domestic animals. In contrast, the Thai resemblances to the Kadai languages and Malayo-Polynesian tend to recur throughout the family, not just in some single language; are basic; do not concentrate in any particular semantic area; and exhibit an independence of form which excludes any particular Kadai or Malayo-Polynesian language as a source.

Borrowing can never be an overall explanation of a mass of recurrent basic resemblances in many languages occurring over a wide geographical area. It is sometimes adduced in this *ad hoc* fashion. Since we find independent sets of resemblances between every pair of languages, among every group of three languages, and so on, each language would have to borrow from every other. A thesis of borrowing to account for resemblances must be specific, pointing out which peoples have borrowed from which, and it must be plausible in terms of the factors just cited. It may be added that the vast majority of languages do not display mass borrowing, and, where it does occur, it is easily detected.

The method for discovering valid relationships described here may be summarized as resting on two main principles—the relevancy of form–meaning resemblances in morphemes to the exclusion of those based on form only and meaning only and the technique of group comparison of languages. Some of the reasons for this latter emphasis have been adduced earlier. There are further considerations which recommend this procedure. Instead of comparing a few or even just two languages chosen at random and for linguistically extraneous reasons, we proceed systematically by first comparing closely related languages to form groups with recurrent significant resemblances and then compare these groups with other similarly constituted groups. Thus it is far easier to see that the Germanic languages are related to

the Indo-Aryan languages than that English is related to Hindustani. In effect, we have gained historic depth by comparing each group as a group, considering only those forms as possessing the likelihood of being original which are distributed in more than one branch of the group and considering only those etymologies as favoring the hypothesis of relationship in which tentative reconstruction brings the forms closer together. Having noted the relationship of the Germanic and Indo-Aryan languages, we bring in other groups of languages, e.g. Slavonic and Italic. In this process we determine with ever increasing definiteness the basic lexical and grammatical morphemes in regard to both phonetic form and meaning. On the other hand, we also see more easily that the Semitic languages and Basque do not belong to this aggregation of languages. Confronted by some isolated language without near congeners, we compare it with this general Indo-European rather than at random with single languages. It is a corollary of the considerations advanced here that if a language has no close relatives, it is more difficult to find its distant relatives. Therefore, we should begin with well-defined groups of more closely related languages and leave such isolated cases to be considered after more widespread families have been constituted. Table 1 will show that it is not mere percentage of resemblances between pairs of languages which is decisive, except for quite close relationships, but rather the setting-up of restricted groups of related languages which then enter integrally into more distant comparisons.

Table 1

	A	B	C	D	E	F	G	H	I
Head	kar	kar	se	kal	tu	tu	to	fi	pi
Eye	min	ku	min	miŋ	min	aš	min	idi	iri
Nose	tor	tör	ni	tol	was	waš	was	ik	am
One	mit	kan	kan	kaŋ	ha	kan	kεn	he	čak
Two	ni	ta	ne	kil	ne	ni	ne	gum	gun
Blood	kur	sem	sem	šam	i	sem	sem	fik	pix

In examining the forms in Table 2, the hypothesis immediately arises that A, B, C, and D form a related group of languages. We will call this 'Group I'. It is also apparent that E, F, and G constitute another related group (Group II), and that H and I are likewise connected (Group III). The hypothesis will also suggest itself that Groups I and II are related. On the other hand, the material cited offers no real support for the relationship of Group III to Groups I and II. If we look more closely, however, we will see that languages B and E show no likely cognates, whereas E has a form for 'one', *ha*, closely

resembling *he* in language H. E therefore shows a higher percentage resemblance to H than to B on the basis of isolated comparison. Yet the hypothesis that E is related to H rather than to B would hardly occur as a realistic one when all the relevant evidence from languages more closely related to E, B, and H is taken into consideration. The tables of percentages of resemblances among pairs of languages which are sometimes cited as evidence can at times be quite misleading, nor can elaborate statistical manipulations of these quantitative data add to their validity.

There is the further consideration that isolated hypotheses are less significant in their culture-historical implications and may even, on occasions, lead to erroneous conclusions. Thus it is no doubt true that Albanian, Bengali, and Swedish are related; but if all the intervening languages are unclassified or stated to be independent, some rather questionable historical deductions would be made. In addition, isolated hypotheses may lead to fruitless controversies, in which both parties have correct but only partial answers. Thus in aboriginal South America, where widespread relationships on a scale hitherto unrecognized actually exist, there are controversies which, transposed in terms of the Eurasian area, might run somewhat as follows. One investigator states that Albanian is related to Greek. The other disagrees and maintains, on the contrary, that it is related to Italian. Both present fairly convincing cases, since their hypotheses are correct, though, of course, a far stronger case could be presented for Indo-European as a whole, with the positions of Greek, Italian, and Albanian defined within it. Other linguists viewing the controversy either come to the cynical conclusion that, with sufficient effort, you can present a convincing case for any relationship, real or fancied, or decide that we need several more generations to gather the data necessary to decide the controversy.

As a heuristic principle, the swiftest and surest method of bringing into play many of the considerations discussed here is the compilation of comparative fundamental vocabularies of all the languages of an extended area. This accomplishes a number of purposes simultaneously. It involves the aspect of language least subject to borrowing outside grammatical elements. The forms are generally of fair length. Semantic straightforwardness is attained by using the translation equivalent of the same term in English or whatever language is used as the language of translation. The tendency of similar forms to appear in a number of languages, as well as the plausibility of descent from a common original, can easily be noted. The presence of recurrent phonetic correspondences can be seen without great difficulty. If, as is often the case, word lists or dictionaries include noun plurals or other morphological facts, even details of morphological combinations and alternations can be taken

Table 2*

	One	Two	Three	Head	Eye	Ear	Nose	Mouth	Tooth
Breton	ünan	dau	tri	penn	lagad	skuarn	fri	genu	dant
Irish	öːn	dɔː	tri	kjan	suːlʲ	kluəs	srɔːn	bjal	fjakəlʲ
Welsh	ɨn	dai	tri	pen	ləgad	kłist	truɨn	keg	dant
Danish	en	toːʔ	treːʔ	hoːdhə	ɔjə	oːrə	nɛːsə	monʔ	tanʔ
Swedish	en	tvo	tre	hüvud	öga	öra	näsa	mun	tand
Dutch	eːn	tveː	driː	hoːft	oːx	oːr	nöːs	mont	tant
English	wən	tuw	θrij	hɛd	aj	ihr	nowz	mawθ	tuwθ
German	ajns	tsvaj	draj	kopf	augə	oːr	naːze	munt	tsaːn
French	œ̃,yn	dö	trwa	tɛːt	œj/jö	orɛːj	ne	buːš	dã
Italian	uno, una	due	tre	tɛsta	okkjo	orekkjo	naso	bokka	dɛnte
Spanish	un, una	dos	tres	kabesa	oxo	orexa	naso	boka	diente
Rumanian	un	doj	trej	kap	okiu	ureke	nas	gurə	dinte
Albanian	nʼə	dü	tre	kokə	sü	veʃ	hundə	goja	ðəmp
Greek	enas	ðjo	tris	kefáli	máti	aftí	míti	stóma	ðóndi
Lithuanian	vienas	du	triːs	galva	akis	ausis	nosis	burna	dantis
Latvian	viens	divi	triːs	galva	atss	auss	deguns	mute	zobs
Polish	jeden	dva	tši	glova	oko	uxo	nos	usta, gēba	zõp
Czech	jeden	dva	tši	hlava	oko	uxo	nos	usta	zup
Russian	adʲin	dva, dvʲe	trʲi	gəlavá	óko	úxo	nos	rot	zup
Bulgarian	edin	dva	tri	glava	oko	uxo	nos	usta	zəb
Serbo-Croatian	jedan	dva	tri	glava	oko	uho	nos	usta	zub
Finnish	üksi	kaksi	kolme	päː	silmä	korva	nenä	suː	hammas
Estonian	üks	kaks	kolm	pea	silm	kõrv	nina	suː	hammas
Hungarian	ed	keːt	haːrom	föː, fej	sem	fül	orr	saːj	fog
Basque	bat	bi	hirür	bürü	begi	belari	südür	aho	orts

* Table 2 is in a broad phonemic transcription. I have largely left it as it is, except to change digraphs to IPA symbols and correct some minor errors. See also §6, Table 7-Ed.

into account. Most important of all, perhaps, is that, where more than one family is represented, as is always the case when the languages examined are from an extensive area, the contrast between the relatively numerous and qualitatively superior resemblances among related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances.

A relationship may sometimes be first suggested by agreement in some strikingly irregular morphological alternation or very full agreement in some set of grammatical affixes. For example, I was first led to entertain the hypothesis of the relation of Zaghawa to Kanuri and Teda, to form the Central Saharan family, by a remarkable agreement in a conjugational paradigm in which the morphemes of the first two persons were suffixed while those of the third person were prefixed. All the personal affixes were,

moreover, phonetically similar. An examination of the fundamental vocabulary of these languages, which followed, amply confirmed the result. As a general procedure, however, the great advantage of vocabulary is the large number of essentially independent items it furnishes which are comparable from language to language and which are always present. Moreover, where little information is available about languages, the data are far more likely to be lexical than grammatical. All available grammatical information should be systematically examined, but vocabulary leads most swiftly to the correct hypotheses as a general rule. The effectiveness of mass comparison of basic vocabulary, for all its apparent simplicity, is illustrated in Table 2 by only a few forms from all the contemporary languages of Europe.[3]

Note that, even by the time the second word has been examined, the correct hypothesis emerges. The subsequent words fully confirm the initial hypothesis again and again. I believe that it is not generally realized how great is the number of different ways in which a given number of languages can be genetically classified. If, for example, there are four languages, A, B, C, and D, the following classifications are possible: (1) into one family in one way /ABCD/; (2) into two families, seven ways, /ABC/D/, /ABD/C/, /ACD/B/, /BCD/A/, /AB/CD/, /AC/BD/, /AD/BC/; (3) into three families, six ways, /AB/C/D/, /AC/B/D/, /AD/B/C/, /CD/A/B/, /BD/A/C/, /BC/A/D/; (4) into four families, one way, /A/B/C/D/. This makes a total of fourteen ways. With the increasing number of languages, the number of distinct ways of classifying increases at a tremendous rate. For eight languages, the number is already 4,140.[4] For twenty-five, the number of languages in Table 2, the possible ways of classifying are 4,749,027,089,305,918,018, that is, nearly five quintillion or 5×10^{18}. Otherwise put, the method of vocabulary comparison, after the examination of two words, has already selected out of nearly five quintillion possibilities exactly that one which is, by universal consent and much other evidence, accepted as the correct one! There must be good reasons for this result. It has been the purpose of this chapter to explain what they are.

The correct hypothesis may not appear quite so quickly in every case, but even supposedly distant relationships, e.g. Algonkian-Ritwan, Austroasiatic, appear fairly soon and are confirmed again and again.

The methods outlined here do not conflict in any fashion with the traditional comparative method. They may be viewed rather as an attempt to

---

[3] The mass comparison of basic vocabulary is actually the oldest method employed. Essentially correct results were obtained in the eighteenth century even from very poor descriptive material. The earliest instance of which I am aware is von Strahlenberg (1730).

[4] The number of possible classifications for $n+1$ languages is obtained recursively from that of n languages by the formula: $p_{n+1} = \sum_{i=0}^{n} \binom{n}{i} p_i$; cf. Ore (1942).

make explicit the first step in that method itself, for we cannot begin systematic reconstruction until we know which languages to compare. The application of the comparative method is a continuous process, and, in principle, there is no sharp break between its initial and its more advanced stages. Thus at the very beginning, under the guise of the apparently synchronic concept of sound resemblance, what is being considered by the experienced observer is the diachronic probability that the compared sounds are independent continuations of the same original sound. This, on the whole, coincides with synchronic similarity on an articulatory basis, since sound changes normally involve the change of a single feature of articulation at a time. Such judgments are further guided by our accumulation of knowledge of attested sound changes in other language groups.

Indeed, the very act of noting form–meaning resemblances involves notions of correspondence and reconstruction. If we compare English 'hænd and German 'hant, we do so on the assumption that the *h* in both forms corresponds, that English *æ* corresponds to German *a*, etc., and not to *h* or to *n*. Moreover, however incompletely, reconstruction of an original sound system is involved. If I equate English *æ* and German *a*, this is on the assumption of a common origin; and the original form, while not precisely determined, is strongly limited to those sounds which could have given rise to both *æ* and *a*. It was very probably some low, unrounded vowel like *a*, far less likely *i*, and certainly not *k*. Moreover, the procedure of mass comparison advocated here helps to make the conjecture regarding the ancestral sound ever more precise by the addition of further forms from additional languages. The test provided by the tendency to converge backward in time as each form is compared within its own subgroup of the larger family which was earlier stated as an integral part of the method determining genetic relationship involves this type of preliminary reconstruction.

The further application of the comparative method resulting in more precise reconstruction is built on a systematic utilization of the etymologies disclosed by preliminary comparison. These etymologies are of varying strength, depending on the following factors: phonetic resemblance, semantic plausibility, breadth of distribution in the various subgroups of the family, length, participation in parallel irregular alternations, and the occurrence of sound correspondences found in other etymologies which are strong on these same grounds. More advanced reconstruction will add some new etymologies and/or invalidate some of the weaker original ones. Those etymologies that are strong on the basis of the criteria mentioned cannot, I believe, be invalidated by the later reconstructions of the sound system. It is rather the efficiency of such reconstructions in explaining these etymologies that is the touchstone

by which such reconstructions are tested. Unless etymologies of this degree of strength existed, we would not have been justified in drawing a conclusion of genetic relationship in the first place.

This is clear from actual practice. The Latin form *quattuor*, 'four', is a first-rate etymology because it is long, exhibits recurrent correspondences in most of its parts, occurs in every branch of Indo-European, and is semantically straightforward. However, the double *t* remains unexplained. The Indo-Europeanist does not therefore reject *quattuor* as a valid etymology. He seeks rather to explain it by other recognized historical processes, such as the analogical influence of other numerals. In other words, reconstruction of an original sound system has the status of an explanatory theory to account for etymologies already strong on other grounds. Between the *ᵛvaida* of Bopp and the *ɣwoidxe* of Sturtevant lie more than a hundred years of the intensive development of Indo-European phonological reconstruction. What has remained constant has been the validity of the etymologic relationship among Sanskrit *veda*, Greek *woida*, Gothic *wait*, all meaning 'I know', and many other unshakable etymologies both of root and of nonroot morphemes recognized at the outset. And who will be bold enough to conjecture from what original the Indo-Europeanist one hundred years from now will derive these same forms? Thus reconstruction is in itself a continuous process, although the human effort may be discontinuous and pause after the first stages through lack of refined descriptive data or qualified and interested specialists; and this process goes onward indefinitely into the unknown future.

References

Greenberg, Joseph H. 1957. 'The definition of linguistic units', *Essays in Linguistics*. Chicago: University of Chicago Press, 18–34.

Ore, Oystein. 1942. 'Theory of equivalence relations', *Duke Mathematics Journal* 1942: 573–627.

von Strahlenberg, Philip John. 1730. *An Historic-geographic Description of the Northern and Eastern Parts of Europe and Asia* (English translation (London, 1738) from German original edition of 1730).

This page intentionally left blank

3

The problem of linguistic subgroupings*

Both German and archaic English have in common an -st suffix indicating the second person singular of the verb: German *du denk-st*; English *thou think-est*. Given the relatively close genetic relationship of English and German, the obvious explanation is one of common origin, that is, that both forms are the continuations of a Proto-Germanic -st, second person singular suffix in both English and German. The known history of these forms, however, shows something quite different. Both German and English inherited a second person singular in -s. In both languages, independently, the suffixation of the independent pronoun in question led to a form in -st, which then spread by analogy to noninterrogative constructions: *þinkes-þu, þinkest; denkes-du, denkest.*[1] The forms are thus the results neither of common origin from an original *-st nor of borrowing; they are convergent developments. Yet convergence in this instance cannot mean accident. Had not both languages possessed an inherited second person in -s, a second person singular independent pronoun beginning with a dental, and an interrogative construction involving inversion, the common end result could not have occurred. A common stage had been set. Small wonder, then, that a similar act ensued. We have, then, a specific resemblance in form and meaning which is a complex resultant of genetic relationship and convergence. Similar instances can be found in nonlinguistic cultural history. Oriental scholars have long been struck by the general similarities of Egyptian and Sumerian cultures, accompanied by only minor instances of resemblances that can have resulted from direct historic contact. In this sense Egyptian and Sumerian cultures are not historically related. Yet it is surely no accident that in the same general area of the world and in the same chronologic period, cities, priesthoods, and

[1] Perhaps other factors entered into the history of the forms. Both in English and in German there were two preterite presents with inherited -st from earlier *t-t in the second person singular, which could have provided an analogical model. These are OE *wāst*, Modern German *weisst*, 'thou knowest', and Modern English *must*, German *musst*. In this case, the basis for convergence is even wider.

a host of other important similar cultural features developed. Here again, on a genetically related base in neolithic culture, like developments occurred independently. This process in language is what Sapir called DRIFT; it may quite simply be defined as convergence among genetically related languages.

Other complex causes of sound–meaning resemblances may be suggested. Bloomfield once cited as an example of the hazards of linguistic reconstruction the possibility of constructing a Proto-Central-Algonkian word for 'whiskey'. These languages, all starting from cognate words for *fire* and cognate words for *water* and in possession of a common Algonkian pattern of compounding, produced words similar in form and meaning which are not the result of common inheritance. This instance differs somewhat from the previous Germanic example, in that the developments were not independent. The semantic pattern of compounding a word *fire-water* for 'whiskey' presumably spread through borrowing of meaning pattern only (semantic borrowing) over a large area of Amerindian languages, including both Algonkian and non-Algonkian languages. Only in the Algonkian languages, however, did they produce sound–meaning resemblances because of the existence of a common genetic basis. The Algonkian forms for 'whiskey' result, then, from a complex of common inheritance and semantic borrowing.

In Portuguese, as spoken in the United States, the term *livraria*, from meaning 'bookshop' as it does in European Portuguese, has taken on the meaning 'library' because of its resemblance in sound to the English word. This is again a resemblance in both sound and meaning. Although Portuguese and English are ultimately related, the existence of 'library' in English is, of course, the result of borrowing from Romance languages. The specific shift in American Portuguese stems from English semantic influence, not direct borrowing. This illustrates still another complex cause of sound–meaning likeness—ordinary borrowing combined with semantic borrowing.

These examples are cited because of their relevance to the problem of subclassification of languages. Did we not know the history of -*st* through written records, this resemblance between English and German, not shared by other Germanic languages, might well be taken as evidence in favor of a common origin of English and German distinct from that of other Germanic languages. That is, we would consider this form as pointing toward a special grouping within Germanic which opposed English and German as descended from a distinct intermediate speech community to one or more other such intermediate communities as ancestors of the other Germanic languages. From this example, it is apparent that convergence among related languages is a different and more subtle problem than simple convergence among unrelated languages. Given the same starting point, we may expect that

similar quite specific results may ensue without historic contact. As in the establishment of genetic relationship, it is sound–meaning resemblances that count; but the entire weighting is different, which is what lends the problem its methodological interest. To cite another instance, if two languages have the form *mata*, 'eye', this is evidence which weighs positively in establishing genetic relationship. If, however, two different Malayo-Polynesian languages both have *mata*, 'eye', this same fact is of practically no moment as an indication of a special relationship between the two languages within Malayo-Polynesian. Here the well-nigh universal distribution of *mata* prevents it from functioning as evidence for any one particular form of subgrouping, while it is precisely this universal distribution which is a cogent indication of the common origin of the family as a whole.

The problem of genetic subgrouping is thus one which is methodologically distinct, though related, to that of the establishment of genetic relationship. It has given rise to far less general discussion. Subgroupings are often done in a casual manner, and differences of opinion in their regard are generally considered of relatively minor significance. Yet the specific historical relationship implied, being more recent in time, may be of far greater interest to the ethnologist and culture historian. The establishment of a large linguistic stock either with incorrect subgrouping or with no subgrouping at all can lead to serious errors of historical interpretation. Such a family as Hokan-Siouan, covering as it does large portions of North and Central America, can lead to a vast variety of conflicting interpretations unless accompanied by detailed and accurate subgrouping. The problem is thus in its way quite as important as the more frequently discussed one of genetic relationship, and it is often far more difficult of solution.

The relative ease or difficulty of the subgrouping problem rests ultimately on the ratios of several time spans. We now no longer ask whether A, B, and C are related. We ask the more subtle question, given the relationship of A, B, and C, is the distance between A and B equal to, or less than, the distance from A to C? If less, then AB forms a group as distinct from C. Our explanation is that the speech community ancestral to all three at one time split into a number of language communities, from one of which A and B have descended by further differentiation and from another of which C has arisen. There are, then, three points of time to consider, or more, if further and finer subgrouping can be carried out. Methodologically, it will be simpler to consider the minimum case of three time points.

Figure 1 indicates that five contemporary related languages—C_1, C_2, C_3, C_4, and C_5—group into two branches, one consisting of C_1 and C_2, the other of C_3, C_4, and C_5. Language B_1 is ancestral to C_1 and C_2; B_2 to C_3, C_4, and C_5,
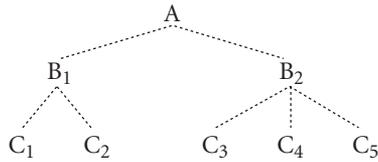
FIGURE 1.   Relative time spans of branches of a family.

while A is ancestral to the entire group. The three points of time are A, B, and C; the two intervals of time A–B and B–C. The comparative length of these two periods A–B and B–C has important consequences. If the earlier period A–B is very long compared to the subsequent period B–C, that is, if the ratio A–B/B–C is large, the problem is easy of solution. During the relatively long period A–B, many independent changes have occurred in the two branches, with little change in the ensuing period B–C to obliterate the results. In this situation, recognition of relationship among the languages of each branch often precedes that of the existence of the family as a whole. For example, the relationship among the Semitic languages was noted long before that of the larger Afroasiatic (Hamito-Semitic) family, which contains Semitic as a branch. We may presume that the groups of languages which at an earlier stage were recognized as independent families are valid branches of the larger family, since the differences among the branches are here so great that each was recognized as a separate entity before the family as a whole came to be accepted. However, grouping errors may arise even here, and the whole problem should be re-examined after the establishment of the larger family.

The opposite extreme is found where the ratio A–B/B–C is small. The earlier period here is short, with the consequent opportunity for only a small number of changes, while the subsequent period is long, allowing for the obliteration of these changes. These are the difficult cases in which, as with the Bantu or Malayo-Polynesian languages, relationship of the languages to one another as a whole was early recognized but where even at present [1957-Ed.] there is no satisfactory subgrouping.

The problem of subgrouping, then, is the recognition of the existence of a set of changes common to a particular subgroup which has occurred between the period of divergences of the family as a whole and that of the subgroup in question (the time periods A–B1 and A–B2 of Figure 1). It is a dynamic problem of the detection of changes. Even when phrased as though it employed criteria based on the synchronic sharing of features, a historical analysis is implied. Thus, if reference is made to an item of vocabulary found in certain related languages but not in others, this static phraseology conceals

the fact that it is the process of replacement of one item of vocabulary by another which is decisive. As has long been seen, the essential factor is shared innovations, since shared retentions can always occur independently without a common period of development. Shared obsolescences can be of some significance when connected with a functional replacement. If two languages replace an earlier word for 'nose' with some other term and also lose the traditional term, this double agreement is of significance for judging the existence of a common historical period.

The detection of such common innovations and obsolescences encounters special difficulties; since the languages all have a common starting point, the chance of convergence—the process of drift mentioned in the earlier part of this chapter—is very great. The problem of borrowing is also multiplied, in that closely allied languages are certainly more likely to borrow even fundamental vocabulary and grammatical affixes than more distant forms of speech. In dealing with the task of subgrouping, we have not four but five causes of sound–meaning resemblances to take into account: chance, symbolism, borrowing, genetic inheritance from the common period of the entire family, and genetic inheritance from the period of common development of the language ancestral to the subbranch. In distinguishing these last two, a typical danger of circular reasoning must be surmounted. Since occurrence in at least two separate branches of a family is the common reason for assigning a feature to the ancestral language of the family as a whole, a resemblance between two languages can be assigned to this early period if the two languages are classified in separate branches. In this case the feature is judged to be a retention and not indicative of a special relation between the two languages. On the other hand, we can consider the resemblance as evidence that the two languages belong to the same branch of the family. In this case the feature is judged to be a common innovation and not to be part of the protolanguage of the family as a whole.

One solution to this problem is the bringing to bear of evidence from languages more distantly related to the family as a whole, where such exist. The entire problem then remains, in a sense, one of subgrouping, but on a wider scale. For example, in Bantu languages, a few terms for parts of the human body, 'ear', 'arm', and 'armpit', are found commonly with the prefix *ku-* and just as commonly with *li-*. Proto-Bantu is generally reconstructed as having the *li-* prefix for the words 'ear', 'arm', and 'armpit'. The argument for *ku-* is, I believe, much stronger. An analogical change from *ku-* to *li-* is easily understandable, since *ku-* is not otherwise used with ordinary nouns, being typically an infinitive and locative prefix, whereas *li-* is very common and includes many other terms for parts of the human body. The motive for a

change from *li-* to *ku-* is hard to discover. Moreover, a few languages have *liku-*, but none have *kuli-*. Here the *ku-* was not understood as a prefix, and *li-* was then prefixed to the whole form analogically. Aside from any judgment of the internal Bantu evidence, however, this question is decided in favor of *ku-* by its frequent appearance in the most widely scattered branches of the vast Niger-Congo family of which Bantu is but a subbranch, contrasted with the non-occurrence of *li-* outside Bantu. Again, there are variant Bantu forms for 'two', *bali* and *bili*, both about equally widespread. We might make a rather uncertain choice in favor of *bali* as earlier, on the assumption that *bali* > *bili* might occur as a sporadic instance of vowel harmony. In fact, the evidence from other Niger-Congo languages is again decisive and in favor of *bali*. The form *bili* is never found anywhere outside Bantu. This, incidentally, allows us to judge *bili* as an innovation, the sharing of which is evidence of the existence of a separate subbranch within Bantu, while the occurrence of *bali*, a retention, has no such implications. These and many other examples which could be cited are of interest because they refute the common belief that more remote relationships should be ignored, while each distinct subbranch is reconstructed separately and independently. In many cases we cannot choose between alternative reconstructions without taking the wider family into account. In fact, in Indo-European the reconstruction of the protolanguage of the family as a whole progressed far more rapidly than did that of the individual branches. Moreover, Proto-Germanic and other comparable intermediate unities have always been reconstructed with one eye backward to Proto-Indo-European and one eye forward to the contemporary Germanic languages.

Let us now consider some of the types of linguistic change with a view to their value in determining subgroupings. Starting with phonology, regular sound changes can be seen to be generally of little value in this regard. The possible number of changes is small, and the probability of convergences high. Thus in Indo-European the voiced aspirates have, without doubt, become unaspirated stops a number of times independently. A conditioned change, i.e. one in which one phoneme has changed to another only under certain stated conditions, is of somewhat greater value, but any single one of these also easily results from convergent developments.

The sharing of a whole series of changes is of greater cogency. For example, the Malayo-Polynesian languages seem to fall into two main subgroups, a western and an eastern. Among other features, the sharing of a whole series of phoneme mergers by the eastern languages is certainly an important item of evidence. Certain western languages, for example, Malagasy, have independently carried out some of the changes, but not all of them.

What appears at first glance as an impressive series of shared phonemic changes may, on closer inspection, turn out not to be at all decisive. A good many of the changes stated by Grimm in his first law, referring to the transition from Proto-Indo-European to Proto-Germanic, also occur independently in Armenian. But the whole series of changes resolves itself into a few interconnected habit changes which have also occurred in other nonrelated languages. The changes $b > p$, $d > t$, and $g > k$ are all the consequences of a single change in the habit of articulation from voiced to unvoiced sounds. The changes $p > f$, $t > þ$, and $k > x$, which are also included in the statement of Grimm's law, likewise involve a single change of habit from stop to fricative articulation. Moreover, as Martinet has pointed out, such changes are interconnected as the result of a general tendency to greater or lesser vigor of articulation at certain historic periods. Indeed, this whole set of changes recurs in Angas of the Chad branch of the Afroasiatic languages and elsewhere.

Sporadic changes, such as individual assimilations and dissimilations, are of greater moment, in that they are less likely to happen independently. However, such common tendencies as the dissimilation of one of two *l* sounds to an *r* can certainly occur convergently. Still, the number of possibilities is greater here, since each is an independent case. The sporadic changes $n > r$ in Aramaic *bar*, 'son', and $l > r$ in *tarten*, 'three' (cf. Hebrew *ben* and *šɔloš* respectively), are practically certain indications of membership of a dialect in the Aramaic rather than the Canaanite branch of Northwest Semitic. To sum up, unconditioned changes, if large in number, shared conditioned changes, and sporadic sound changes are all evidence of subgrouping. A single unconditioned change is of practically no value at all.

In the area of morphological change, the most important single process is analogy. In general, shared analogies are of little help for the problem under consideration, since the pressure of more frequent on less frequent patterns is everywhere great and likely to lead to similar changes. On the other hand, sharing of a highly irregular alternation, which, as we have seen, is of great weight indeed as an indication of genetic relationship, is useless in the present instance. A highly irregular formation which has withstood analogy must be very old. It is a common retention, not an innovation, and therefore irrelevant for such grouping. Thus we arrive at the somewhat discouraging conclusion that both analogical change, if it follows the dominant pattern, and absence of analogical change are equally indecisive as indications of subgroupings. An illustrative example is furnished by the common Indo-European demonstrative and third person pronoun. In the nominative singular masculine, some languages have reflexes of an original *so (e.g. Greek

ho, 'the'); other languages of an original **sos* (e.g. Sanskrit *sas*). The masculine singular nominative without the usual *-s* is an anomaly. There is no analogical pattern that could lead to the loss of *-s* independently in a number of instances, whence Indo-Europeanists unanimously reconstruct **so* as Proto-Indo-European. The agreement of two languages in retaining reflexes of **so* is therefore merely a common retention and no evidence for a separate group. On the other hand, the pattern of a nominative masculine singular in *-s* is predominant in other forms, so that the addition of *-s* is a natural analogic development which can easily occur in separate instances.

The following example of a rare analogical pattern is much less likely to be the result of convergence. The perfect *y-t-n*, 'give', shared by Phoenician and Ugaritic is a quite strong argument for the affiliation of Ugaritic with the Canaanite branch of Semitic to which Phoenician belongs. Elsewhere in Semitic the perfect is *n-t-n*. It was probably the possession of a common imperative and imperfect formation without initial consonant in both verbs with initial *n* and *y* that provided the model: *šb:tn* (imperatives) = *yšb:tn* (perfect). Since both initial *n-* and initial *y-* verbs are uncommon and since the shift of membership has occurred only in this single verb and in the same direction—from the *n-* to the *y-* class—in both Ugaritic and Canaanite, it is highly unlikely to have occurred independently in the two cases. In view of the geographical proximity of the two languages, borrowing as an alternative explanation is much more likely than convergence if this irregularity is to be rejected as evidence of the Canaanite affiliation of Ugaritic.

A morphological construction, for example, a periphrasis of verb root and auxiliary to form a tense involving the existing elements not hitherto combined or, better still, at least one element which is uncommon elsewhere in the family, is powerful evidence for subgrouping. The common possession of a future in *-b* by Italic and Celtic languages is a case in point. Such a combination is unlikely to have occurred independently, although even this is not impossible; one of the common theories derives this formation from a periphrasis with **bhu*, 'to become'. The existence of another, more common future in *-s* elsewhere in Indo-European suggests that the *-b* future is an innovation. In other words, where a replacement has occurred not involving some common analogical formation, the possibility of convergence becomes small. That is, this would be so, were we sure that there was functional replacement. Proto-Indo-European might have had two futures with different semantic functions of which only one survived in any given language. The danger of circular reasoning again rears its head. The appearance of *-b* in Italic and Celtic can be evaluated as evidence that the formation is Proto-Indo-European if an Italo-Celtic branch is not accepted.

Lexical innovations are of great value because convergence is practically ruled out. That, for example, two Indo-European languages would independently make up a new verb 'to take' with the form *nem* is extremely unlikely. Though convergence is thus virtually ruled out, borrowing becomes an important alternative explanation, since it is among lexical items that dialect borrowing is most frequent. In detecting lexical innovations, the existence of as complete materials as possible for each language concerned becomes of major importance. If the material is very limited, it may appear that a particular group of languages shared a common term not found elsewhere. Fuller evidence may disclose the existence of a cognate with slightly different meaning in some other language in the family. What appears to be a lexical innovation thus becomes merely a semantic shift in at least one language or group of languages. While this is also evidence, it is far less convincing than a complete lexical innovation.

This brings us to the topic of semantic change. It is obvious that meaning changes are strongly subject to convergence. Moreover, it is often difficult to know what is retention and what is innovation, for a semantic shift which takes place in one direction can often just as easily occur in reverse fashion. A term for 'day' often becomes 'sun', but likewise a term that means 'sun' frequently comes to mean 'day'.

From this review of some of the more common types of linguistic changes, it will be seen that there is hardly a feature shared by certain related languages and not others for which convergence or borrowing is absolutely excluded as an explanation. Nevertheless, when in even more difficult cases the evidence is examined closely and in the light of general comparative reconstruction of the linguistic history of the entire family, certain groupings will normally emerge. Although, as has been noted, no single resemblance is ever completely decisive, it will be found that certain languages share with one another a far larger number of features which may be innovations than they do with related languages outside the subgroup and that among these are some of those which are least likely to be the result of convergence, including shared sound shifts *en masse*, sporadic sound changes, new morphological formations, shared analogical shifts, including some of the rarer ones, and true lexical innovations. It is the sheer number of such resemblances, together with the inclusion of some of the types most likely to be innovations, that excludes convergence or borrowing as an overall explanation.

The present chapter outlines the type of evidence to be considered in arriving at subgroupings. It by no means follows that such evidence will always be found. The reason may be that, while such subgroupings do exist,

the period of common development of each subgroup is short and its period of subsequent differentiation long, as has been noted. In difficult cases, such as this, the lack of sufficient first-rate descriptive grammatical and lexical material for the languages and of adequate linguistic reconstruction prevents us from assembling and judging adequately the evidence for the groupings which are actually present. In other cases no grouping may exist. There is no *a priori* reason for denying this. If linguistic innovations in a speech community spread in a random way, such as that envisaged by the classical wave theory, then there would be gradual transitions only and no sharp breaks anywhere to give to distinct groupings. Wherever migrations or the intrusions of peoples speaking nonrelated languages occur, innovations encounter a barrier which cannot regularly be overcome, and sharp groupings inevitably result. This has often occurred, e.g. the separation of Rumanian from the rest of the Romance speech community. With or without actual movements of people of this sort, language families often display clear divisions into branches. The pure wave model therefore cannot be universally correct. Although, as has been seen, absence of subgrouping is a possibility, it seems more likely that the nonlinguistic forces which produce differentiation into separate tribes, political states, and economic regions must have a profound effect on spoken communication, producing weakness in lines of communication which gives rise to dialects and eventually separate languages and language subgroups as the process continues.

The problem of discovering subgroupings is, in the more difficult cases, quite arduous, far more so than the discovery of genetic relationships. The principles discussed here are not essentially new. They may be found, for example, in the classic treatment of Brugmann (1883). I do not believe that there are any short cuts. Recently glottochronological methods (see for example Elbert 1953) have been used. No doubt in less difficult cases this will lead to accurate results. However, the mere counting of the number of cognates shared, without attention to morphological or phonologic evidence and without consideration of the general distribution of each form for its bearing on the question of innovation, is a relatively crude method which disregards much relevant evidence. If, for example, we were using the method of glottochronology to group the Indo-European languages, under the number 'four' we would have, among other entries, English *four*, German *vier*, Danish *fir*, Italian *quattro*, Spanish *cuatro*, French *quatre*. Since these are all cognate, we would simply score this as a single agreement among all the languages concerned, and it would contribute no information toward the problem of subgrouping. Yet English, German, and Danish here share innovations in the form of the word, loss of *t*, $qu > f$, etc., which Italian

Spanish, and French do not. Hence valuable relevant features are being overlooked.

This criticism of the adequacy of glottochronological methods for this problem is not meant to detract from its possible value as an approximate quantitative measure of the periods of time involved, once the subgrouping problem has been solved by conventional methods.

In cases of obvious subgrouping, the correct results will be very quickly evident from comparative vocabulary inspection, as described in the previous chapter on genetic relationship. This method will show lexical innovations, as well as some new morphological combinations, widespread sound changes, and sporadic shifts. In more difficult cases, vocabulary inspection should also furnish an answer, but only after meticulous examination of the distribution of each form and of the relevant phonological and semantic factors. However, the examination of morphology, if the languages have a complex morphological system, and considerations from the development of the sound system of each language based on reconstruction will in such cases often lead to more rapid results. Insofar as the data and state of historical knowledge allow, all types of evidence should be considered. The results in all these domains will necessarily agree if the evidence examined is relevant and due weight is accorded each item.

References

Brugmann, Karl. 1883. 'Zur Frage nach den Verwandschaftsverhältnissen der idg. Sprachen', *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1: 225–56.

Elbert, Samuel H. 1953. 'Internal relationships of Polynesian languages and dialects', *Southwestern Journal of Anthropology* 9: 147–73.

This page intentionally left blank

4

The general classification of Central and South American languages*

The classification of Central and South American languages set forth in the appendix [§4.1-Ed.] to this paper is provisional in some respects. The eight families listed are to be considered branches of a more inclusive stock which probably includes also all the remaining American languages except Na-Dene and Eskimauan. Among the groups listed here only Otomanguean, which has not yet been thoroughly investigated by the writer, is considered at all not likely to belong to this great family.[†] The problem of the classification of American languages becomes, therefore, for the most part a vast problem in the subgrouping of this single dominant family. It follows that, if this view proves correct, it will not be sufficient merely to show that a certain set of languages are related in order to prove that they form a stock. We must demonstrate that all of the languages within the group are more closely related to each other than any are to any languages outside the group. We distinguish, therefore, between the establishment of a valid relationship and that of a valid genetic group at whatever level. For example, Swedish, Albanian, and Hindi are all related but they do not form a valid genetic group, and culture-historical conclusions drawn from treating them in this manner would be highly misleading.

If the thesis presented here is correct it will be possible to show that certain apparently contradictory theses concerning genetic relationships are not, in fact, incompatible, and we may thus avoid useless controversies. For example, Freeland and Sapir considered certain languages in Mexico to be Penutian; Whorf and McQuown added others, notably Mayan, in a family which

[†] Greenberg (1987) retains Otomanguean as a subgroup of Amerind [Ed.].

likewise includes Azteco-Tanoan. It is my opinion that these languages are related to both California Penutian and to the Azteco-Tanoan languages but more closely to the former. The problem thus becomes one of subgrouping within the larger stock. This situation obtained to such a degree in South America that the writer considers that, whereas the groupings designated here by capital letters, e.g. Paezan, Andean, and Macro-Ge, are for the most part certain as valid genetic groupings, it is by no means inconceivable, though unlikely, that, for example, Andean might turn out to be closer to Macro-Ge in III than to Equatorial in II, or that Paezan is closer to Andean than to Chibcha proper, although this alternative was carefully considered and rejected. A degree of uncertainty in subgrouping exists also at the lowest level. Thus I am by no means entirely certain that Bororo might not be considered a Ge language proper rather than as having a separate status within Macro-Ge. This should not be too suprising. After a century or more of comparative Indo-European studies the existence of an Italo-Celtic grouping is still being debated.

Of the eight groups outlined here, the greatest uncertainty exists in the case of the two new vast groupings in South America, Andean-Equatorial, and Ge-Pano-Carib. It should be emphasized that the doubt does not pertain to the relationship among all of these languages but rather to the correctness of these two assemblages of languages as valid genetic groupings in the sense described above.[‡]

The time at my disposal is, of course, too brief to discuss the methodological problems with any degree of thoroughness. These will be treated fully in a projected series of articles in which it is expected that the classification will be set forth in detail and each family treated separately.[§] A few important methodological considerations may, however, be pointed out. Only those resemblances which involve both sound and meaning simultaneously are considered relevant for historical connections. When the morphemes involved are roots this is called lexical comparison, when they are affixes, grammatical. There is no contradiction in the results attained by lexical and grammatical comparison and both methods are employed as far as possible. For purely practical reasons it is easier and more fruitful to begin with the comparison of basic vocabulary items.

[‡] Greenberg (1987) retains Ge-Pano-Carib but seperates Andean from Equatorial-Tucanoan (all three are grouped together in the appendix to this chapter). Also, Greenberg (1987) alters a number of further subgroupings [Ed.].

[§] These articles never appeared; eventually, the entire proposed Amerind family was presented in Greenberg (1987) [Ed.].

A second important consideration is that comparison should be on the widest possible scale and utilize data from all the languages for which material exists. It might be thought that surer results would be attained by comparing only a small number of languages based on hypotheses of very limited scope. It will not be possible to discuss here all the reasons which indicate that this is not the case, but a few relevant factors may be indicated. It is a basic principle of comparative linguistics that a form is an inherited one if it reappears in languages of other subgroups of the same stock, while there is no prima facie case for this if it does not. This principle is of primary importance in eliminating purely accidental resemblances between two languages. Moreover, by considering all the more closely related languages on either side, the form can also be tested to see if the tentatively reconstructed forms for both groups are similar. The comparative method has always been applied to groups of languages rather than languages in isolation.

If we choose a small number of languages and compare them on a narrow basis not only are our results less reliable, but even when they are correct they are less significant and even misleading to archaeologists and ethnologists because they are unlikely to be valid genetic units in the sense described earlier. Finally it may be pointed out that the evidential bearing of valid etymologies depends on their distribution. Thus, on an isolated hypothesis connecting Panoan with the distant Totonac, we might note the striking resemblance of Totonac *makan* and Panoan *meken*, both meaning 'hand'. However, we might interpret this form, which is actually found in a number of major stocks, as evidence for a special relationship between Totonac and Panoan, if we do not consider it in terms of the total relevant distribution.

With these methodological factors in mind, a brief description of the actual procedures will be given. The first was in the nature of a preliminary survey designed to provide initial hypotheses concerning groupings in South America. This consisted in the compilation of about forty vocabulary items which experience had shown to be among the most stable, for approximately forty languages. Among these languages were a number of the Arawak, Tupi, and Carib languages on the assumption that larger groups such as these provide greater depth for comparative purposes than single isolated languages or small language stocks. From this first comparison a number of groups emerged, notably the Andean, Macro-Ge, Macro-Panoan, and Equatorial. Each of these was then assigned a separate notebook and a list of over three hundred words was compiled for each language, insofar as they

were obtainable. New languages were compared with each of these groups with emphasis not on stray resemblances with single languages but on the occurrence of forms derivable by known types of sound changes from the tentatively reconstructed originals based on recurrent forms. Where languages did not fit into any of the established groups, they were assigned to new separate groups and entered in other notebooks. As new languages were examined, adding to the precision with which the basic lexical fund of each group was known, it became evident that certain of these groups, for example Macro-Ge and Macro-Panoan, were particularly close to each other and formed parts of still larger groupings. Material from over 250 languages was entered into notebooks at this stage.

The third procedure involves the use of index cards, each assigned to one semantic sphere, e.g. sun and day. Utilizing the material in the notebooks, entries are made in separate paragraphs for each probable etymology involving any languages in the area covered. It is thus possible to see at a glance the distribution of each set of probably related forms. This part of the work is still in its initial stages. When it is completed it will be possible to marshal the evidence in systematic form to solve the more difficult problems of [sub]grouping.

A fourth procedure has been the compilation of grammatical information for approximately thirty languages. The information in this area is, of course, less extensive in the existing literature than the lexical. What has been recorded thus far does not contradict conclusions based on lexicon, and in some cases adds striking confirmation. Thus a common system of singular pronominal prefixes seems to run through the entire Ge-Pano-Carib group and includes an irregular alternation in the third person.

It is hoped that other linguists will independently try the method of mass comparison suggested here in order to test the writer's conclusions. The ultimate test is a pragmatic one. Those parts of Sapir's scheme such as Penutian and Na-Dene which are valid have proved fruitful in that workers have been able to carry on more advanced comparative investigation within their framework. On the other hand, various suggestions regarding Central America contained in Sapir's *Encyclopedia Britannica* article have, in general, been ignored. My own methods indicate that they are, for the most part, incorrect. In the same fashion, if the present classification is correct, it will prove its usefulness in future more advanced comparative investigations, and arguments raised against it will be disregarded. By the same token it cannot be saved by the most ingenious argumentation if it fails the crucial test of practice.

4.1 Appendix: Tentative linguistic classification of Central and South America

   I. Macro-Chibchan
      A. Chibchan proper
          1. Chibcha-Duit, Tunebo group, Aruaco group, Cuna-Cueva, Guaymi-Dorasque, Talamanca group, Rama-Guatuso
          2. Misumalpan, Paya, Xinca, Lenca
          3. Shiriana
      B. Paezan
      Choco, Cuaiquer, Andaki, Paez-Coconuco, Colorado-Cayapa, Warrau, Mura-Matanawi, Jirajira, Yunca, Atacameno, Itonama
  II. Andean-Equatorial
      A. Andean
          1. Ona, Yahgan, Alakuluf, Tehuelche, Puelche, Araucanian
          2. Quechua, Aymara
          3. Zaparoan (including Omurano, Sabela), Cahuapana
          4. Leco, Sec, Culle, Xibito-Cholon, Catacao, Colan
          5. Simacu
      B. Jibaro-Kandoshi, Esmeralda, Cofan, Yaruro
      C. Macro-Tucanoan
          1. Tucano (including Auixira), Catuquina, Ticuna, Muniche, Auaque, Caliana, Macu, Yuri, Canichana, Mobima
          2. Puinave
      D. Equatorial
      Arawak (including Chapacura-Uanhaman, Chamicuro, Apolista, Amuesha, Araua, Uru), Tupi (including Ariqueme), Timote, Cariri, Zamuco, Guahibo-Pamigua, Saliban, Otomaco-Taparita, Mocoa, Tuyuneri, Yurucare, Trumai, Cayuvava
 III. Ge-Pano-Carib
      A. 1. Macro-Ge: Ge, Caingang, Camacan, Machacali, Puri, Patacho, Malali, Coropo, Botocudo, Chiquita, Guato, Fulnio, Oti (prob.)
          2. Bororo
          3. Caraja
      B. Macro-Panoan
      Tacana-Pano, Moseten, Mataco, Lule, Vilela, Mascoy, Charrua, Guaycuru-Opaie
      C. Nambicuara
      D. Huarpe
      E. Macro-Clarib
      Carib (including Pimenteira and Palmella), Peban, Witotoan, Cucura (prob.)
      F. Taruma
  IV. Oto-Mangue

    V. Tarascan
   VI. Hokan including Jicaque, Yurumangui (prob.)
  VII. Penutian including a Mexican branch Mixe-Zoque, Huave, Mayan, Totonac
VIII. Azteco-Tanoan

Reference

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford: Stanford University Press.

5

The methodology of language classification*

The present volume [Greenberg (1963)-Ed.] contains a complete genetic classification of the languages of Africa. Since the results differ in important respects from previously current schemes of classification, a brief methodological foreword seems in order.

There are three fundamentals of method underlying the present classification. The first of these is the sole relevance in comparison of resemblances involving both sound and meaning in specific forms. Resemblances in sound only, for example the presence of a tonal system as such, or in meaning only, as in the existence of morphemes (meaningful) forms indicating sex gender but without phonetic similarity, are irrelevant. The second principle is that of mass comparison as against isolated comparisons between pairs of languages. The third is the principle that only linguistic evidence is relevant in drawing conclusions about classification. This last is so self-evident when stated that it would seem unnecessary. In fact, disregard of this principle is very common and a subtle source of errors in classification in Africa and elsewhere as will be shown later.

The emphasis on sound–meaning resemblances does not imply in principle exclusive reliance on either root morphemes (vocabulary) or affix morphemes (grammar). Related languages, of course, show resemblances both in vocabulary and grammar and contradictions no more appear in Africa than they do in Europe or the Near East, provided only that irrelevant resemblances in form or meaning only are excluded.

An example of the use of sound only or meaning only is the method of Meinhof which is primarily typological with evolutionary overtones. One sample of Meinhof's method is perhaps worth more than pages of abstract analysis. In treating the extraordinarily diverse languages of the southern hills of Kordofan west of the Nile, Meinhof paid no attention to any factor other than the presence or absence of noun prefixes. If a language had noun

prefixes, he called it pre-Hamitic, which to him was a stage between the nonclassifying, nongender, isolating Sudanic and the inflective Hamitic with its system of sex gender. If it did not have a system of noun prefixes, it was Sudanic.

Three among the languages of Kordofan—Tagoi, Tumale, and Tegali—are very similar. Meinhof classified Tagoi and Tumale as pre-Hamitic, but Tegali he assigned to Sudanic simply because it did not have noun prefixes.

A comparison of vocabulary shows that numerals [sc. of Tegali] completely agree with those of Tumele. Moreover they are for the most part identical with the Tagoy numerals. Besides, a number of word stems and some verb forms of Tegele are identical with those of Tagoy and Tumele. But the grammatical structure of the noun indicates that Tegele is a Sudanic language because noun classification is absent while Tagoy and Tumele have clear noun classes. Apparently there has been a mixture of two diverse elements. (Meinhof 1916: 110).

Here, as elsewhere, there is the mechanical application of structural criteria, or even a single criterion. There is no recognition of the living realities of language change, no understanding that prefixes are not essential qualities inherent in an unchangeable species. Characteristic also is the facile assumption of mixture where typological criteria lead to a contradiction. The MacDiarmids, observers with a minimum of formal linguistic training, group the languages of Kordofan, with essential correctness, into eight stocks. They naively classify Tegali along with Tegoi and Tumale (MacDiarmid and MacDiarmid 1931).

The importance of mass comparison as opposed to isolated comparisons between pairs of languages has become clear to me as a result of certain questions of a general nature raised by a number of critics. Basically these criticisms come to two: a doubt as to the feasibility of drawing genetic conclusions of wide scope and the belief that the present classification is arbitrary in the sense that just as good evidence might be assembled for other classifications cutting across the one presented here. The former objection, which appears to be widespread among linguists, is understandable as a reaction against the previous catch-all use of such terms as Hamitic and Sudanic. It should be realized, however, that no concrete evidence of the kind which documents this work was ever assembled for the total assumed ranges of these language stocks. When resemblances can be assembled which are recurrent in many languages, which extend over vast and widely separated geographical areas and which encompass elements with morphological functions, pronouns, and the most stable parts of the vocabulary (most stable on *a priori* grounds and on the basis of our knowledge of areas with written records), then common origin is the only adequate explanatory hypothesis.

The importance of resemblances recurrent in a large number of languages as plausible outcomes of some single ancestral form as elicited by mass comparison is of very great evidential power in excluding either chance or borrowing as explanations. Considerations derived from the elementary theory of probability help to make this explicit. In fact, I had no such considerations in mind when I first undertook the present classification, since the importance of widespread specific resemblances was evident even when not formulated in these terms.

Let us assume even that accidental resemblances between two languages can be rather high, say 20 per cent. The chance that some single meaningful form will appear with similar sound and meaning is then 1/5. The chance that this same element will appear also in some third languages is the square of 1/5, that is 1/25. In general, given n languages the chance that a resemblance will occur in all of them will be $(1/5^{n-1})$. It is easy to see that this ratio rapidly becomes infinitesimal with the increase in n, in other words, resemblances running accidentally through a large number of languages will occur only with the utmost rarity.

Many linguists acquainted with only two or three languages of a family are thus not in a position to assess the importance of the resemblances they note because of this neglect of mass comparison. Much of the relevant evidence for relationship will not appear in such a procedure. Suppose one were to compare English with Hindustani only. The tendency of these resemblances to recur in other languages of the Indo-European family would not be realized. If they were not related, very few such recurrences would be found. Moreover, many facts about English which point to its being Indo-European would not be taken into account if these traits happened to be among those lost in Hindustani and vice versa. For example, a comparison between English and Russian would reveal many resemblances not found in Hindustani while a comparison of Hindustani with Russian would reveal many resemblances not found in English. But any facts which showed that English belonged to some larger group or that Hindustani belonged to the same larger group is relevant to the common origin of English and Hindustani. Each additional language thus brought into the comparisons adds further connecting links tending to establish the validity of the structure as a whole.

In regard to the second problem, the supposed arbitrariness of such a classification, the best test is the pragmatic one of actual performance. If someone will produce a comparable set of proposed etymologies and morphological elements cutting across the families found here and having the properties of widespread distribution, semantic plausibility, and appearance of regular phonologic development from a common original, the arbitrariness will have

been proved. I do not believe that this is any more possible than to demonstrate that Finnish, Basque, and German belong to a single family, while Hungarian, Greek, and Hebrew belong to another.

The most effective manner of applying mass comparison as a method is a table of pronouns, grammatical elements, and vocabulary for the languages involved. Such a table when prepared for European languages leads to the universally accepted classifications regarding relationship, even with a very small number of items. The following table has been prepared for a very small number of African languages and a very few items of vocabulary. The forms I consider to show likelihood of common origin are indicated by similar symbols.

	one	*two*	*three*	*hand*	*eye*	*ear*	*mouth*
1.	saŋ	● su	soti	yuŋ	● siŋ	–	● a
2.	wate	■ iba	● tati	ju(le)	no(do)	● to(go)	yabodo
3.	● toro	● ču	agozo	● daho	● samo	■ sumo	■ či
4.	ili	■ iwa	● ita	ilɔ	ewu	● ɔtɔ	◆ ɛnu
5.	mwe	bali	● tato	■ (li)-to	(le)-iso	● (ku)-toi	◆ (mu)-nywa
6.	● tilo	ndi	yasko	kela	● sim	■ sumo	■ či
7.	kiet	■ iba	● ita	■ ɛte	enyin	● utoŋ	◆ inua
8.	lakoi	● swe	we	● taha	i	kebbe	● a

Grammatical prefixes and suffixes are enclosed in parentheses since what we are comparing are morphemes, minimal units having a meaning. Even the first three words lead to a separation of these languages into two groups: I—1, 3, 6, 8; II—2, 4, 5, 7, which is confirmed by the other words. The two families are I. Saharan, II. Niger-Congo. The languages are 1. Berti, 3. Teda, 6. Kanuri, 8. Zaghawa, 2. Kotopo (Adamawa), 4. Ahlõ (Togo), 5. Proto-Bantu, 7. Efik. The addition of several hundred Niger-Congo languages and many additional words or grammatical forms would lend continuous further evidence for this same division.

The final principle seems the most obvious of all, namely that languages should be classified on linguistic evidence alone. All that counts is the degree and types of similarity in linguistic forms. The most common source of error in this regard is the fact that prominence of a language or group of languages because of practical importance, extent of population and territory, or literary cultivation tends to lead to separate status in classification. Two important instances of these in African classification is the independent position assigned to Bantu in previous classifications and the special status accorded Semitic in the Hamito-Semitic (Afroasiatic) family as against the remaining four

branches which were generally lumped together in a pseudo-entity Hamitic. It would be well if languages could be compared without a knowledge of such extraneous facts so that only the degree and type of linguistic resemblance would be taken into account as in the above table of comparison of Niger-Congo and Saharan in which languages are referred to by number.

References

MacDiarmid, P. A. and D. N. 1931. 'The languages of the Naba Mountains'. *Sudan Notes and Records* 14: 15–48.

Meinhof, C. 1916. 'Sprachstudien im ägyptischen Sudan'. *Zeitschrift für Kolonial-sprachen* 7: 36–133.

This page intentionally left blank

Part II

Classification, sound correspondences, and reconstruction

This page intentionally left blank

6

The principles of genetic linguistic classification*

My purpose in this chapter is to discuss genetic classification; but I hope that the discussion will also help to explain and justify a deviation from what has become virtually a compulsory practice among American Indianists: the use of sound correspondence tables and asterisked reconstructed forms.

In proceeding in this manner, this volume will resemble my first published classificational work, namely, that on African languages (Greenberg 1963). In that study I did not use a single asterisk or a single table representing a reconstructed sound system; and although I made occasional reference to particularly striking sound correspondences, these figured in no essential way as part of my method. There were, however, extensive lists of proposed etymologies, both lexical and grammatical, and shared grammatical irregularities. Yet it is reasonable to assert that this classification has won general acceptance and has become the basis for a considerable body of comparative work on African languages.

These aspects of my methodology earned me a fair share of criticisms, of course, even from those who accepted and built on my results in their own investigations. The following are representative quotations. The first are the views of William Welmers (1973: 5, 6, 15, respectively):

Greenberg has not, to be sure, demonstrated the existence of regular sound correspondences among all of the languages in any of the four language families he posits for Africa, although it has already been implied that such correspondences are the only real proof of genetic relationship. In fact, evidence that falls short of clear demonstration of regular phonetic correspondences may nevertheless be overwhelming. . . . But the nature of the similar forms with similar meanings which Greenberg cites, and the number of them, is such that the fact of genetic relationship can be considered established. . . . For all practical purposes the validity of the four families can be considered established.

Several years earlier, at a conference held at Aix, Luc Bouquiaux (1967: 156) made the following statement:

I do not assert that I accept the totality of his conclusions, but for the languages of the Jos Plateau in Nigeria which I know, my studies have in every instance confirmed the classification he has proposed. It is possible in fact that his method may not be absolutely correct in regard to the regularity of sound correspondences, but I cannot but pay tribute to his intuition, which was later verified in every instance, although he often had at his disposal materials of very unequal value.

In a general review of the state of African linguistics, Paul Schachter (1971: 34) virtually stumbles on the correct solution in the following statement. (My italics indicate the decisive point.)

Certainly much more work of the kind begun by [J. M.] Stewart will be needed before the same regularity of correspondences as that found within the Bantu family can be claimed for Niger-Congo as a whole, or for that matter for any of its branches, none of which has to date been accorded the kind of scholarly scrutiny with which the Bantu languages have been favored. In the meantime, however, it seems appropriate to ask what conclusions *other* than genetic relationship between Bantu and West African languages can be drawn from an objective examination of the data cited by Greenberg and his supporters; e.g. Greenberg's extensive lists of strikingly similar forms, with shared meanings, attested over the entire Niger-Congo areas, or the detailed morphophonemic similarities noted by Welmers.

But in an empirical science, how much more can be reasonably required than that the evidence be 'overwhelming' and 'the fact of genetic relationship [be] established' (Welmers), or that there be no 'other conclusion than genetic relationship' (Schachter) or that 'the classification ... proposed' be confirmed 'in every instance' (Bouquiaux)?

Welmers's mention of demonstration, a term traditionally associated with Euclidean geometry, is appropriate in mathematics and logic, which were once described as consisting of 'surprising tautologies'. The notion that regular sound correspondences can fittingly be called demonstrative in this sense, although this and similar terms have often been used, will be shown in the course of this chapter to be illusory. As we shall see, what is in question is not just the nature of the truth exhibited by sound correspondences, but the still more basic question of what is *meant* by these and similar expressions, which are often used by linguists as though their meanings were self-evident.

There are indications that some investigators in the field of African languages have begun to realize that my work not only produced certain specific results, but also employed a revolution in methodology, as Edgar Gregersen (1977: 5), for example, has noted. Having made this point, Gregersen then quotes approvingly the following statement by Paul Newman (1970),

a Chadic specialist: 'The proof of genetic relationship does not depend on the demonstration of historical sound laws. Rather the discovery of sound laws and the reconstruction of linguistic history normally emerge from the careful comparison of languages already presumed to be related.' Actually what is involved is not so much a revolution as a return, in certain essential respects, to an earlier point of view, as will be noted later in this chapter. In fact the Neo-Grammarians of the late nineteenth century, the very school that proclaimed the regularity of sound change as their central doctrine, never made the claims for it that have grown up since, and that have been accepted by many linguists as a virtually indisputable dogma—though never, I would add, either stated with clarity or reasonably proved, but simply taken for granted as axiomatic.

In discussing this doctrine critically, let me say at the outset, if it is not obvious, that my remarks are not intended as an attack on the validity of comparative linguistics or on the importance of undertaking reconstruction. Rather, the discussion is meant constructively as a way of taking first steps where the comparative method has not been applied for want of an assured basis in valid genetic classification.

I do not wish to claim that all of the points I shall raise in this discussion are original with me. Several have been made by others, and in the pages below their work will be duly noted. However, as far as I can see, the only persons who have thought along basically similar lines are Sydney Lamb, Aaron Dolgopolsky, and, to a lesser extent, the famous anthropologist Alfred Kroeber and his co-worker Roland Dixon.

Basically, the wrong question has been asked, namely, when are languages genetically related? Sometimes in fact it is phrased as follows: when are two languages genetically related? What should be asked is, how are languages to be classified genetically? Note that in all of the quotations above, the problem is stated in terms of relationship. As Lamb (1959: 33) notes, 'To many linguists the classification of languages and the determination of relationships seem almost synonymous.'

Consider this example. A linguist proposes the following classification for certain languages of Europe: (1) Swedish, Sicilian, and the Laconian dialect of Greek; (2) Norwegian and Provençal; (3) Bulgarian and Icelandic; and so forth. In every one of these groupings the languages are related, since they are all Indo-European. Moreover, we may credit our hypothetical classifier with caution, for if he proceeds in like fashion a large number of independent stocks will be proposed. What is absurd, of course, is that none of these groups is a valid genetic unit. By a valid genetic unit is meant a group at any level whose members are closer to each other genetically than to any form of

speech outside the group. No doubt Bulgarian is related to Icelandic, but we are dealing here with a pseudo-entity from which strange cultural-historical conclusions would be drawn, and which does not constitute a reasonable unit for historical comparative investigation.

The concept of classification into valid genetic units in a hierarchy of various levels is a far richer notion than mere relationships. From such a detailed classification many statements of relationship of differing degrees can be deduced. Statements of relationship are thus mere consequences of classification, but not vice versa.

Note also that the above definition of valid genetic unit contained the phrase 'closer to each other genetically than to any form of speech outside the group'. The occurrence of 'any' in this definition requires that one look exhaustively outside the group, since such external evidence is relevant to determining the validity of the group. Those, therefore, who focus on a limited group determined by accidents of expertise, and anywhere else they just happen to look, are anything but cautious. For what is more incautious than to disregard relevant evidence, as any trained historian will attest?

We may distinguish two kinds of lack of caution in these matters, asserting and denying. In the former, two languages or low-level groupings are compared to the exclusion of other languages at least equally closely related, as would happen if one compared Swedish and Albanian in isolation and asserted their relationship. Equally incautious is to deny a relationship while disregarding relevant evidence, as for example when an expert in a particular Hokan language who is skeptical of its Hokan affiliation, or indeed the existence of a Hokan group at all, looks at only one other Hokan language. A comparable case would be that of an Armenian specialist who, when told that Armenian is an Indo-European language, compared it only with English. With such a procedure, the specialist may well be overwhelmed by the differences, unable to evaluate the similarities for lack of a comprehensive comparative framework, and unaware of important pieces of evidence for Armenian being Indo-European because they happen not to appear in English, a point that will be developed in detail later in this chapter.

In light of the distinction between relationship and classification, the statement sometimes made that you cannot disprove the relationship of two languages becomes uninteresting. No doubt you cannot *disprove* that Nahuatl is related to Swahili, but you can disprove that Nahuatl is closer to Swahili than to Pima. It is to account for such comparative degrees of resemblance that one posits that Nahuatl and Pima must belong to some valid genetic group (in this instance Uto-Aztecan) that does not include Swahili.

We see that from this point of view the problems of subgrouping and classification are closer than has generally been realized. Indeed if all the languages of the world are related, the problems become identical. Classifying the languages of the world becomes simply a matter of subgrouping a single large stock.

But in subgrouping it is once more the distribution of similarities that counts. The significance of a particular similarity, insofar as it bears on classification, becomes apparent only when we know where else it is found. Put another way, the significance of distribution as the essential basis of historical inference is known to all historically oriented anthropologists, and language is merely a special case.

Note that all that has just been said is based on the notion of evaluating resemblances, and the point has sometimes been made that the notion of resemblance is vague. However, what is involved in classification is not the registering of a resemblance, but a noting of the comparative degree of resemblance. Is a form A more like B than it is like C? Given, for example, *pan/fan/ezuk*, who would hesitate? What is meant, moreover, by greater resemblance is diachronic resemblance, that is, the probability that A and B derived by changes from a common source, as compared with C's having derived from a common source at greater remove (e.g. *four/vier/cuatro*) or from a different source altogether (e.g. *hand/Hand/mano*).

We may distinguish synchronic from diachronic resemblance even though they are enough alike that they can be largely equated in the heuristics of classification. Sounds and meanings by and large change to other sounds and meanings that are synchronically similar, e.g. the change from p to b, which involves a single feature difference. But s and h may be said to be diachronically similar because of the frequently attested change $s > h$, whereas in synchronic analysis they differ by a whole series of features. This example also illustrates another characteristic of diachronic resemblance, its frequent asymmetry, since $h > s$ is not known to occur, whereas synchronically, by definition, s is as similar to h as h is to s.

Further, if we find three forms that all look very different from each other, no judgment of differential similarity is required. We base our classification on the strong predominance of similarities in one language, or set of languages, in comparison with another. There will no doubt be marginal cases, but even the most sophisticated techniques of the comparative method cannot decide *all* etymologies, as the reading of even a page or two of standard etymological dictionaries will show.

Also to be taken into account is the fact that as the number of languages ultimately known to belong to a grouping at some level increases, the

precision of our judgement increases both in regard to the lower level of decisions, mere cognation, and in regard to at least some higher-level deductions regarding the shape and meaning of the source form.

Now we turn to the question of sound correspondences. Suppose there were a test that, when applied to two or more languages, always gave a definite answer. Let us suppose it is like a litmus test. The paper turns red when the languages are related, blue when they are not. Faced with, say, 1,000 languages in Africa, we begin to apply it. But even with pairwise comparison, there are $1,000 \times 999/2$, or 499,500 pairs we could choose. And even if such a test existed and gave valid results, the vast work of subgrouping would remain.

In fact, probably no one claims that we can devise a classification by regularity of sound correspondence, only that we can test hypotheses that have already been proposed. We therefore need a method of forming hypotheses. The number of ways of classifying n objects into one set, two sets, etc., up to n is called the partitions of n. Even without subgrouping, the number of partitions as a function of n increases astronomically with increasing values of n. The number of ways of classifying merely 20 languages is already $5,172 \times 10^{10}$, i.e. over 51 trillion.[1] For 1,000 languages, of course, the number is far more staggering. How this is to be dealt with is discussed later in this chapter.

Those who have realized that as an initial step one must first choose some hypothesis in order to test it by regularity of correspondence maintain, then, that the comparative method is not a method of *arriving* at a classification, but a method of *proving* a classification already hypothesized. What is not taken into account is the truly astronomical number of possible classifications, as just noted. No method is given for choosing a hypothesis except 'inspection' or perhaps intuition, as mentioned above by Bouquiaux.

Basically what I am denying is that there really are two separate steps. This possibility has been noted by some well-known philosophers of science. For example, in a discussion of Norwood R. Hanson's theories, Peter Achinstein (1977: 358) states: 'Any of the reasons Hanson mentions for suggesting a hypothesis can also be, and often are, reasons for accepting it. Take Hanson's retroductive reasoning, the fact that a hypothesis offers a plausible explanation of the data can be a reason for accepting it once it has been suggested. There is no such thing as a *logic* of discovery as opposed to a logic of justification' (italics in the original).

Returning to our litmus-test analogy, we must, however, ask if any such test exists. Considering that such expressions as proof, demonstration, and

---

[1] For a discussion of the mathematics of this function, see Greenberg (1957: 44) [§2-Ed.]. I am indebted to George Collier for providing a computer program that calculates the values of this function.

certainty constantly recur in the literature, one can reasonably ask for a rigorous procedure. But in fact a variety of versions occur, usually not worked out in any detail, but alluded to as if generally understood, and equated in some fashion with the methodology of comparative linguistics as developed by the nineteenth-century Neo-Grammarians. I shall therefore set up a number of models that can be constructed on the basis of the sorts of statements that are commonly made in the literature. They seem to arrange themselves between two poles, an emphasis on regularity of correspondence at the one pole, and on the reconstruction of an ancestral or protolanguage at the other.[2]

An extreme example of the insistence on the regularity of correspondence is István Fodor's book (1966) on the problems of African linguistic classification. Fodor appears to demand that whenever a particular phoneme x is found in one language, we should always find a particular phoneme y in the other. But clearly a simple regular merger in one of the languages not occurring in the other, or regular conditioned change resulting in a split in one not occurring in the other, will produce results that do not meet this test. As in virtually all these discussions, it is not clear what the consequence is of discovering an irregularity in correspondence, however defined. There are four possibilities. The most drastic is to deny the relationship; the next most drastic is to reject that particular etymology. A third possibility is to accept the etymology if it seems strong on other grounds, e.g. length of the form, semantic plausibility, and widespread distribution in other languages that display numerous other similarities with the languages under consideration, so that they appear to be, or are accepted as forming, a valid linguistic stock. If the etymology, along with the relationship, is accepted, one may seek to explain what appears to be a discrepancy by employing one of numerous strategies, nine of which are discussed below. The last possibility is to accept the etymology as valid without explaining all the related forms, and simply to admit that certain developments may become clear in the light of subsequent knowledge or may even remain indefinitely inexplicable.

Fodor, it seems, opts for the most drastic: rejection of relationship. But on that basis, we can disprove the relationship of New York City English to Philadelphia English. In New York City *bɔːd*, *ɔː* corresponds to Philadelphia *ɔr* in 'board,' but *ɔː* in 'bawd'. This is no doubt a straw man, though I believe it is Fodor's doctrine. It does, however, make an important point: we must distinguish between processes and the results of processes. To some extent the distinction is made when the term 'sound law' is reserved for instances in which a particular sound changes, either unconditionally or under stated

---

[2] For a clear statement of this view, see Gleason (1955).

phonetic conditions, in a particular language. Whenever this occurs in two or more languages stemming from a common source, the result will be a regular correspondence, although, as we have seen, it by no means needs to result in the sort of simple situation posited by Fodor. Indeed, it is a commonplace of reconstruction that the same protophoneme is reflected by a whole series of different correspondences under varying original phonetic conditions. Moreover, many other processes are admitted by even the strictest comparatists, e.g. metatheses and sporadic distance and adjacent assimilation and dissimilation, very often of liquids, nasals, or sibilants.

The comparative method of reconstruction is essentially the application of what we know about the processes of change in general in order to reconstruct the probable historical course of events, starting from a hypothetical reconstructed original. Of these processes, regular sound change is but one, even in phonology.

A less naïve view than Fodor's is expressed by Franz Boas (1929: 15) when he states that 'until definite phonetic shifts can be provided by a sufficient number of parallel forms, the question of relationship must be open.' This is, of course, exceedingly vague; for a start, we are not told what is a sufficient number of parallel forms. But Boas's view, if I understand it correctly, is broader than Fodor's, since he would not exclude, for example, an *s* in one language in some instances corresponding to an *s* in another language but in other instances to an *š*, provided they were recurrent (i.e. occur more than once). In the absence of a statement to the contrary, this would presumably not exclude the occurrence of a unique correspondence in a valid etymology, requiring only that any relationship have at least some recurrent ones.

Robert D. Levine (1979), in his critique of Edward Sapir's Na-Dene hypothesis, goes beyond Boas to reject all etymologies in which there is at least one unique correspondence. However, this position fails to take into account the fact that uniqueness and regularity are not necessarily related. A resemblance may be unique but regular. On the other hand, it can be recurrent, yet in one or more of its instances not a genuine correspondence at all. Let us examine both phenomena.

In the first case, a unique correspondence may be regular if it reflects a rare protophoneme that happens to survive in one etymology common to the two languages, or if there have been conditional splits in two languages, but in differing environments. In the simplest instance, a single but different split in each language, there will be four correspondences, at least one of which may be under such restricted conditions that only one instance will occur. In addition, a unique correspondence may occur if any of a number of

well-attested but sporadic processes is involved, e.g. dissimilation of liquids in one of the languages.

In the second case, consider the phonetically unusual correspondence English *d* = French *f*, as in 'do' and *faire* 'to make', 'to do' (in fact an established etymology in which the initial consonant is traditionally reconstructed as *dh, but which appears to be the only etymology containing this correspondence if we confine ourselves to English and French). We may then happily note *head* = French *chef* (*h/š* is in fact recurrent) and thus find a second instance of *d* = *f*. However, the further comparison in Germanic of English 'head' with German *Haupt*, Danish *hoved*, etc., and on the French side with the Rumanian plural *capete*, which preserves the *t* of Latin *caput*, shows us that *d* in English corresponds not to *f* but to *t, lost in French, while the *f* in French *chef* corresponds to the *p* of *Haupt*, etc., lost in English. This last example is typical in showing how reconstruction is always an approximation, and how, not surprisingly, the approximation becomes more precise as further evidence from other related languages is introduced.

Thus far we have cited those whose emphasis is on the regularity of correspondence. At the other pole, and this is characteristic of much recent work in the Amerindian languages, one can emphasize the carrying out of a reconstruction as proof. A straightforward expression of this point of view is found in Henry A. Gleason's workbook for his textbook on descriptive linguistics. After giving a comparative list of 25 words in Swahili, Kikongo, Zulu, Persian, Malay, and Tagalog, he cautions: 'Do not attempt to reconstruct. To prove your conclusions this would be necessary, but adequate data have not been provided here' (1955: 58).[3]

Let us call those who put their trust in sound correspondences strict constructionists and those who emphasize reconstruction as proof, loose constructionists. Then, to put it epigrammatically, the really rigid strict constructionists can never succeed, and the really loose constructionists can never fail. Besides the later Boas, we may cite Harry Hoijer as a strict constructionist who started out believing in a relatively simple theory of sound correspondences and ended up profoundly skeptical of the validity of all except the very lowest level of Amerindian classifications. For those who see reconstruction as proof, there are so many quite legitimate ways of explaining what are apparently irregular correspondences that there is no empirical way of disproving a reconstruction.

---

[3] Gleason is also typical here in another respect. Instead of asking the student to classify the six languages, he asks for pairwise judgements of relationship: 'For each pair of languages between which vocabulary similarities were found, determine whether they may be considered related by re-examining the pairs of words that are similar' (ibid.).

Before we consider this topic in more detail, it will be helpful to make a number of distinctions, some of which have been made in previous discussions, but some of which have not. To begin with, we must distinguish sound changes, which when regular are often called sound laws, and their results, which when we compare them in related languages are regular correspondences. Thus if, from a reconstructed protophoneme $*k$ (there can of course also be cases historically attested from documents) in language A, there was a regular unconditional change $*k > x$ such that k became x in all instances, whereas in language B k became $\check{c}$ regularly before front vowels but otherwise remained k, there would be three sound laws, $k > x$, $k > \check{c}$, $k > k$, but two correspondences, $x : \check{c}$ (before earlier front vowels) and $x : k$ (everywhere but before front vowels). Since the correspondences result from the changes, there is a tendency to use the terms almost interchangeably, which can lead to confusion.

A further distinction is that between a recurrence and a correspondence. A recurrence is a surface phenomenon. Thus, in our earlier example, the d of English 'head' and the f of French *chef* are an instance of a recurrence, but not of a correspondence. There is a further distinction between a correspondence in general and a regular correspondence. By a regular correspondence is meant an occurrence in two or more languages, relative to a particular theory of reconstruction, of a set of sounds that came from a common ancestral sound only by regular conditioned or unconditioned sound changes. Thus the p of Latin *pater*, the p of Sanskrit *pitā*, and the f of English 'father' are regular correspondents that derive from Proto-Indo-European $*p$. But the m of Latin *novem* 'nine', though it corresponds to the second n of English 'nine' because it derives from the same original sound, does so with a change from n to m owing to the influence of other numerals, such as *septem* 'seven' and *decem* 'ten'. The m is in correspondence with the second n of 'nine' in the sense that it comes from the same original sound, but it includes a process other than regular change in its derivation, contamination. Hence it is not a regular correspondence. The most common such process, of course, is morphological analogy. Such correspondences will be called nonregular, when it is necessary to distinguish them from regular ones.

Correspondences (both regular and nonregular) are relative to a particular theory of reconstruction. For example, certain recurrences between Germanic, on the one hand, and other Indo-European languages, on the other, were viewed as nonregular correspondences until the late 1870s because they did not accord with Grimm's law. They became regular with the discovery of Verner's law, which showed that they were the result of regular changes

under certain conditions statable in terms of the position of the Proto-Indo-European pitch accent as found, basically, in Greek and Sanskrit. Had Verner rejected the forms as unrelated because of the nonregularity or, more drastically, denied the validity of Indo-European, he would never have made his discovery.

These various distinctions will be useful in discussing my thesis that the loose constructionist can never fail. Faced with what looks like a nonregular correspondence, a term that, as just noted, is always relative to a particular reconstruction, the comparatist has at least ten options, all of which can be documented from proposals actually made in the literature of Indo-European studies or that of other linguistic families. Since different solutions to the same problem have often been offered by different linguists, the same examples will in some instances occur under different rubrics.

The first option, often used by the Neo-Grammarians, is to give additional conditioning factors. In this way no new protophoneme is postulated, but of course the number of examples of each correspondence is reduced, and the number of distinct regular correspondences associated with the specific protophoneme increases. If one makes the conditions fine enough, one can account for all apparent deviations, but some may be improbable on phonetic or other grounds. An example is Fortunatov's law relating to Proto-Indo-European $*r$ and $*l$. Indic has examples of both l and r recurring in words with either r or l in other languages. In other words there are four sets of recurrences: $r = r$, $r = l$, $l = r$, and $l = l$. Fortunatov seeks to show that the distinction of $*r$ and $*l$ survived in Indic. According to him, $*r$ + dental remained unchanged, but $*l$ + dental resulted in the loss of l and the modification of the original dental to a retroflex consonant. There is an alternative and more generally accepted explanation, as we shall see later.

A second option is in a sense the exact opposite of the first. Instead of refraining from further complicating the protolanguage and positing new, later changes under differing phonetic circumstances, one attributes the recalcitrant cases to different protophonemes that survived only in the given language and merged everywhere else. It is not even always necessary to add a new phoneme to the protolanguage. We can attribute the differing sounds to varying sequences of phonemes already posited for the ancestral speech. For example, corresponding to what has usually been reconstructed as Proto-Indo-European $*y$, Greek has h in some instances and z in others. G. Schulze and others, including Karl Brugmann, attributed this to two different phonemes that merged everywhere except in Greek. Sapir, using the laryngeal phonemes based on Ferdinand de Saussure's original theory and the more recent evidence of Hittite, which seems to preserve them, at least in some

instances, as ḥ deduced initial *h* from a voiceless laryngeal (which he needs anyway) followed by *y*, with *z* the result in other environments.

As with the first option, we can always make this one work, but at the cost of constantly adding to the number of protophonemes or their combinations. However, a protolanguage with, say, 125 phonemes is completely implausible on typological grounds. This shows in any case that if we are to consider a reconstruction proof of relationship, we need tests for the plausibility of the reconstruction itself, in which typological factors will figure prominently.

Suppose we find that in two highly similar tonal languages both have two pitch phonemes, high and low, but that every possible equation is a recurrence: H = L, H = H, L = H, and L = L. Will this prevent reconstruction and thus disprove the relationship? We can simply posit four tonal protophonemes as the source of each correspondence, and in this case no typologically cogent objection can be brought, since four-tone languages are well attested.

A third option is not to attribute the apparent irregularity to phonological processes alone, but to resort to what has been called reverse analogy. Suppose there is morphophonemic alternation in the original language as the result of earlier phonetic changes. In each word in which it occurs the alternation can be eliminated by the analogical spread of one of the alternants. However, different words in the same language, and the same original word in different languages, can analogize in different ways.

This form of explanation is well known and constantly resorted to by comparatists dealing with morphologically complex languages. Thus the *e* of Latin *ped-* and the *o* of Greek *pod-*, an example of a nonregular correspondence, are not attributed to a single earlier phoneme. Rather, as part of the Ablaut (vowel alternation) system of Proto-Indo-European, *e* occurred in certain forms of the paradigm and *o* in others. Thus from an original alternation *ped/pod*, Latin generalized *e* and Greek *o*. Many Indo-Europeanists treat *e* rather as a reduced grade that replaces a theoretical zero grade when unaccented or when unacceptable clusters result. The principle, however, remains the same. Evidently reverse analogy can lead to 2^n sets of recurrences across *n* languages. With two languages, as in the foregoing example, *e/e, e/o o/o*, and *o/e*, are all possible sets of recurrences.

A further example is once more Verner's law in Germanic, according to which one pair of alternating phonemes arising from an earlier change conditioned by the Indo-European accent appear in English and German as *r/z*. In *hare* English generalized *r*, while in *Hase* German generalized *z*; in *lose* English generalized *z*, while German generalized *r* in *verlieren*. In many other instances *r* = *r* and *z* = *z*. Survivals of the original alternation are English *was/were* and *lose/forlorn*. The first alternation is of course lost in *r*-less dialects.

A kindred but distinct source of nonregular correspondence is morphophonemic analogy. Suppose, as a result of earlier changes, two morphophonemic alternations, X ~ Y and Y ~ Z, have survived in parallel fashion in two languages. Because of the equivocal status of Y it will often happen that X replaces Z, or vice versa. The result is a nonregular correspondence, which may well be recurrent, of the type X/Z, or Z/X, respectively, in which the sounds are not reflexes of the same original phoneme. If $Z > X$ or $X > Z$ in both languages, we will reconstruct the wrong phoneme.

A fourth option is somewhat like the preceding except that the condition of the alternation is presumed to be a word-boundary phenomenon ('word sandhi'). In initial or final position it is supposed that phonetic variants depended on the nature of the preceding or following word, respectively. An example of the former is so-called Indo-European *s*-movable. Roots with an initial consonant often appear with or without a preceding *s-* in a completely sporadic way, e.g. Greek *tegos* 'roof', but Lithuanian *stogas* with the same meaning. One commonly espoused theory is that the forms with *s-* arose in contexts in which the previous word ended in *-s*, a common final in Proto-Indo-European, with a reinterpretation of the position of the word boundary so that *s-* was considered part of the following word, giving rise to two variants for each word. These generalized in different ways in different words across languages. A similar phenomenon in word final position is the existence of forms with final long vowels followed by sonants (*n, r, w,* etc.) in some instances and not in others, e.g. Sanskrit *çvā* 'dog' as against Greek *kuōn.* A common explanation is that the sonant was lost before the initial consonant of the next word, and thus sandhi alternants arose that were generalized differently in different instances. Lane (1968) wishes to limit the loss of sonants to instances in which the following word begins in a sonant. There is, of course, a difference in principle between these two examples. In the first an element is transferred to a word as a result of the reinterpretation of the place of the boundary. In the second an element is *lost* at the boundary. They could therefore be considered examples of somewhat different processes.

A fifth option is allied to the preceding two in that in order to account for apparently random variations across languages, two alternants are assumed to have existed in the ancestral language, and individual languages chose one or the other independently in each instance, giving rise once more in n languages to 2^n correspondences, some of which may be unique. But whereas in the preceding type, the origin was phonological, here a formerly meaningful element has 'faded' so that it has become a purely conventional element, either devoid of function or with a secondary acquired function.

Some Indo-Europeanists have suggested such a theory to explain the origin of *s*-movable, namely, that it was a formerly meaningful prefix.[4] In an earlier work (Greenberg 1978) I have described the process by which a demonstrative becomes first a definite article, then a combination of definite and indefinite article (stage II article), and finally an empty marker. In the last stage, stage III, instead of always remaining on the noun or adjective, or always disappearing in a specific language, we may have random survival for each individual word. An example of this is a prefixed *k*- found in random fashion, distributed differently for each word across languages in all nouns that originally began with a vowel in Nilo-Saharan. In Greenberg (1963) it is called *k*-movable and has from all appearances developed from an old stage II article already in the protolanguage. Since this *k*- did not differ for noun class or gender, it is a marker of nominality as such in its final stages.[5]

An example closer to hand is the free and semantically empty use of diminutive suffixes in Romance languages resulting in such cognates as Italian *sole* 'sun' < Vulgar Latin *solem*, as against French *soleil* < Vulgar Latin *soliculum*, literally 'little sun', with of course many examples of the opposite kind. A widespread North American phenomenon of this kind is 'diminutive consonantism'.

Under the sixth option I include a whole series of sporadic processes that are well known and accepted as explanatory by even the strictest Neo-Grammarians. Some are purely phonetic, e.g. assimilation and dissimilation, which are especially common in liquids and sibilants, or metathesis, the exchange of position of two different sounds. For example, English *wasp* and Lithuanian *vepsa* are universally recognized as cognates. In English *wasp* comes from an earlier *waps* (cf. the variants *ask* and *axe* in English). Such a change is nonregular, or sporadic, there being no general rule that all earlier *ps* sequences became *sp* in English. Other sporadic changes are generally attributed to semantically related forms, e.g. Vulgar Latin *grevis* 'heavy' as against Classical Latin *gravis* through the influence of *levis* 'light'. Again Latin -*m* in *novem* 'nine', as against final -*n* elsewhere, is not attributed by anyone to a third protonasal besides *m* and *n*. The solution is obvious when one takes into account the inherited final -*m* of *septem* 'seven' and *decem* 'ten'.

A seventh option in dealing with cases that are deviant on the basis of a specific reconstruction is the method Malcolm Guthrie heavily relies on in his important study on comparative Bantu (1967). Faced with the usual deviations, Guthrie, a firm believer in the complete exceptionlessness of sound

---

[4] For the literature on this theory, see the references in Szemerényi (1970: 88).

[5] On this point, see Greenberg (1981). The 'Penutian Parallel' referred to in the title of this article concerns a suffixed -*s* of similar origin. [see also §10-Ed.]

laws, simply reconstructs numerous variant forms of what are clearly the same word. In a spot check, I found that about 25 per cent of the words he reconstructed required at least two variant, and phonetically similar, proto-forms. An extreme case is the word for 'day', which is reconstructed as *cîku, *cûgu, *cûku, *tîku, *tûku, and *tûkû. Guthrie does not, of course, maintain that these six forms are six different words for 'day', all synonyms that happen to be phonetically similar; instead he calls them 'mutations', merely a polite term for irregularities. A resort to this method on a still vaster scale is found in Calvin Rensch's reconstruction of Oto-Manguean (1976). What is posited in these cases is presumably free variation among competing forms in the protolanguage. This is sometimes a convincing explanation if we have reason to believe that a particular sound was just in the process of change at the time of the dialectal break-up of the ancestral language.

An eighth option is to explain irregular correspondences by positing dialect mixture. This is a now widely accepted explanation for the Indic liquids r and l, which, as we have seen, Fortunatov explained by positing more conditioning factors. Many now believe that *r and *l merged in earlier Indic, with some dialects taking r for the liquid that resulted, and others l. The dialect represented in literary Sanskrit would, on this view, have taken some of its forms from an r-dialect and some from an l-dialect.

A ninth option is to reject the difficult etymologies altogether as not being valid cognates. W. S. Allen (1953) saw this possibility clearly. One can prove the complete regularity of sound laws simply by eliminating any exceptions, thus tautologically producing regularities. Many believe that lack of conformity to regular correspondences routinely results in the elimination of etymologies. Actually such cases are very rare because of all of the other resources just described. If we resorted to it on a wholesale scale, we would need another theory to account for the existence of so many near misses.

The tenth and final option is frequently the wisest policy. This is neither to reject recalcitrant cases as invalid etymologies nor to accept an explanation where none looks very plausible. It is now recognized that all, or very many, sound changes leave unexplained residues. Perhaps some day an explanation will be found by greater ingenuity or by the discovery of new data. In some cases—for example, Indo-European problems that have not yielded a satisfactory explanation after over a hundred years of intensive effort—we must perhaps simply say *ignorabimus*.

Up to this point we have proceeded as if there were no question that sound changes are regular and have simply discussed the various other processes that could account for apparent exceptions. But in fact the Neo-Grammarians, who at first proclaimed the absence of exceptions to sound laws, soon gave

ground and stated this as a mere hypothesis. As Berthold Delbrück (1880), a leading Neo-Grammarian, put it, sound laws without exception are only to be found in heaven. According to both Schachter and Fodor, Carl Meinhof 'proved' that the Bantu languages were related by reconstructing Proto-Bantu. But Meinhof himself took the later Neo-Grammarian position, writing (1932: 20) that 'there are and always will be a number of exceptions for which no explanations are found.'

In fact, although sound changes are largely regular, there are well-attested cases of wholesale violations. An example is that one of the tones of early Chinese, the Ju-Shêng, gave rise to each of the four tones of modern Mandarin under conditions that are not completely understood. R. A. D. Forrest (1950) gives a sort of statistical solution to this problem. For example, after initial unaspirated consonants we find the second tone in 52.1 per cent of the cases, whereas after aspirated consonants we find the fourth tone in 62.1 per cent, but for both classes all four occur. One could, of course, take the second of the ten earlier options and posit a much more complex proto-Chinese system. But there is no evidence for this. In doing it, ironically, one would have 'proved' the relationship of Mandarin to the other Chinese dialects, hardly a necessary exercise, by an invalid reconstruction.

The foregoing example refers to tone, which some linguists may think of as a marginal feature. But no one presumably would deny the centrality of the vowel system. Yet in Finno-Ugric, a family whose comparative treatment is still earlier than that of Indo-European and whose validity is not doubted by even the most conservative, scholars have not been able to reconstruct the ancestral system of vowels. At best they have succeeded only in determining whether vowels of the initial have front- or back-vowel harmony. József Szinnyei (1910) notes that the decision is based mostly on Hungarian, Finnish, Mordvin, Cheremiss, and the Tavda dialect of Vogul. Where they disagree, he simply goes with the majority. For noninitial syllables he concludes that since proto-Finno-Ugric probably did not have vowel harmony, the vowels of the noninitial syllable must have been independent of the initial vowel. But, he adds, 'We cannot say anything beyond this in regard to the vowels of the second and subsequent syllables' (Szinnyei 1910: 52).

The common doctrine thus inverts scientific logic. It purports to prove etymologies that are strong on other grounds by a reconstructed sound system and a set of historical changes, whereas it is rather the success of a particular theory in explaining the best-established etymologies that tests the plausibility of the reconstruction itself. Even in the best-studied families, such as Indo-European, the theories may undergo drastic changes while the main body of etymologies remains unchanged. Edgar Sturtevant (1942: 37) notes in

reference to English *fire*, Greek *pyr*, etc., after seeking to account for all the variations in different Indo-European languages, 'These details are in part mere suggestions, but the underlying etymology is beyond question.' That great apostle of common sense, Otto Jespersen, was well aware of this point: 'Nowhere have I found any reason to accept the theory that sound changes always take place according to rigorous or "blind" laws admitting no exceptions. . . . It is very often said that if sound laws admitted of exceptions there would be no possibility of a science of etymology. There are, however, many instances in which it is hardly possible to deny etymological connection though the phonetic laws are violated' (Jespersen 1922: 295). After citing some examples, he adds: 'All this goes to show (and many more cases might be instanced) that there are in every language words so similar in form and meaning that they cannot be separated, though they break the sound laws.'

In fact there is no necessary logical connection between complete regularity of sound change and the possibility of genetic classification. If there were, classification would be largely impossible. All that is required is that the change not be random and/or enormously rapid.

Consider the following conceptual experiment. Suppose that words in languages consisted of sequences of five digits, and that there are two unconnected languages generated by independent random selections of these digits. Let us call these languages A and B. Suppose both split into two languages, C, D and E, F, respectively. Further suppose that after a fixed period, say 1,000 years, the probabilities of changing to the next higher digit is 25 per cent, the next lower 25 per cent, two higher 15 per cent, two lower 15 per cent, and three higher 10 per cent, three lower 10 per cent—in other words, an approximation to a normal probability distribution. It is clear that for an enormous period of time C will resemble D more than it does E or F, and E will resemble F more than it resembles C or D. In fact, as we will see later, reasonable analogues of such a situation occur in other fields in which the genetic branching model applies, and lead to results easily as reliable as those obtained in linguistics.

Regarding changing theories of reconstruction, Georg von der Gabelentz (1891) remarked that in the short period from August Schleicher (1850s) to Karl Brugmann (1880s) the views on the Indo-European protolanguage had changed quite a bit. He could be forgiven for failing to envisage that far greater changes were to come. After more than a century of effort, which has involved the majority of the linguistic community in the nineteenth century and a substantial group of specialists in the twentieth, it is not too much to say that the only matters on which everyone agrees in regard to the sound system of Proto-Indo-European are that there were at least four points of

articulation for the stops, including at least labials and dentals or alveolars, at least three manners of articulation for stops, and at least one sibilant and one vowel.

When it was discovered that Hittite had a back fricative ḫ in some of the positions in which de Saussure had posited a *coefficient sonantique* in his famous *Mémoire* (1879), most Indo-Europeanists accepted the existence of one or more laryngeals in Proto-Indo-European that disappeared in all the languages except Hittite and the other Anatolian languages. (Survival has, however, also been asserted for Armenian and Albanian.) But almost every number from zero (Kronasser, Bonfante, and others who reject laryngeal theory) to ten or more has been posited. The most popular have been two and three, but even in respect to these, different phonetic properties and conditioning effects are assumed.

In the last five decades or so, we may note, Hittite, Luwian, Lycian, Palaic, and other Anatolian languages have been universally recognized as Indo-European, as has Tokharian with two dialects, Tokharian A and B. In no case did anyone publish the sorts of articles with tables of correspondences and asterisked forms so common in the pages of the *International Journal of American Linguistics*, which are believed to reflect the methodology of Indo-European comparative linguistics. In the case of Tokharian, one must believe that all three (or four) consonant manners had merged into one, and then each phoneme, now the sole representative of the position of articulation, split into two under phonetic circumstances that no one has been able to state.

Why, then, have all these languages been accepted as Indo-European? The reason is the existence of a considerable number of word-stems resembling those that are widespread in Indo-European, and a number of highly characteristic grammatical formatives involving sound and meaning. Thus the existence of even a few forms such as Hittite *eszi* 'he is' and *asanzi* 'they are' (cf. Latin *est/sunt*, Sanskrit *asti/santi*) is quite sufficient to exclude accident.

It is widely believed that true cognates can be distinguished from borrowings because they exhibit regular sound correspondences. In Greenberg (1957) [§2-Ed.] I pointed out, and I was not the first to do so, that borrowed words sometimes showed equal or greater regularity. In an otherwise not always favorable review in *Language* (1959), Isidore Dyen, a rigorous comparatist of impeccable credentials, agreed with me, and others have subsequently made the same point.

In Table 1 we have most of the cognates from French and English, taken from etymological dictionaries. French and English are, of course,

TABLE 1 French-English Correspondences

1 pje = fʊt	29 sæk = fajv	57 lɛ:vr = lɪp
2 pɛ:r = 'faðər	30 si ~ sis = sɪks	58 lõ ~ lõg = lɔŋ
3 pwasõ = fɪš	31 sɛt = 'sɛ:vən	59 lu: ~ lu:v = wʊlf
4 plæ ~ plɛ:n = fʊl	32 sɔlɛ:j = sʌn	60 leš(e) = lɪk
5 pu:r = fɔr	33 sɛl = sɔlt	61 lɛ:n = wʊl
6 frɛ:r = 'brʌðər	34 sã = 'hʌnd(rɪd)	62 rasin = ruwt
7 fɛ:(r) = duw	35 sœ:r = 'sɪstər	63 katr = fɔr
8 fy = bij	36 sa:bl = sænd	64 ki = huw
9 fãd(r) = bajt	37 syœ:r = swɛt	65 kə = hwat
10 vã = wɪnd	38 (rə)səv(wa:r) = hæv	66 kã = hwɛn
11 vɛ:r = wʌrm	39 šod = hat	67 kœ:r = hart
12 vul(wa:r) = wɪl	40 šjæ ~ šjɛn = hawnd	68 kɔrn = horn
13 vœ:v = 'wɪdow	41 šɛf = hɛd	69 ku:r = jard
14 vif ~ vi:v = kwɪk	42 šɛ:r = howr	70 gɛ:p = wasp
15 vən(ir) = kʌm	43 žənu = nij	71 græ = korn
16 vɛt(i:r) = wer	44 žug = jowk	72 ɥi ~ ɥit = ejt
17 mwa: = muwn	45 nœf ~ nœv = najn	73 œj = aj
18 mɛ:r = 'mʌðər	46 ne = nowz	74 ɔrɛ:j = ir
19 mwajæ = mɪd	47 nɥi = najt	75 õ:gl = nejl
20 trwa = θrij	48 nœf ~ nœ:v = nuw	76 ø ~ œf = ɛg
21 etwa:l = star	49 nõ = nejm	77 (a)s(wa:r) = sɪt
22 tɔnɛ:r = 'θʌndɛr	50 ni = nɛst	78 æ ~ yn = wʌn
23 ty = ðaw	51 nɛ:ž = snow	79 žə = aj
24 twa = θæč	52 nõbri = 'nejvəl	80 nu ~ nuz = ʌs
25 trɛ:r = drɔ	53 ny = 'nejk(əd)	81 mwa = mij
26 dø = tuw	54 nwa = nʌt	82 (sɛ)t = ðə
27 dis = tɛN	55 lã:g = tʌŋ	
28 dã = tuwθ	56 ly(i:r) = lajt	

Indo-European languages of different branches. I do not vouch for the completeness of the list, and there may be a few errors. Still it is hardly deniable that it gives a general view of what these cognates are like. The tables that follow (2–4) break the Table 1 list down by initial and noninitial consonant correspondences and vowel correspondences.

An examination of the initial consonants, the most regular of all, shows for example that *k* in French corresponds to English *f* (once), *h* (three times), *hw* (twice), and *j* (once); and so for all the other initial consonants to a lesser or greater degree. The vowels are close to statistical randomness, and in this they are rivaled by the final consonants. In the light of the earlier discussion, the reasons should be obvious. Even regular sound change, when conditioned, gives a variety of correspondences. As time goes on these accumulate, along

TABLE 2 Initial Consonant Correspondences in Table 1

p = f, 1, 2, 3, 4, 5	š = h, 39, 40, 41, 42
f = b, 6, 8, 9	ž = zero, 43, 79
f = d, 7	ž = j, 44
v = w, 10, 11, 12, 13, 16	n = n, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54
v = kw, 14	n = zero, 80
v = k, 15	l = t, 55
m = m, 17, 18, 19, 81	l = l, 56, 57, 58, 59, 60, 61
t = θ, 20, 22, 24	r = r, 62
t = t, 21	k = f, 63
t = ð, 23, 82	k = h, 64, 67, 68
t = d, 25	k = hw, 65, 66
s = f, 29	k = j, 69
s = s, 31, 32, 33, 35, 36, 37	g = w, 70
s = h, 34, 38	g = k, 71

TABLE 3 Vowel Correspondences in Table 1

i = ɛ, 27, 50; zero, 62; uw, 64	ø = uw, 26
i ~ i: = ɪ, 14, 30	œ: = ɪ, 13, 35; ɛ, 37; a, 67, 73
e = zero, 21; ow, 46; ɪ, 60	ə = ʌ, 15; æ, 38; zero, 43, 79; a, 65
ɛ = ɛ, 31, 41; ɔ, 33	ã = aj, 9; ɪn, 10; uw, 28; ʌn, 34; ɛn, 66
ɛ = a, 2, 70; zero, 4, 61, 74; ʌ, 6, 11, 18; uw, 7; e, 16; ɪ, 57; ə, 22; ɔ, 25; ow, 42, 51	ã: = ʌ, 55
	æ = ɪ, 19; aj, 29; n, 71
a = ij, 20; uw, 62; ɔ, 63	õ = ejm, 49; ej, 52
a: = æ, 36	õ: = n, 75
ɔ = ʌ, 22, 32; o, 68; i, 74	wa = ɪ, 3; uw, 17; a, 21; æ, 24; ʌ, 54; ij, 81
u = ɪ, 12 ij, 43; ow, 44; ʌ, 80	
u = ɔ, 5; a, 69; ʊ, 59	jɛ = aw, 40
y = ij, 8; aw, 23; w, 37; ej, 53; aj, 56; ʌ, 78	je = ʊ, 1
	ɥi = aj, 47; ej, 72

with various sporadic phonetic processes, grammatically induced analogy, and new derivational affixes, etc., while lexical replacement reduces the number of cognates. All of these processes work to produce more and more diversity of correspondence and virtually never to reduce it.

In Table 5 I have reconstructed with great ease a pseudo-entity, Proto-Turco-Arabic, based on loanwords in Turkish from Arabic. But how do we know that Turkish is not in fact a Semitic language like Arabic? If anything, much current doctrine would force the conclusion that the case for Turco-Arabic is better than that for French-English.

Table 4 Noninitial Consonant Correspondences in Table 1

p = p, 70
b = n, 36; v, 52
f = d, 41; g, 76
v = zero, 13; v, 38; p, 57; f, 59
f ~ v = k, 14; n, 45; zero, 48
t = r, 16; v, 31; zero, 63; t, 72
d = t, 9
s = š, 3; n, 27; s, 30; t, 62; zero, 70
z = s, 80
š = k, 60

ž = zero, 51
k = v, 29
g = k, 44; ɲ, 55, 58; zero, 75
r = r, 2, 5, 6, 11, 18, 20, 22, 25, 35, 42, 63, 67, 68, 69, 71, 74; t, 37; l, 52; zero, 57
l = l, 4, 12, 33, 75; r, 21; n, 32
n = zero, 4; m, 15; n, 22, 40, 43, 68, 78
j = zero, 19, 74

Table 5 'Proto-Turco-Arabic' Correspondences

Arabic	Turkish	Proto-Turco-Arabic	Arabic	Turkish	Proto-Turco-Arabic	Gloss
b	b	*b	ʕajiba	ajibe	*ʕajibe	marvel (n)
b	p	*b	ṭālib	talip	*ṭaleb	student
f	f	*f	farāḥ	ferah	*ferah	spacious
w	v	*w	dawāʔ	deva	*dɛwaʔ	medicine
m	m	*m	mufahham	müfahham	*möfahham	illustrious
t	t	*t	tablīǧ	tevlig	*tɛbliǧ	communication
d	d	*d	dirham	dirhem	*derhɛm	dirham, a coin
d	t	*d	ʕābid	abit	*ʕabed	worshipper
n	n	*n	ʔajnabī	ejnebi	*ʔejnɛbi	foreign
s	s	*s	sākin	sakin	*saken	stationary
ṣ	s	*ṣ	ṣaff	seff	*sɛff	row, rank
z	z	*z	ʕazīz	aziz	*ʕaziz	powerful
š	š	*š	kaśf	keśf	*kɛśf	exposure
th	s	*th	mathal	masal	*mathal	story
dh	z	*dh	dhamm	zemm	*dhɛmm	blame (n)
ṭh	z	*ṭh	manṭhur	manzur	*manṭhur	considered (a)
ḍh	z	*ḍh	xamiḍh	hamiz	*xamɛḍh	acid (a)
r	r	*r	raybiya	reybiye	*rɛybiɛ	skepticism
l	l	*l	lāḥim	lahim	*laḥem	carnivorous
y	y	*y	yābis	yabis	*yabes	dry
j	j	*j	ʔajala	ajele	*ʔajɛlɛ	haste

One obvious consideration, of course, is that these loanwords are not basic vocabulary items. But we do not even need such a hypothesis. The most powerful proof is, once more, distribution across languages. Turkish and Arabic are not mutually intelligible and are obviously distinct languages.

Hence, if Turkish were really a Semitic language, it would show some independence within that family. But Turkish never has a Semitic morpheme unless it occurs in close to the same form in Arabic. In the absence of direct historical evidence, which is of course present in this case, this is the most powerful evidence for borrowing.

In Table 6 I have listed just three of the eight Bantu languages Meinhof used for his reconstruction of Proto-Bantu. Among them we see no less than ten correspondences, all representing his reconstructed *g. Coupez subsequently posited two phonemes, *g and *ɣ, and this is now generally accepted, but it will not account for all the facts. The addition of others of the hundreds of Bantu languages will, of course, greatly increase the number of additional and mostly unique correspondences. Recall that Schachter and Fodor both stated that Meinhof had 'proved' the validity of Bantu by his reconstructions.

TABLE 6 Bantu Correspondences

Swahili	Konde	Sango	Proto-Bantu (Meinhof)	Gloss	Correspondences
imba	imba	lu-yimbo	*gimba	sing	0/0/y
mw-ona	gɔna	ɔna	*gona	sleep	0/g/0
gumu	uma	yuma	*guma	hard	g/0/y
ganja	iky-anja	li-ganja	*ganja	lake	g/0/g
uki	ul-uki	iny-usi	*guki	sweetness	0/0/0
gawa	ya'ba	—	*gaba	divide	g/y/-
el-elza	gɛla	jela	*gela	measure	0/g/j
gwa	gwa	gwa	*gua	fall	g/g/g
fyag-ia	phyag-ila	fyaj-ila	*piaga	sweep	g/g/j
m-kuyu	u-khuyu	—	*kugu	fig tree	y/y/-

In Table 7 I have listed a few basic words for twenty-five languages of Europe.* The number of ways of classifying twenty-five languages, even without specifying subgrouping, is $.4639 \times 10^{19}$, that is, over a quintillion. Yet the correct classification and even subgroupings and intermediate groupings (e.g. Balto-Slavic) are apparent from just a cursory glance at two or three words. The power of a method that looks at everything at once, instead of testing isolated hypotheses, is thus immense. This method may be called multilateral comparison.

---

* Table 7 was intended to be the same as Table 2 in §2. However, the original Table 7 includes orthographical forms instead of a broad phonemic transcription, and contains a number of errors. I have reproduced here Table 2 from §2-Ed.

Table 7

	One	Two	Three	Head	Eye	Ear	Nose	Mouth	Tooth
Breton	ünan	dau	tri	penn	lagad	skuarn	fri	genu	dant
Irish	ö:n	dɔ:	tri	kjan	su:lʲ	kluəs	srɔ:n	bjal	fjakəlʲ
Welsh	ɨn	daɨ	tri	pen	ɬəgad	klist	truɨn	keg	dant
Danish	en	to:ʔ	tre:ʔ	ho:dhə	ɔjə	o:rə	nɛ:sə	monʔ	tanʔ
Swedish	en	tvo	tre	hüvud	öga	öra	näsa	mun	tand
Dutch	e:n	tve:	dri:	ho:ft	o:x	o:r	nö:s	mont	tant
English	wən	tuw	ɵrij	hɛd	aj	ihr	nowz	mawɵ	tuwɵ
German	ajns	tsvaj	draj	kopf	augə	o:r	na:ze	munt	tsa:n
French	ɶ̃, yn	dö	trwa	tɛ:t	œj/jö	orɛ:j	ne	bu:š	dã
Italian	uno, una	due	tre	tɛsta	okkjo	orekkjo	naso	bokka	dɛnte
Spanish	un, una	dos	tres	kabesa	oxo	orexa	naso	boka	diente
Rumanian	un	doj	trej	kap	okiu	ureke	nas	gurə	dinte
Albanian	n'ə	dü	tre	kokə	sü	veʃ	hundə	goja	ðəmp
Greek	enas	ðjo	tris	kefáli	máti	aftí	míti	stóma	ðóndi
Lithuanian	vienas	du	tri:s	galva	akis	ausis	nosis	burna	dantis
Latvian	viens	divi	tri:s	galva	atss	auss	deguns	mute	zobs
Polish	jeden	dva	tši	glova	oko	uxo	nos	usta, gēba	zōp
Czech	jeden	dva	tši	hlava	oko	uxo	nos	usta	zup
Russian	adʲin	dva, dvʲe	trʲi	gəlavá	óko	úxo	nos	rot	zup
Bulgarian	edin	dva	tri	glava	oko	uxo	nos	usta	zəb
Serbo-Croatian	jedan	dva	tri	glava	oko	uho	nos	usta	zub
Finnish	üksi	kaksi	kolme	pä:	silmä	korva	nenä	su:	hammas
Estonian	üks	kaks	kolm	pea	silm	kõrv	nina	su:	hammas
Hungarian	ed	ke:t	ha:rom	fö:, fej	sem	fül	orr	sa:j	fog
Basque	bat	bi	hirür	bürü	begi	belari	südür	aho	orts

If, following the arrangement of Table 7, words are listed in the horizontal dimension and languages in the vertical, we may say that the method advocated here is the vertical. The commonly advocated one is horizontal. We are accustomed to looking at a few languages across many words rather than at many languages across a few words. Let us say that the vertical method is synoptic, like scanning an entire forest from the air; then the horizontal method can be likened to wandering about in the dark undergrowth, here and there espying a similarity at random without any possibility of understanding the totality.[6]

[6] The vertical method has a superficial resemblance to glottochronology, but glottochronology gives us only pairwise percentages. A full-scale critique of glottochronology is not within the scope of this work; on the fundamental weakness of pairwise comparison, see Gleason (1959).

Kroeber, using a format he called tabular (identical with that of Table 7) for Uto-Aztecan, whose unity was still in doubt for many, stated (1906–7: 162):

As to the conclusions to be drawn from this table there can be no question. The evidence of the genetic relationship of all the languages represented from Nahuatl to Luiseño is overwhelming and leaves room only for wonder how the fact could have been doubted. Others have perhaps had the experience of comparing some particular Shoshonean dialect with Nahuatl on the strength of relationship currently announced, and of being disappointed at the small number of positive resemblances visible; but the present collocation in compact and unified form from all dialect groups alters the condition thoroughly so that identities, which before could only be suspected and seemed exceedingly doubtful, are revealed with entire certainty.

Confronted by the numerous independent stocks in California posited in the Powell classification, Kroeber himself once told me that after fruitlessly making individual paired comparisons, he one day conceived the notion of looking at all of them simultaneously. He and his co-worker Dixon immediately noted that two major groupings accounted for a large majority of these families, to which they gave the names Penutian and Hokan; and these groupings, later expanded to include many languages outside of California (some subsequently noted by Kroeber himself), have stood the test of time.[7]

Another way of seeing the importance of multilateral comparison, and incidentally showing the weakness of pairwise percentages as used in glottochronology, is the following. Suppose we were to compare in isolation English and Hindi, known to be related since both are Indo-European. As we compared each pair of words, we would often encounter quite dissimilar forms, and this would seem to argue against a connection. But in fact one or the other might be Indo-European, as would be shown from a wider comparison, and hence relevant to proving the relationship. We thus see that the evidence for relating two languages may be in only one of them, or indeed in neither but found elsewhere within the larger stock. We may state these considerations in the following way. English is a Germanic language. Hindi is an Indo-Iranian languages. Both Germanic and Indo-Iranian languages are Indo-European languages. Hence English is related to Hindi.

The introduction of other languages is also relevant to the evaluation of those resemblances that are found in both languages. Continuing with the

---

[7] See Dixon and Kroeber (1913: 649) on the Penutian languages of California: 'This relationship would have been recognized previously were it not that attention was directed chiefly toward phases of structures, which, while conspicuous, were not very typical of the group in question; and especially because comparisons had been instituted between single languages instead of the whole five.' In the first part of this citation, Dixon and Kroeber are clearly referring to typological criteria.

English-Hindi example, we might for instance note the somewhat vague resemblance between English 'tooth' and Hindi *dā:t*. Comparison with other Germanic languages, e.g. German *tsa:n* and Dutch *tand* will suggest that there was formerly a nasal consonant before English *-th*. Comparison of Hindi *dā:t*, on the other hand, with Kashmiri and Sindhi *dand* will suggest that the Hindi nasalized vowel reflects a short vowel + nasal consonant. If we introduce other branches, the additional evidence, e.g. of Italian *dente*, Spanish *diente*, makes the connection between the English and Hindi forms a virtual certainty.

Having thus arrived at an approximate source form, *dant or *dent, we now turn to the Baltic languages, where we find Lithuanian *dant-is* 'tooth' (*-is* is the nominative singular inflection), and to Modern Greek, where we find *ondi*, both easily derivable from a form stemming from the initial Germanic-Indic comparison. That all this is accidental is completely improbable. Thus even from this single comparison we will arrive at the hypothesis that Germanic, Romance, Indic, Baltic, and Greek form the nucleus of a genetic group.

Since any form can be replaced, the divergent forms in Slavic (e.g. Russian *zup*, Czech *zub*) do not disprove the affiliation of Slavic. In addition to the numerous other comparisons of the kind just discussed for other lexical items, there is obviously a special resemblance between Baltic and Slavic. Genetic relationship is plainly transitive, so that if Baltic is related to Slavic, and also to Germanic, Romance, and Indic, then Slavic must be related to Germanic, Romance, and Indic.

The addition of evidence from other highly plausible cognates, with varying distributions over the languages, at once serves to define the stock and to add powerful new evidence. We have two dimensions along which probabilities are being simultaneously multiplied. One is that of individual lexical items. Given the basic arbitrariness of the sound–meaning relationship, each word that is semantically dissimilar (e.g. 'tooth' and 'three') gives independent evidence. Using a metaphor based on the arrangement of Table 7, we may call this the horizontal dimension. In the other, vertical dimension we have the constant recurrence of the same set of languages in various combinations.

Two further observations can be made on the foregoing example of 'tooth'. First, we have compared contemporary forms. Yet many eminent comparatists claim that were it not for the existence of early written evidence from languages like Greek, Latin, and Sanskrit it would be impossible to frame or confirm the Indo-European hypothesis. The second observation is that within the etymology for the word 'tooth' there is not a single recurrent correspondence, yet it should be clear that what we have done is the initial step in the comparative method itself. We have compared related forms and posited an approximate original form and subsequent changes. We also have

correspondences. Clearly we are comparing the *t* in English 'tooth' with the *d* in Italian *dente* and not with any of the other sounds in that word.

All this runs against the current widely held view that there are two methods of classification (to which glottochronological classification is sometimes added). One is inspection, which is prescientific. By sheer luck, then, many of the classifications we hypothesize turn out to be correct, although, as has been noted earlier, the probability of chance success makes discovering a needle in a haystack a relatively simple task by comparison. Then, to be sure our hypothesis is correct, we must use the comparative method and make a reconstruction. Only with that in hand will we be sure we are correct.

I believe, however, that the preceding discussion shows the superficiality of such a view. One must look at the operations performed and not be led astray by vague terms like inspection and comparative method. To inspect languages pairwise, or at a half-guess, is a different thing from a multilateral comparison undertaken with a consciousness of the types of resemblances that are likely to bespeak common origin. As we have seen, this is the initial and in fact indispensable first step in the comparative method itself. Nor is the comparative method defined by the external, and no doubt impressive, trappings of correspondence tables and asterisks.

The preceding discussion also helps to show that the following frequently adduced negative argument, based on glottochronology, is invalid. The empirically noted fact is that in the 100-word glottochronological list, about 80 per cent of the vocabulary remains after 1,000 years. Hence, if loss is independent, the expected resemblance between two languages is .64 at this point. In the next 1,000 years, the loss for each will be $(.80)^2 = .64$ and the expected resemblance $(.64)^2$ or .4096. Continuing in this fashion, we can calculate that after 8,000 years the resemblances will be .0281, and after 10,000 years, approximately .012, which is presumably less than chance. Hence relationships of this degree of depth cannot be discovered.

One obvious flaw in these calculations is that the number of cognates will be larger than the above percentages because certain words will drop off the list owing to semantic change, but will still exist in the language, e.g. English *hound*, which is cognate with the ordinary German word for 'dog', namely *Hund*. There is also what is called the 'dregs phenomenon'. The words in the list are surely not of equal stability. Hence those that have stood the test of 8,000 years of change are far more likely to be retained during the next 1,000 years than the words of the original list during the first 1,000. Martin Joos (1964) has suggested a plausible mathematical modification that will take this factor into account.

But what is far more important, from my point of view, is the fact that through multilateral comparison we can extend glottochronological theory to account for resemblances not between two languages, but among any number of languages. For example, if we compare three languages, A, B, and C, we can ask, for a given time period, how many resemblances will be found between A and B, A and C, B and C, and A, B, and C. Clearly every word found in at least two languages of the stock can be recovered by comparison.

It is possible to combine the Joos function with the calculation of expected recoverable vocabulary by extending glottochronology from two to *n* languages. For the mathematics of these calculations, and a table of recoverable vocabulary (which is actually even higher since the factor of semantic change is not considered), the reader is referred to Appendix A[§6.1-Ed.] From this it can be seen, for example, that with only ten languages, even after 10,000 years about 42 per cent of the original vocabulary is recoverable. More languages will, of course, greatly increase these values.

Since most of the original vocabulary is thus recoverable, it is possible to carry out multilateral comparisons with other, similar stocks. There is, of course, no reason to compare just one large stock with another. As the great principle of uniformitarianism suggests, many of the linguistic stocks of an earlier period, like the present ones, could have had up to ten or more branches. Consequently, there is no theoretical limit to the depth at which classification can be carried out when the number of languages examined is large. Only at the final stage, if no subgroupings appear, will we have to resort to such considerations as the sheer number of similarities or shared grammatical irregularities.

One reason that linguists have not in general employed the methods discussed here is that in including all the languages for which material is available, much poorly recorded data will be used. This is, of course, what Bouquiaux has in mind when he refers to 'materials of unequal value'. But, one may ask, if such materials cannot be used, how could the correct and reliable results he mentions have been attained? The fact is, the method of multilateral comparison is so powerful that it will give reliable results even with the poorest of materials. Incorrect material should have merely a randomizing effect. If a clear pattern emerges, the hypothesis is all the more likely to be correct.

Moreover, it is not only possible to classify a language with very poor material; it is often possible to classify one with very little material. For in the context of a broad comparison we can discover the diagnostic items that distinguish each family and grouping, and even a very few resemblances of this sort are highly probative.

Not out of vanity, but because important questions of methodology are at stake, I should like to point to a further example to demonstrate that reliable results can be obtained from desperately poor data. In Greenberg (1953) [§1-Ed.], I briefly outlined a general classification of Australian languages. An examination of the results given there with those obtained by Stephen Wurm (1972), based on far fuller and more reliable data, shows detailed agreement. For example, my general Australian correlates exactly with his Pama-Nyungan. Yet all I had at my disposal were Edward M. Curr's very poorly recorded vocabularies (1886–7) and the two studies of A. Capell (1939–40; 1941–3), which, though more reliable phonetically than Curr, gave only minor lexical information and fragmentary grammatical data, and that for only a relatively small portion of the continent.

In all of the preceding discussion, it has been assumed that resemblances involving sound and meaning simultaneously are the only relevant ones for classification. This would include both lexical items and specific grammatical agreements, e.g. the fact that both English and German use comparative -*er* and superlative -*est* for adjective comparison. The choice of basic vocabulary for initial comparison is made for certain practical reasons. It is often the only material available. It involves that aspect of language, next to morphological markers, that is least subject to borrowing, with the advantage of length, relatively easy semantic comparability, and the essential historical independence of each item, whereas morphological systems involve relatively short items and are subject to morphological leveling. There is, however, no opposition between the two. They should and do lead to the same results, and both should be employed.

Agreement in irregularities and evidence from survivals of grammatical markers that have become petrified are worthy of special consideration and are used in the present work. An agreement like that between English *good/ better/best* and German *gut/besser/best* is obviously of enormous probative value. However, subject as such agreements are to analogical pressure, their absence is not negative evidence, and their presence tells us that there is a relationship, but not at what level. They are psychologically reassuring in showing that we are on the right track and inherently interesting, but not really necessary.

The famous comparativist Antoine Meillet, precisely because he was so aware of all the processes of diversification that give the kind of results we find in Tables 1–4, but who never thought of the simple expedient of mass comparison, believed that if we did not have earlier languages like Latin, Greek, and Sanskrit to work with, we could prove Indo-European only by such agreements in irregularities, e.g. the third person singular and plural of

the present tense of the verb 'to be': German *ist/sind*, Spanish *es/son*, Polish *jest/s*, etc.[8]

If, as I believe, our current ideas are based on a fundamental mis-understanding of Neo-Grammarian doctrine, how did these mistaken ideas arise? Only a few points can be noted here. In Terence H. Wilbur's collection of documents (1977) on the great *Lautgesetz* (sound laws) controversy of the 1880s, one finds not a hint that the genetic relationship of the Indo-European languages is involved or would be shaken if it turned out that sound laws had exceptions. Consider, moreover, the following statement by Delbrück (1904: 121–2)* in what is frequently looked on as the basic manifesto of the Neo-Grammarians:

My starting point is that specific result of comparative linguistics that is not in doubt and cannot be in doubt. It was proved [*erwiesen*] by [Franz] Bopp and others that the so-called Indo-European languages are related. The proof [*Beweis*] was produced by juxtaposing [*Nebeneinanderstellung*] 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 and uninflected words agree in their lexical parts, the assumption of chance agreement must appear as absurd.

Nowhere in this whole passage do we find the terms *Lautgesetz* 'sound law' or *Lautentsprechung* 'sound correspondence', or even *Vergleich* 'comparison', which would imply something more elaborate than mere *Nebeneinander-stellung* 'juxtaposition', seemingly the key word. Nor is the *Ursprache*, or reconstructed protolanguage, mentioned. The relationship, according to Delbrück, was proved by Bopp in the early nineteenth century before this concept was current, as it later became with the work of Schleicher.

If one considers three major examples of the reconstruction of protosound systems of non-Indo-European languages, all undertaken by linguists squarely in the Neo-Grammarian tradition, Carl Meinhof for Bantu, Otto Dempwolff for Austronesian, and Carl Brockelmann for Semitic, the results are the same. As we have seen, Schachter and Fodor, among others, believed that Meinhof demonstrated the relationship of the Bantu languages by his reconstruction. A reasonably careful reading of what Meinhof and the others say shows that this idea never occurred to them. Meinhof simply states (1932: 18): 'First of all,

---

[8] Resemblances involving sound only, or meaning only, and relations of order are of typological, not genetic relevance. This point is, I believe, generally understood at present and is therefore not discussed here. For previous discussions see Greenberg (1953, 1957, 1963) [§§1, 2, 5-Ed.].

* Greenberg cites the fourth edition in the references but uses the date and the title of the first edition; the quotation is found only in the (renamed) fourth edition of this work. The citation and reference are corrected here (and in §11).-Ed.

the characteristics which were found to be common to a great number of languages in Europe and Asia led to these languages being grouped together as the Indo-European "family" of languages. Since then we have also recognized the existence of other language families in Europe and Asia as well, e.g. the Semitic, the Malayo-Polynesian, the Bantu.' As noted earlier, Meinhof believed there will always be a number of exceptions to sound laws for which no explanation can be found.

In the first volume of his great comparative work on Austronesian, Dempwolff (1934: 17) talks about 'sound correspondences that in many cases are regular.' When his reconstruction, based on Indonesian languages in the first volume, is applied to Fijian in the second volume, he finds some remarkable deviations. These mostly involve consonants, which are lost in word final position, but reappear when suffixes are present. He notes (1937: 133) that 'statistically, if one considers all the words likewise found in proto-Indonesian, there are 59 cases of correct correspondence and 49 deviations.' In fact, for two of the consonants the number of exceptions is greater than the number of regular cases. Why does he then call them regular? Because they are phonetically plausible and coincide with correspondences in other word positions. He then states: 'On the basis of these facts, these irregularities in Fijian are to be interpreted as false analogy.' This term is the equivalent of what is now usually called analogy. That he should abandon the relationship to Indonesian because of this never even occurs to him.

As for Brockelmann, in his comparative Semitic grammar of 1908, far from thinking that he has proved that the Semitic languages are related by means of reconstruction, he denies its very reality. This fictionalist, or conventionalist, view of reconstruction, also found in Meillet, sees reconstruction as merely a convenient set of formulae to show the relations among forms. An allied, but somewhat different, view is that the reconstruction is hypothetical—our best, but provisional, approximation to reality on the basis of present knowledge. It thus approaches the truth with increasing knowledge, but can never be completely certain. As noted by Allen (1953: 70), 'This example shows how the addition of fresh languages adds to the comparative formulae.'

Even if one were to find documentary evidence of what looks like a candidate for the *Ursprache*, it would surely not coincide exactly with the current reconstruction, showing either that the reconstruction is not entirely correct or that it is not really the *Ursprache*, but some extinct side branch. This is the case with Geez (Classical Ethiopic), earlier viewed as equivalent to the ancestral language of the present Semitic languages of Ethiopia. However, because it shows some innovations not reflected in the modern languages, it is now generally believed not to be ancestral.

That reconstruction is not necessary to prove relationship is shown by the interesting case of Albanian. Its status as an Indo-European language is universally accepted. Earlier, Delbrück was cited as stating that Bopp and other pioneers had proved that the Indo-European languages formed a family. In an immediately preceding passage he lists these and includes Albanian. There had, in fact, been doubt concerning Albanian, but it was agreed that Bopp had proved the case for its Indo-European affiliation in a monograph published in 1854. An examination of this work shows a number of things of interest in the present connection. Bopp rests his case mainly on specific morphological resemblances, although he cites a number of lexical items in the course of his exposition and obviously considers them of great importance too. He does not derive Albanian forms from reconstructed Proto-Indo-European. The notion of family tree and protolanguage both originated with Schleicher, but only several years later.

Bopp realizes that Sanskrit has in some instances undergone changes, but evidently considers that it represents the original form in almost all instances, although in etymologies he generally cites Latin, Greek, Gothic, and Lithuanian alongside Sanskrit. He has no term for sound law or sound correspondence. He realizes that there are certain common correspondences, but it is clear that he considers them merely usual and is not disturbed by deviations. In fact, he generally does not even notice them.

In place of regular sound change, he operates with three concepts, all of them value-laden, and none of them implying regularity. If, in comparison with Sanskrit, the Albanian form is longer, then there has been an addition that is *unorganisch* 'inorganic'. If it is shorter it is *verstummelt* 'mutilated'; if it is of the same length, but has a different sound from Sanskrit in some position it is *entartet* 'degenerate', a term used in plant and animal breeding.

We might expect that given this, to us, quaint conceptual apparatus, Bopp's etymologies would be commonly invalid. Not so. Of the first twenty-five etymologies I encountered in the text, twenty-three appear in Walde-Pokorny's etymological dictionary of Indo-European. The other two are not just accidents but borrowings—from Latin in one case, Italian in the other. Actually, Bopp was quite aware of these borrowings and often identifies them correctly.

More than thirty years after Bopp's monograph, Brugmann noted, of Albanian, that 'the historical treatment of this language is beset with many difficulties and is still in its infancy' (1886, vol.1: 7). In the 1880s and 1890s Gustav Meyer began his fundamental work on the historical grammar of Albanian. Yet in Brugmann's *Kurze vergleichende Grammatik* (1902) Albanian appears only in the table listing reflexes of the stressed vowels, largely with

unexplained multiple values for each proto-Indo-European vowel, which are neither referred to nor explained in the text. Meillet simply takes no account of Albanian in his well-known *Introduction*, even in the last edition actually prepared during his lifetime, namely, the seventh edition of 1934, just eighty years after Bopp's original study.

Are recurrent correspondences and reconstruction of no value at all, then, as a superficial reading of this chapter might suggest? The answer can be put as follows. The existence of the same correspondence in several different etymologies certainly adds to the probabilities of each being correct. Moreover, such correspondences are our chief methodological tool in reconstruction. However, what many linguists fail to appreciate is that anything approaching a complete and highly convincing reconstruction on the basis of recurrent correspondences is in general possible only with languages so closely related that it is unnecessary anyway. Even here we have cases like Athabaskan in which the reconstructions not only differ, but are confined to initial and nonfinal consonants. Where the separation is greater, as we have seen, the reconstructions are so underdetermined by the data that deviations from a particular theory of reconstruction can be accommodated by a whole series of strategies. Some etymologies will always remain uncertain. In others the lesser claim of cognation can be maintained, even though the reconstructive explanatory theory remains uncertain or in dispute. A far more convincing refutation is to show an incorrect morphological analysis, not deviation from a predicted phonological outcome.

It will give us a broader perspective on the problems discussed in this chapter if we consider that, whereas linguists have treated the problem of historical classification as unique to linguistics, and have in part at least suggested idiosyncratic solutions, the same basic model occurs in a whole series of other cases in different disciplines; a comparison can prove enlightening.

One of these is biological classification. Both in evolutionary biology and in comparative linguistics, hierarchical classifications are explained by the dynamic process of successive splits from common ancestors of differing degrees of historical depth. This was well recognized in the nineteenth century by Darwin and Lyell on the biological side, and by Schleicher on the linguistic side. In his *Descent of Man* (1871: 20) Darwin notes that 'the formation of different languages and distinct species, and the proof that both have developed through a gradual process, are curiously parallel.' Since then Otto Jespersen, Raimo Anttila, the present writer, and no doubt others have discussed this similarity.

I do not believe that anyone would argue that biological classification is less advanced than linguistic classification. Yet nothing in biology seems

equatable with sound correspondence. Moreover, even in that science common ancestors can be reconstructed only in the most tentative way. The biologist of course has the direct evidence of fossils (though a zoologist of my acquaintance once told me that fossils were always surprising). Beyond that, the biologist, much like the linguist trying to identify an ancestral language from earlier written forms, faces the problem of deciding whether a particular fossil represents the ancestor of a particular taxonomic group or a side branch that is now extinct, a process fraught with the same sorts of difficulties encountered by the linguist.

One may note that biologists are very clear on the distinction between hierarchical classification and genetic relationship, and the primacy of the former. No biologist has yet, to my knowledge, written a treatise directly comparing a beaver to a mackerel in order to prove their genetic relationship, both being vertebrates. Fortunately, Linnaeus, and even his predecessors, got biological classification off on the right foot by boldly surveying and classifying all forms of life as part of the same operation.

Another comparison that might be made is with textual criticism. Henry Hoenigswald even suggests that Schleicher, the first to represent linguistic classifications by a branching tree, was inspired by his training in classical philology under Friedrich Ritschl, who depicted the relationship of manuscripts of the same work as a stemma or family tree. The splitting of a language over time into distinct languages corresponds to the various copies of the same manuscript, which thus form a subgroup defined by this process.

To say that manuscripts are genetically related without specifying their position in the fully worked-out stemma, which corresponds to genetic linguistic classification, is simply to say that they are manuscripts of the same work. A statement parallel to that often made by linguists would be the following. Though they look suspiciously similar, we really have no right to say that two manuscripts of, say, Sophocles' play *Ajax* are really manuscripts of the same play, that is, go back to a common original, until we have reconstructed the original text. But, of course, that they are manuscripts of the same play is so obvious that no one even discusses the question, while the problem is precisely to restore the original text. This is the goal of textual criticism, and one that will probably never reach a completely certain and universally accepted outcome.

One can also make the profound observation that whenever there is an epsilon at a particular position on a particular line of one manuscript, one normally finds one at the corresponding place in the other manuscript. A few very rigorous thinkers would then claim that if this sometimes does not hold,

there is an irregular correspondence; and thus we could never prove that they are copies of the same work.

In actual fact, however, it is the classification of manuscripts in the form of a hierarchical tree or stemma, in which those copies from the same earlier manuscript form a subgroup, that is basic to reconstruction. The copies within a subgroup do not give independent witness regarding the original text, and if we happen to have the manuscript from which they were copied, they can be eliminated. Moreover, it is precisely the agreement in random errors of the type unlikely to occur independently that provides the basis for classifying manuscripts in the same group. These errors correspond to arbitrary irregularities in language, not to recurrent regular correspondences.

Another example close to linguistics is writing systems. These can also be classified genealogically. Thus, all alphabets descended from the North Semitic version of the alphabet can be distinguished as a group from those like Minean and Ethiopic, which derive from the South Semitic alphabet, as anyone can see from mere inspection. Yet the arbitrary pairings of sound and visual symbol are very few in number compared with the lexicon of a language, the changes of the letters do not seem to exhibit any recurrent regularities, and the original alphabet cannot be reconstructed.

In fact, even in linguistics, only for *phonological systems* is it claimed that regular correspondence or reconstruction is necessary for proving relationships or verifying classifications. Such requirements are never invoked for morphology, syntax, or semantics, yet in Indo-European it was the numerous points of specific contact in *morphological systems* that played the major role at an early stage in regard to these questions. Although the morphological systems clearly are related and must derive from a common original, probably no one would claim that we have a reliable and detailed reconstruction of the ancestral morphological systems. With the discovery of Hittite, the situation worsened. While, on the one hand, Bedrich Hrozný's original demonstration leaned heavily on both morphology and vocabulary (and most of his morphological equations have stood the test of time), on the other, the addition of Hittite had a devastating effect on the validity of many of the details of the original morphological system as reconstructed at that time.

There has grown up, as a corollary of present doctrine, the notion that one must first reconstruct the protolanguage of each lower grouping, thus proving its validity, before proceeding to the reconstruction of the higher-level groupings. Such a stepwise procedure appears to be very virtuous, but in fact is an illusion. The reconstruction will itself be a poorer approximation to the truth if it is confined to a restricted group. In the model case of Indo-European, it was the broader group that was first reconstructed. In fact, many

phenomena of narrower groups can only be understood historically by outside evidence from within the broader stock. An example is the Germanic consonant alternations that arose as a result of the change expressed in Verner's law. These alternations require a knowledge of the Indo-European pitch accent that did not exist in Germanic in the historic period. After all, in historical matters, the earlier explains the later. If the earlier is not directly attested, one looks backward by looking sideways, which is precisely the comparative method.

The broad approach advocated here does not require the reckless positing of risky and uncertain etymologies. All that is needed is to show decisively more cognates than those of any rival hypothesis. Actually, it is better to start with a relatively small number of first-rate etymologies. These are the ones that establish the classification. One may then discover others, given a valid initial hypothesis. In such work a wider comparison often invalidates etymologies erected on too narrow a base.

In the absence of a principled theory on how many etymologies are necessary, many comparatists seek to produce as many as possible. I myself have in several instances worked out etymologies for specific groups of Amerindian languages only to find later that someone has compared at least some of these languages and published a set of etymologies, complete with asterisks and tables of correspondences. I have found in every case that I have fewer etymologies, and that almost all, if not all, are contained in the larger list. This is partly because I have not included much in the way of kinship terms or zoological and botanical terms in my word lists, but also because I have excluded many words that I consider too venturesome or improbable in the light of wider evidence.

Of course, most sound changes are regular, and some competing theories of reconstruction are more plausible than others. Though there is much indeterminacy in reconstructed sound systems, some features do seem quite secure. Besides, for all of these uncertainties, I am by no means saying that comparative linguistics is an idle endeavor. On the contrary, it is an indispensable means toward the attainment of two major goals. The first lies at the very heart of linguistics as a scientific enterprise. This is the understanding of linguistic change, with the ultimate aim of diachronic generalizations. Without the comparative method, we would be confined to the relatively narrow limits set by the existence of written records and to the here-and-now of internal reconstruction.

The other goal is the contribution that the comparative method can make to nonlinguistic history, a matter of greater inherent interest to linguists than might meet the eye. After all, language is our major instrument for dealing

with the external world. Such topics as semantic change cannot be fruitfully pursued without a knowledge of this world. But again, though such endeavors as genetic classification are, in a sense, a kind of scientific obligation that we owe to the related historical disciplines, they are not a mere favor to them, but an indispensable part of linguistics itself.

6.1 Appendix: A generalization of glottochronology to *n* languages[*]

With a large proportion of the vocabulary recoverable by comparison even after a great lapse of time, it becomes possible to compare genetically defined groups of languages with other, similar groups to arrive at still deeper classifications.[9] The general mathematical model outlined below shows how multilateral comparison increases the proportion of recoverable vocabulary as the number of languages compared increases. In fact, multilateral comparison is an even better tool than an examination of the model might suggest, for the 'semantic criterion' is applied strictly in glottochronology; in a standard list of 100 or 200 basic words, only those pairs that are the 'best translation' of a given item are counted as cognates. But because semantic change, like other processes of change, has a cumulative effect over time, many more cognates are in fact recoverable even when they have undergone semantic shifts. An example is the so-called *Hund*-'dog' phenomenon. In English, the word *hound*, cognate with German *Hund*, still exists but does not appear in a glotto-chronological comparison.

The theory behind glottochronology is that if we take a list of, say, 200 basic words (and such a list has been devised), the rates of retention of words on this list in their original meanings over a given time period will be reasonably similar in historically independent cases. By examining a number of documented cases, a retention rate variously calculated at .80–.85 for 1,000 years has been ascertained (with a probable error based on the variability among the test cases); the value .80 is the one used here. If, after 1,000 years, 80 per cent of the word list has been retained, then in the next 1,000 years 80 per cent of what is left will be retained, so that the rate of retention of the original vocabulary is .80 × .80 or .64 ($.80^2$). In general, if t is the number of millennia, the retention rate is the exponential function $.80^t$.

If we now compare two languages that have a common origin, each replacing vocabulary independently at this rate, then after t millennia the common vocabulary will be $.80^t \times .80^t$ or $.80^{2t}$. All glottochronological calculations made hitherto are

[9] All calculations in Appendix A [§6.1-Ed.] and Appendix B [in Greenberg (1987)-Ed.] were carried out on an HP 11C and all the programs were written by myself. I am indebted to James A. Fox for independently carrying out all the calculations in Appendix A and those in Tables B.1–B.24 (as well as the generating function in Appendix B), using BASIC as his programming language.

based on this formula (with the constant varying, as indicated, from .80 to .85). It is possible, however, to generalize from two to n languages. Suppose, to take the simplest case, that n is three. Call the languages A, B, and C. Then, after t millennia, the original vocabulary falls into 2^3, or eight parts. First, there is the vocabulary not found in any of the three languages. Let us call this the *unrecoverable vocabulary*. Then, we have the vocabulary surviving in A but not in B or C, in B but not in A or C, and in C but not in A or B. This vocabulary, which cannot be recovered by comparison, is the *submerged vocabulary*. The fifth, sixth, and seventh parts are the vocabulary surviving in two languages, that is, A and B, A and C, and B and C. Finally, as the eighth class, we will have the vocabulary surviving in all three languages, A, B, and C.

Note that the division of the original vocabulary into eight parts for $n=3$ is $\binom{3}{0} = 1 + \binom{3}{1}$ (i.e. 3 words taken 1 at a time) $= 3 + \binom{3}{2}$ (i.e. 3 words taken 2 at a time, or $3 \times 2 / 1 \times 2) = 3 + \binom{3}{3} = 1$, so that $1 + 3 + 3 + 1 = 8$. But this is simply the well-known binomial expansion for the special case $n=3$, for which the general formula is the following:

$$1 = p^n + np^{n-1}q + (n/2)p^{n-2}q^2 \ldots npq^{n-1} + q^n$$

If we take p as the retained vocabulary and q as the replaced vocabulary, then npq^{n-1} represents the submerged vocabulary and q^n the unrecoverable vocabulary. The vocabulary recoverable by comparison, then, is the total vocabulary minus the sum of the last two terms, i.e. $1 - (npq^{n-1} + q^n)$. Clearly, as the number of languages n increases, more and more of the lost vocabulary becomes merely submerged, and more and more submerged vocabulary becomes recoverable by survival in the additional languages.

Another relevant factor can be included in the calculation. We have assumed that all the items on the vocabulary list are equally stable. But this assumption is clearly improbable. If languages have been separated for, say, 5,000 years, what they still have in common will generally be the more stable elements. Hence a smaller proportion of items should be lost in the next 1,000 years than in the previous 1,000 years. In other words, we have a constantly decelerating rate of vocabulary loss. Failure to take this so-called dregs effect into account results in an underestimation of the retention rate.

A proposal to deal with the dregs effect is found in Joos (1964). The proposal is to split the standard list into eight sublists according to a normal frequency distribution. The empirically ascertained rate of .80 for 1,000 years is then the sum of 2 per cent of the list with r (retention rate) $= .96$; 7 per cent with $r = .93$; 17 per cent with $r = .89$; 24 per cent with $r = .84$; 24 per cent with $r = .78$; 17 per cent with $r = .71$; 7 per cent with $r = .63$; and 2 per cent with $r = .54$.

The binomial theorem can be combined with the Joos function into a function that takes account of both variables, the number of languages, n, and the time in

millennia, *t*. To illustrate, I give the first two terms of the function for five languages after 3,000 years (i.e. $n = 5$, $t = 3$):

$$1 - \{(.02 \times [5 \times .96^3 \times .04^4 + .04^5]) + (.07 \times [5 \times .93^3 \times .07^4 + .07^5]) + \cdots\}$$

Table A.1 shows the proportion of vocabulary expected to be recovered for two to ten languages after periods of 1,000 years to 10,000 years, as calculated by the Joos function. The retention rates for 20,000 years and for twenty languages are shown to give a sense of the expected outcomes after still longer periods of time and with larger numbers of languages. In Table A.2, the expected proportion of recoverable vocabulary without compensating for the dregs effect is calculated for purposes of comparison.

TABLE A.1 Recoverable vocabulary based on the Joos function

| Years | Number of languages | | | | |
	2	3	4	5	6
1,000	.648	.882	.959	.985	.994
2,000	.440	.700	.835	.906	.945
3,000	.310	.542	.693	.788	.851
4,000	.226	.420	.563	.666	.740
5,000	.168	.328	.456	.555	.631
6,000	.128	.259	.370	.461	.534
7,000	.100	.207	.302	.383	.451
8,000	.079	.167	.248	.319	.381
9,000	.063	.136	.206	.268	.323
10,000	.051	.112	.171	.226	.274
20,000	.010	.023	.038	.053	.068

| Years | Number of languages | | | | |
	7	8	9	10	20
1,000	.998	.999	1.000	1.000	1.000
2,000	.967	.979	.987	.991	1.000
3,000	.892	.921	.941	.955	.995
4,000	.794	.834	.865	.889	.976
5,000	.690	.737	.775	.806	.937
6,000	.594	.643	.683	.717	.883
7,000	.508	.556	.597	.632	.819
8,000	.433	.479	.519	.554	.751
9,000	.371	.413	.450	.484	.683
10,000	.318	.356	.391	.422	.618
20,000	.083	.096	.110	.122	.220

Table A.2. Recoverable vocabulary based on a homogeneous replacement rate

Years	Number of languages				
	2	3	4	5	6
1,000	.640	.896	.973	.993	1.000
2,000	.410	.705	.864	.940	.975
3,000	.262	.518	.705	.827	.901
4,000	.168	.366	.541	.679	.781
5,000	.107	.252	.397	.528	.685
6,000	.068	.170	.282	.393	.495
7,000	.044	.114	.196	.283	.369
8,000	.028	.075	.134	.198	.266
9,000	.018	.049	.090	.136	.187
10,000	.011	.032	.060	.093	.129
20,000	.000	.000	.001	.001	.002

Years	Number of languages				
	7	8	9	10	20
1,000	1.000	1.000	1.000	1.000	1.000
2,000	.989	.996	.998	.999	1.000
3,000	.945	.970	.984	.991	1.000
4,000	.854	.903	.937	.959	1.000
5,000	.726	.796	.849	.889	.996
6,000	.585	.662	.728	.782	.981
7,000	.450	.525	.593	.653	.943
8,000	.333	.399	.461	.519	.872
9,000	.258	.293	.345	.396	.770
10,000	.168	.209	.250	.293	.649
20,000	.003	.004	.005	.006	.022

References

Achinstein, Peter. 1977. 'History and philosophy of science: a reply to Cohen', in Frederic Suppe (ed.), *The Structure of Scientific Theories* (2nd edn.). Urbana: University of Illinois Press, 350–60.

Allen, W. S. 1953. 'Relationship in comparative linguistics', *Transactions of the Philological Society* 1953: 52–108.

Boas, Franz. 1929. 'Classification of American Indian languages', *Language* 5: 1–7.

Bopp, Franz. 1854. 'Über das Albanische in seinen verwandschiftlichen Beziehungen', *Akademie der Wissenschaften Berlin, Abhandlungen Phil.-Hist. Klasse*: 459–549.

Bouquiaux, Luc. 1967. 'Le système des classes nominales dans quelques langages (birom, ganawuri, anaguta, irigwe, kaje, rukuba, appartenant au groupe "Plateau"

(Nigeria centrale))', in Gabriel Manessy (ed.), *La classification nominale dans les langues négro-africaines*. Paris: Centre National de la Recherche Scientifique, 133–56.

Brockelmann, Carl. 1908. *Grundriss der vergleichende Grammatik der semitischen Sprachen*. Berlin: Reuther & Reichard.

Brugmann, Karl. 1886. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*, vol. 1. Strassburg: Trübner.

—— 1902. *Kurze vergleichende Grammatik der indogermanischen Sprachen*, vol. 1. Strassburg: Trübner.

Capell, A. 1939–40. 'The classification of languages in north and northwest Australia', *Oceania* 10: 241–72, 404–33.

—— 1941–3. 'Languages of Arnhem Land, north Australia'. *Oceania* 12: 364–92; 13: 24–50.

Curr, Edward M. 1886–7. *The Australian Race*. Melbourne: Ferres.

Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. London: Murray.

Delbrück, Berthold. 1904. *Einleitung in das Studium der indogermanischer Sprachen*. (4th and renamed edn. of *Einleitung in das Sprachstudium*, 1880.) Leipzig: Breitkopf & Härtel.

Dempwolff, Otto. 1934–7. *Induktiver Aufbau eines indonesischen Wortschatzes* (2 vols.). Berlin: Reimer.

Dixon, Roland B. and Kroeber, Alfred L. 1913. 'New linguistic families of California', *American Anthropologist* 15: 645–55.

Dyen, Isidore. 1959. 'Review of Greenberg, *Essays in Linguistics*', *Language* 35: 527–52.

Fodor, István. 1966. *The Problems in the Classification of the African Languages*. Budapest: Hungarian Academy of Sciences.

Forrest, R. A. D. 1950. 'The Ju-Sheng tone in Pekingese', *Bulletin of the School of Oriental and African Studies* 13: 443–7.

Gabelentz, Georg von der. 1891. *Die Sprachwissenschaft*. Leipzig: Weigel.

Gleason, Henry Allen. 1955. *Workbook in Descriptive Linguistics*. New York: Holt.

—— 1959. 'Counting and calculating for historical reconstruction', *Anthropological Linguistics* 1/2: 21–32.

Greenberg, Joseph H. 1963. *The Languages of Africa*. Bloomington: Indiana University.

—— 1978. 'How does a language acquire gender markers?' in Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravesik (eds.), *Universals of Human Language, vol. 3: Word Structure*, Stanford: Stanford University Press, 47–82.

Gregersen, Edgar A. 1977. *Language in Africa*. New York: Gordon & Breach.

Guthrie, Malcolm. 1967. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of Bantu Languages*, vol. 1. Farnborough, Eng.: Gregg Press.

Jespersen, Otto. 1922. *Language: its Nature, Origin and Development*. London: Allen & Unwin.

Joos, Martin. 1964. 'Glottochronology with retention rate homogeneity', in Horace G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists*. The Hague: Mouton.

Kroeber, Alfred L. 1906–7. 'Shoshonean dialects of California', *University of California Papers in American Archaeology and Ethnology* 4: 65–165.

Lamb, Sydney M. 1959. 'Some proposals for linguistic taxonomy', *Anthropological Linguistics* 1/2: 33–49.

Lane, George S. 1968. 'The beech argument: a re-evaluation of the linguistic evidence', *Zeitschrift für vergleichende Sprachforschung* n.f. 81: 198–202.

Levine, Robert D. 1979. 'Haida and Na-Dene: a new look at the evidence', *International Journal of American Linguistics* 45: 157–70.

Meillet, Antoine. 1934. *Introduction à l'étude comparative des langues indo-européennes.* Paris: Hachette.

Meinhof, Carl. 1910. *Grundriss einer Lautlehre der Bantusprachen,* (2nd edn. revised and expanded). Berlin: Reimer.

——— 1932. *Introduction to the Phonology of the Bantu Languages.* Berlin: Reimer/Vohsen.

Newman, Paul. 1970. 'Historical sound laws in Hausa and in Dera (Kanakuru)', *Journal of West African Languages* 7: 39–51.

Rensch, Calvin. 1976. *Comparative Otomanguean Phonology.* Bloomington: Indiana University Press.

Saussure, Ferdinand de. 1854. *Mémoire sur le système primitif des voyelles dans les langues indo-européennes.* Leipzig: Teubner (1879 edition).

Schachter, Paul. 1971. 'The present state of African linguistics', in Thomas A. Sebeox (ed.), *Current Trends in Linguistics, vol. VII: Linguistics in Sub-Saharan Africa.* The Hague: Mouton.

Sturtevant, Edgar H. 1942. *The Indo-Hittite Laryngeals.* Baltimore: Linguistic Society of America.

Szinnyei, József. 1910. *Finnisch-Ugrische Sprachwissenschaft.* Leipzig: Goeschen.

Welmers, William E. 1973. *African Language Structures.* Berkeley: University of California Press.

Wilbur, Terence H. (ed.). 1977. *The Lautgesetz Controversy: A Documentation (1885–86).* Amsterdam: John Benjamins.

Wurm, Stephen A. 1972. *Languages of Australia and Tasmania.* The Hague: Mouton.

This page intentionally left blank

7

On lumping and splitting in linguistics and biology*

As a preliminary and purposely vague characterization of the opposition between lumping and splitting, a currently popular terminology in a number of fields, we may say that in all instances we have to deal with classification of a set of objects which have some common defining feature, e.g. being a language, being a species, etc. Lumping, then, is a tendency to classify into relatively few groups, each with a relatively large membership, while splitting is a tendency to distinguish many groups each with a relatively small membership.

The two major fields in which this terminology is used are biological taxonomy and linguistic 'genetic' classification into families, each of which consists of a group of languages whose resemblances are to be explained by differential development from a common original, the heart of Darwinian biological evolution as opposed to varying versions of creationism. This kind of explanation actually appeared earlier in linguistics in that the commonalities of the Indo-European, and even earlier in the late eighteenth century, of the Finno-Ugric languages, received a commonly accepted explanation in terms of dynamic development from a common original before such explanations were accepted in Darwinian biology.

The vogue of the lumping vs. splitting analogy is much more recent and does not, as far as my knowledge goes, antedate the twentieth century and first appeared in biology. Thence it was transferred to linguistics. However, as has hardly ever been noted, this analogy as it occurs in biology, differs from its application in biology in one very basic way. In biology, even the most inveterate splitter does not deny the common origin of truly vast assemblages of superficially very different life forms. Such a group, for example, is the eukaryotes, which includes groups as different as horses and trees. Indeed, the existence of a common origin of all life forms is probably accepted by the

majority of biologists, nor does an assertion of this sort produce outraged statements by those who may differ on this matter and do not consider such a unity proved, as does an analogous statement in linguistics.

The lumping vs. splitting opposition, then, in biology rather refers to the fineness of distinction required to assign different taxa taxonomic status within recognized larger groups. The situation in linguistics is very different. Those who split groups of languages from each other deny their provable ultimate common origin, unlike the situation in biology. Moreover, for many linguists lumping and splitting has been identified with scientific lack of caution and caution respectively. Indeed, for certain linguists, apparently all 'splitters', these represent two different types of linguists. For example, Matisoff contrasts microlinguists and macrolinguists and he takes pride in being one of the former. Doerfer, the leader of a movement to dismantle Altaic, labels any attempt to compare larger groups of languages with each other as *Omnicomparatismus*, presumably a hateful and proscribed activity. All such 'splitters' apparently consider that classification to be preferred which posits as many independent entities as possible on the grounds that it is the most 'cautious' and hence the most scientifically accurate.

But is the positing of as many independent families as possible the most cautious? Caution, it should be obvious, is not correlated with the number of independent families being proposed. Caution is a relation between evidence and conclusions. If one dismantled Indo-European and said that there were approximately sixteen historically independent entities this would not be caution but simply error. The vast number of intercrossing resemblances in sounds and meaning among the Indo-European subfamilies would simply remain unexplained even though an obvious and now universally accepted way of accounting for them exists, namely that they are branches of a single family. And why stop there? A classification in which each Romance language was a separate and independent entity would be even more cautious, and one would add, even more incorrect.

The mere number of independent entities posited is not then an indicator of caution and hence scientific certainty. In my first work on African languages I posited sixteen separate stocks for the continent. Later I reduced these to four by a comprehensive comparison of these sixteen stocks. These are now universally accepted. Was I a splitter who then became a lumper by reducing the number of stocks? The reduction was made because without it numerous and significant resemblances would not be accounted for. Neither a splitter nor a lumper be, rather try to account for the facts by a classification which best explains the evidence. This does not by any means exclude a fine-grained set of divisions of languages within the larger stock by a hierarchical

set of divisions of successively lower levels which in principle is like the biological taxonomy of successively finer grades or levels such as kingdom, phylum, class, order, family, genus, species, and subspecies, reflecting the age of the common ancestor from oldest to most recent. So, my classification of the Bantu group as a subgroup of a subgroup of a subgroup of a subgroup of a subgroup within Niger-Kordofanian, involving a detailed splitting with more and more circumscribed subgroupings, was not carried out in a similar way in any other major subgrouping of Niger-Kordofanian because the special case of Bantu, which occupies roughly one-third of southern Africa, was particularly significant from a historical point of view while the difficulty of doing it was trivial although it shocked Bantuists who lacked a historical frame of reference (though not archaeologists or biological anthropologists!). Considering the evolving state of linguistic classification during the present period, I would prefer to keep the relative remoteness of common descent in a hierarchy without freezing the grades by adapting biological terms, until perhaps our taxonomy is complete as we trust it ultimately will be with profound consequences for our knowledge of human history.

The general notion of relative fineness of taxonomic distinction, which is only partly captured by the notion of lumping and splitting which over-simplifies by setting up polar opposites, is of course applicable in all areas of human thought where classification, a truly basic human mental activity, is involved. Thus, in archaeology, when we set up a classification of artefacts by relative similarity and probable function, we also encounter as opposite tendencies the putting together of never completely identical artefacts into numerous smaller groupings as against a smaller number of broader groups. Here also there are historical interpretations involved. Archaeological cultures are defined by commonalities in their typologies. What one archaeologist who tends to posit fewer groups and larger differences within groups, may define as a widespread culture with possible correlates in physical type, language, etc., may for another archaeologist represent a larger number of independent and historically unconnected cultures.

It is clear that, as noted before, the setting up of classifications involving a number of taxonomic levels is a fundamental human thought process and is a basic activity in a variety of fields. Each field deserves its separate and appropriate analyses. The relation of this activity to historical process is ultimately our basic reason for carrying it out (except of course if we are creationists!).

This page intentionally left blank

8

The concept of proof in genetic linguistics*

8.1 The notion of proving a relationship

Virtually all historical linguists share a common notion which may be stated in something like the following terms. One starts with a hypothesis of relationship between two or more languages or groups of languages. If groups of languages are involved these in turn are conceived of as a number of languages which already have been proven to be related, e.g. the Indo-European family. Taking as examples binary hypotheses (these are the most frequent in the literature) we may illustrate the three possibilities by actual examples. An instance in which we deal with two single languages is the Japanese-Korean hypothesis, of a single language with a group of languages, the Eskimo-Indo-European hypothesis and of one group with another group the Indo-European-Uralic hypothesis. As an attentive reader may already have noted, in formulating it this way, what we call a group or distinct language is often a relative matter. Virtually every language has internal dialect divisions and the point at which we talk of a group or family as distinct from a single language is to some extent arbitrary. Moreover, some of these hypotheses as usually stated take some single language as part of the hypothesis because it is more important or better known and disregard the group of which it is a member. Thus Uhlenbeck who championed the Eskimo-Indo-European hypothesis treated Eskimo *as though* it were a single language using West Greenlandic, the best studied form, as representative of Eskimo as a whole. In fact, Eskimo contains at least two units, Inuit and Yuit which certainly deserve to be called separate languages. In addition, Eskimo is universally recognized as having Aleut as its nearest relative in a language family called Eskaleut. It is noteworthy in these typically binary hypotheses, that less important

languages are generally disregarded. Thus, Aleut is almost never compared with anything else while the more prominent Eskimo is. So also for the purposes of Japanese–Korean comparison, the dialect group found in the Okinawa Islands and which certainly deserves the status of a separate language is generally disregarded in comparisons of Japanese with other languages or language groups.

8.2 Probability versus demonstration

When hypotheses such as those mentioned in the previous section are advanced, what is sought is 'proof', and indeed phrases like 'to demonstrate' the relationship of A to B, a terminology which is, of course, ultimately taken from geometry, are frequently used. What is sought is certainty of the kind attainable in mathematics and logic. I believe there is at least a dim realization that in all empirical sciences, as against logic and mathematics, in which truths flow infallibly and tautologically from definitions, all that we *can* get are results so close to certainty that for all practical purposes we can consider them certain, that is, a hypothesis which is overwhelmingly better than any other. In the search for infallibility, certain criteria have been advanced which some linguists think will bring them results which are tantamount to certainty. The two most popular are the existence of sound correspondences and the reconstruction of a proto-language from which the languages hypothesized to be related can be derived. Both of these will be considered later. For the moment, it is to be noted that the two approaches involve somewhat different metaphors from different nonlinguistic fields. Sound correspondences are often called laws, or more exactly, the sound changes that give rise to the correspondences, are the so-called 'sound laws'. Here the analogue is to the laws of nature, which like sound changes brook no exceptions. They thus vicariously share the certainty and prestige of such laws as the laws of gravitation, one which is of course empirically founded. The metaphor of reconstruction is more like that of logic. The reconstructed forms are like postulates, and the changes like laws of deduction. Thus, the forms to be explained are derived by a procedure which has some analogy to that of deductive logic and in this way participates in its prestige and certainty.

8.2.1 *Analytic and synthetic truth*

What we have been sketching is, of course, akin to the famous Kantian distinction between analytic propositions, true by definition and synthetic, drawn from experience and therefore only subject to lesser or greater degrees

of confirmation. All the twisting and turning in the world cannot make reconstruction or sound laws give the complete certainty attaching to the truths of the disciplines of logic and mathematics.

It is crucial to note that the view of genetic linguistics as concerned with proving relationships and the notion of certain procedures as providing such proof are shared by the 'conservatives' who believe that there are a large number of independent, or at least not provably related families in the world, and those who undertake long-range comparisons like the Nostraticists. It is just that the 'proofs' of the latter are not accepted by the former as adequate. The Nostraticists, in particular, tend to belong to the camp of the reconstructionists, that is, those who believe that a reconstruction of an ancestral language proves the relationship of the descendant languages.

8.2.2 *The relativity of proof*

The existence of numerous points of dispute concerning language classification so well-known to all historical linguists has just been alluded to. What it shows is that comparative linguists do not have a well-formulated and generally agreed-on notion of what in fact constitutes proof of a hypothesis of relationship.

If one extends one's view backward from contemporary linguistics to earlier periods the differing notions of what proof is become even more striking. Consider the following statements from an earlier but not remote period regarding the Finno-Ugric family, actually along with Samoyed, a subgroup of Uralic, a family universally accepted.

The first comes from Kai Donner (1901: 129), an eminent Finno-Ugricist and the founder of the journal *Finnisch-Ugrische Forschungen*: 'Through J. Sajnovics (*Demonstratio* 1770) and S. Gyarmathi (*Affinitas* 1799) Strahlenberg's well-founded assertion regarding the Finno-Ugric (Uigur) group was proven once and for all.' The reference here is to a work published by Strahlenberg in 1730 [see §2-Ed.].

Later than Donner, a well-known Finno-Ugricist, Ravila (1935: 21), stated that:

In 1770 the Hungarian Sajnovics published his famous *Demonstratio*, in which, using quite modern methods he proved the relationship of Hungarian to Lapp. By this and by the *Affinitas* of Sámuel Gyarmathi which appeared in 1799 the Finnish-Hungarian relationship was regarded as finally established.

That this view is widely held by Finno-Ugricists up to the present is shown further by a review of an English translation of Gyarmathi written by

Z. McRobbie (1986: 159) in which Gyarmathi's achievements are summed up in the following terms:

Sámuel Gyarmathi was the first scholar to analyze all of the Finno-Ugric languages on a broad basis by emphasizing systematic lexical and morphological comparisons. He was able to define degrees of linguistic affinity pointing out that Vogul and Ostyak are the closest relatives to the Hungarian language. And although Gyarmathi did not utilize regularities in sound correspondences he was nevertheless able to establish a number of Finno-Ugric etymologies still valid today.

Gyarmathi's work of course preceded the major work of Bopp in 1816 generally viewed as the beginning of comparative linguistics in relation to Indo-European. It preceded the first reconstruction of Proto-Indo-European by Schleicher in the 1850s and of course the first formulations by the Neogrammarians of the notion of exceptionless sound laws in the 1870s. Yet, as we have seen, Gyarmathi's work has been widely viewed by Finno-Ugricists as having proved the relationship of these languages.

At the other extreme, and very recently, Callaghan (1991: 131) in a review of a dictionary by Harvey Pitkin of Wintun, a Penutian language, remarks:

There has been a long debate about whether all or part or *none* (italics mine) of the Penutian hypothesis is valid. A preliminary step to the determination of remote relationships is competent reconstruction of protolanguages at a lowest level.

Presumably, Penutian includes here California Penutian. Except for the extinct Costanoan whose likeness to Miwok is obvious, we have one or more dictionaries and grammars of all the languages Miwok, Yokuts, Wintun, and Maidu. What more evidence do we need to decide the validity of California Penutian at least, which was proposed by Dixon and Kroeber in 1919 more than seventy years before Callaghan's statement? Moreover, Callaghan (1958: 193) herself had stated that a consensus of opinion among the four field workers including herself concerned with these languages was that California Penutian does indeed constitute a linguistic stock. In the very same issue of the *International Journal of American Linguistics* there was an article by Pitkin and Shipley, two of the field workers alluded to in the preceding statement, which contained over 400 etymologies of California Penutian with the usual apparatus of sound correspondence and reconstructed forms.

I have purposely counterpoised here two extremes, the 'prescientific' work of Gyarmathi on Finno-Ugric widely accepted as 'proof' of the validity of that stock, and the evidence apparently accepted by Callaghan in 1958, but rejected by her in 1991 as inadequate since she states that possibly none of the Penutian hypothesis is correct.

From this we see what constitutes proof of relationship varies tremendously depending on the particular language group involved and the historical period. There is then no consensus as to how much and what sorts of evidence is sufficient to 'prove' relationships. What we actually have at present, seen in encyclopedias and general reference works on linguistics, is a product of historical accident. Objectively, the California Penutian languages are probably closer to each other than Finnish is to Hungarian. If a stock was obvious enough and especially if it became established early enough it was then and generally is now an accepted grouping. If it was proposed later and in an atmosphere in which terms like the 'rigorousness' of the comparative method received more and more emphasis it became controversial or was rejected largely, again, for accidental historical reasons, this time changing fashions and the vicissitudes of the *Zeitgeist*.

8.3 Relationship versus classification

The reason for what has been called in the previous section 'relativity of proof' is that the wrong question is being asked. If we ask whether A is similar to B, since everything in the universe is similar to everything else in some way, one never knows how similar something must be to something else (here, of course, one language or language group to another) to draw any conclusions. What is lacking is a standard of comparison, the standard being in this case other languages. But this already brings with it multilateral comparison, degrees of similarity and thus, when carried to its ultimate conclusion, a classification of languages.

Ironically, this point has often been seen, but not carried through with thoroughness or impartiality by opponents of particular hypotheses of classification. An interesting case in point is that of Yukaghir, in regard to the Indo-European-Uralic hypothesis. Yukaghir was up to that time, and still is now lumped by many with other languages in Siberia spoken by small populations in a Paleo-Siberian group with the proviso that none of them are provably related to any other.

Paasonen in 1907, in order to refute the hypothesis that Indo-European was related to Uralic, pointed out there were at least as many resemblances between Uralic and Yukaghir as between Uralic and Indo-European and everybody *knew* that Yukaghir was not related to Uralic. The hidden syllogism of Paasonen's reasoning might be stated as follows. (1) Yukaghir is not related to Uralic. (2) Yukaghir shows at least as many resemblances to Uralic as Indo-European does. (3) Therefore, Indo-European is not related to Uralic. By calling attention to the numerous resemblances of Uralic to

Yukaghir, Paasonen inspired Collinder (1940) and others to advance the notion that Uralic was related to Yukaghir. This, of course, does not logically exclude the idea of a relationship between Uralic and Indo-European, only it is a more distant one. For A and B to be related in this broader framework means that they are more similar in essential ways requiring genetic explanation than some other group or groups of languages. Thus, Indo-European and Uralic are more similar to each other than either is to Sino-Tibetan. At this point, we do not posit a relationship between Indo-European and Sino-Tibetan until a still larger picture emerges.

Once we proceed in this way, we end up with a taxonomy, a classification involving degrees of relationship, one in its abstract structure quite similar to that of a biological taxonomy. To make it complete we look for all the languages belonging with A at a particular degree of relationship. Such an entire group we traditionally call a linguistic stock and it corresponds to the biologist's notion of a taxon.

An essential difference between the approach through relationship as against that through classification is that comparison in the former case might be called decontextualized. We compare only A and B as though no other languages in the world existed. Hence, no standard of comparison exists leading to the difficulties that have just been seen.

8.3.1 *The orthodoxy of the classification approach*

The superiority of the classification approach should be obvious. From classifications we derive numerous statements of relationship of various degrees but it is difficult to see nor has, I believe, anyone shown how we can derive a taxonomy solely from statements of relationship. Moreover, a taxon at any level is a natural subject for comparative study and reconstruction. The protolanguage thus reconstructed represents a real, historically valid entity. It implies a population who spoke it and we can derive important information about human history from hypotheses regarding its place of origin and from the reconstructed forms indicative of the culture of the speakers.

Contrary to some contemporary opinions, this has always been the basis of traditional comparative linguistics. Indo-Europeanists do not take two branches at random, say Slavic and Italic, and make a restricted comparison between them to reconstruct a Slavic-Italic protolanguage. On the other hand, there does exist a field of Balto-Slavic studies because of the fairly large number of features common to Baltic and Slavic compared to other Indo-European languages. This leads to the positing of an intermediate Balto-Slavic node between Proto-Indo-European on the one hand and Proto-Slavic and Proto-Baltic on the other. Buck (1933) in his comparative

grammar of Greek and Latin makes it clear from the outset that there is no linguistic reason that would justify such a comparison, only the cultural importance of Greek and Latin literature and the fact that study of the two of them is commonly found in a single department of Classics.

The approach through multilateral comparison is merely an attempt to make explicit what the assumptions were that led to the first and universally accepted classifications into linguistic stocks. It is an effort to use the terminology of the logical positivists, to explicate the notions of linguistic stock and family tree of languages. Such an explication involves the elimination of decontextualized isolated comparisons as described above and its replacement by systematic multilateral comparison in order to produce a valid taxonomic scheme of languages.

8.3.2 *Regular correspondences as proofs of relationship*

We now consider the two most frequently used criteria, as mentioned in §8.2.2, which are supposed to prove relationship, namely regular sound correspondences and the reconstruction of protolanguages. In fact, these two criteria are ultimately related. Textbooks of comparative linguistics, in their methodological chapters, start with sound correspondences and then show how, by using them, one can reconstruct the ancestral language. Viewed in this light they are part of a larger whole, the initial and final stages respectively of the comparative method. Hence, one often encounters the statement that the validity of some particular stock has been demonstrated by the comparative method. Nevertheless, these two criteria deserve separate treatment.

Regarding sound correspondence, the detailed treatment in Greenberg (1987) [§6-Ed.] can be summarized as follows. The statement that languages are related if they show regular sound correspondences is not so much denied as shown to be refutable in specific cases because there are so many alternative explanations of the sort universally accepted in historical linguistics, the most important probably being analogy. Here individual linguists may differ in regard to preferences for particular types of processes as explanation. There are those who believe that sound laws have no exceptions and who are, therefore, likely to favor regular sound changes to account for the difference in forms which appear similar enough to be likely etymologies. On the other hand, there are those who refuse to multiply sound correspondences and either reject certain etymologies accepted by the first group, or resort more frequently to other explanations including sporadic assimilation and dissimilation, metathesis, and of course morphological analogy more frequently than the first group. These are, of course, tendencies rather than two clearly defined camps with fundamental methodological differences.

There is a further point which has been seen and accepted for some time now by many historical linguists, namely that where there have been extensive borrowings from one language into another, the two languages can show great regularity of sound correspondence, indeed, often greater than between cognate languages. A now famous case is that of Wolfenden who assigned Thai to Sino-Tibetan on the basis of borrowings from Chinese into Thai. His thesis is now universally rejected but it is not always realized that in a broader classificational attempt it would be unacceptable, not only because Thai basic vocabulary is hardly affected but because every Thai resemblance to Sino-Tibetan points directly to Chinese as a source. If it were really Sino-Tibetan it would in its vocabulary frequently display resemblances to [vocabulary in] languages like Tibetan and Burmese but not found in Chinese.

8.3.3 *Reconstructions as proof of relationship*

The notion of reconstruction easily derives from that of sound correspondence. Suppose that among five related languages there is correspondence such that three have p and two have f. If we take the further step of assuming that the original sound was $*p$ we are making an additional assertion, namely that the original p became f in two of the five languages. Suppose that four of the languages had f and only one had p? Should we go with the majority and reconstruct $*f$? Clearly new and basically typological factors arise here. One has to do with the diachronic typology. Is the change $*p > f$ more likely than $*f > p$? Most linguists would agree, I believe, and reconstruct $*p$. Suppose there were another correspondence in which all of the five languages had f. Since the reconstruction of $*f$ here would be overwhelmingly favored, the reconstruction of $*p$ for the first correspondence would be strongly preferred even if only one of the languages had p. This would be for a different and synchronic typological reason. The protolanguage would otherwise have two $*f$ phonemes. We would also require unconditional split, the sin against the Holy Ghost in comparative linguistics.

There are a fair number of instances e.g. Blackfoot in Algonkian, Cherokee in Iroquoian, and Kamchadal in Chukotian, in which reconstructions have not been possible for the languages indicated yet everyone admits that they belong to these respective families. Given this and the possibility of reconstructing a perfectly good protosystem in some instances for loan words, reconstruction becomes neither a necessary nor a sufficient condition for the languages to be related.

Strictly speaking, I believe that reconstruction in cases such as Blackfoot and Cherokee is possible by positing a very large and typologically implausible number of phonemes for the protolanguage.

In one well-known instance, the vowel system of Uralic, actually that of the initial syllable where a full set occurs, there have been two major theories: that of Steinitz and that of Itkonen. The former makes the system like that of Proto-Ostyak as he reconstructs it and the latter quite like that of Finnish. Steinitz posits a whole series of *Ablaut* variations in the protolanguage which saves the regularity by assuming inherited vowel variations, while Itkonen is more tolerant of irregularities but in either case a large number of instances are irregular. At one point, Décsy said that in Finno-Ugric the second part of Voltaire's famous *bon mot* concerning etymology holds, namely that the vowels count for nothing.

Since, as we have seen, Finno-Ugric was discovered even earlier than Indo-European and no one doubts its validity, the question arises as to whether lack of regular sound correspondences could ever disprove a relationship based on numerous and obvious similarities. The answer seems to be that it could not. In principle, there should not be any difference between consonants and vowels in this regard. If there was ever a case for rejecting a hypothesis on the basis of irregular correspondences it would be Finno-Ugric, but it is universally accepted. A theory not disprovable by any data is not an empirical theory.

8.4 The internal logical structure of the comparative method

In the above discussion there has been an implicit assumption apparently shared by all historical linguists that there exists either a set of sound laws or a valid reconstruction that is, as it were, timelessly true. We need to distinguish here two things. One is the real history of investigations by historical linguists as they take place over time. The other is *logical* priority of certain steps in the process in relation to other [steps]. We have already seen something of the notion of logical priority in the discussion of the relation of sound correspondences to reconstruction. In fact, the logically prior is likely to occur first historically. Sound correspondences are noted before protophonemes and reconstructed forms are set up. The two kinds of priority both figure in the following statement of Delbrück (1884: 47), the collaborator with Brugmann in the great comparative grammar of Indo-European:

Since obvious etymologies are the materials from which sound laws are drawn, and this material can always be increased or changed, therefore new sound laws can continually be recognized and old ones transformed.

First, Delbrück is obviously portraying the development of linguistics over time. New sound laws arise and old ones are changed. The initial sentence, however, contains the key notion that sound laws are in the first place drawn

from obvious etymologies. But such obvious etymologies have then a kind of logical priority and must somehow be recognized by properties not involving sound laws since sound laws are derived from them.

Others have made essentially the same point. Goddard says, 'In general, the establishing of phonological correspondences goes on within a family of languages known to be related . . .' (Goddard 1975: 25; quoted with approval by Campbell and Mithun 1979: 52). Somewhat earlier, Newman had similarly stated, 'The proof of genetic relationships does not depend on the demonstration of sound laws. Rather, the discovery of sound laws and the reconstruction of linguistic history normally emerge from careful comparison of languages known to be related' (Newman 1970: 39). Regarding Newman's observation, Watkins, a distinguished Indo-Europeanist, has said more recently (1990: 292), 'As to the mystique of sound laws on the other hand, Greenberg is quite right to quote with approbation the Africanist Paul Newman (1970).'

None of these citations faces the problem of the initial step, namely how we recognize 'a family of languages known to be related' (Goddard) or 'languages already presumed to be related' (Newman). What I believe does emerge here is the recognition of at least three stages, which are in order of logical and usually historical priority the following: (1) recognition of a family of related languages; (2) the discovery of sound laws; (3) the application of the comparative method, starting with sound laws, leading to the reconstruction of linguistic history and the protolanguage. In this process, sound laws are often revised or abandoned and new ones found. A core of basic etymologies furnishes the starting point (cf. Delbrück above) and it is rare for any of these to be abandoned. New ones are found, and some which are proposed become a matter of debate with some linguists accepting and some rejecting. Etymology will never cease to exist as a field of study in any linguistic stock simply because more etymological problems will always exist.

8.4.1 *The protocomparative method*

By the term 'protocomparative method', I do not mean the total method of constructing a protolanguage, but rather an initial stage of the comparative method which logically and in actual practice but without explicit recognition, precedes what is usually taken to be the initial stage in the comparative method, namely, the discovery of sound correspondences.

The pre-sound-correspondence stage was illustrated briefly in Greenberg 1987 (especially 24–6) [§6-Ed.]. There in Table 7 equivalents of nine common words are given for twenty-five languages in Europe. By the time one has gotten to the second word a threefold division into Indo-European, Finno-Ugric,

and Basque and even the major subgroupings of the first two (the third being an isolated language) become apparent. In comparing for example the third item, equivalents for the English word 'three', we are making many judgements of differential phonetic similarities, the meaning here being kept constant. For example, we judge that Welsh *tri* is more similar to Italian *tre* than either is to Finnish *kolme* and in that, in turn, Finnish *kolme* is more similar to Estonian *kolm* than either is to the Welsh and Italian forms. Even these four equivalents for 'three' give us a grouping of Welsh and Italian (Indo-European) as against Finnish and Estonian (Finno-Ugric). But how do we arrive at these judgements? In doing so, we have as it were applied the comparative method in embryonic form. We compare the *t* of Welsh *tri* with the *t* of Italian *tre* and not with *e* in the latter word. So also the two *r*'s and the vowels are being compared. This set of equivalents for 'three' and the groupings it gives are already important evidence leading to the correct classification and, of course, there is no recurrence of sounds within the words so there is no correspondence of *t* to *t* or *r* to *r* in the usual sense. When, however, we have found more words such correspondences will be found and the *t* and *r* in the words for 'three' will be examples of them. We might, therefore, call them protocorrespondences.

There is further in the same section in Greenberg (1987) a discussion of the equivalents for the Proto-Indo-European word for 'tooth' as leading back quite naturally to a single form that might be characterized as approximately *dant* or *dent*. This once again is based on exiguous data, without yet assuming anything like a complete original sound system or a regular set of sound changes. Such relatively amorphous hypotheses are the actual and the logical precursors of more fully elaborated but never conclusive formulations resulting from the application of the comparative method. They are the true first steps in the method, never recognized in textbooks of comparative linguistics but worthy of fuller study and consideration.

8.4.2 *Genetically relevant criteria*

In all the preceding discussion there has been frequent reference to similarities and degrees of similarity among languages e.g. that English is more similar to German in genetically relevant ways than it is to French. However, there has been no explicit treatment of the question as to what kinds of similarities are involved. After all, languages can be like or unlike each other in innumerable ways. However, there has been an implicit answer to this question in the immediately preceding section, similarity in specific word forms, or more exactly morphemes, involving sound and meaning simultaneously. Resemblances of other kinds, e.g. in sound only as in the existence of

glottalic ejectives in both languages, or meaning only as in agreement in a tense system including a future versus nonfuture opposition, or in word or morpheme order are irrelevant and would normally be considered typological. Of course, languages can have typological agreement with other languages of the same stock which is part of their common inheritance from their common linguistic ancestor. However, such characteristics do not figure in our carrying out the classification. They are, as it were, a bonus deriving from it and useful in considering the relative stability of typological traits and their historical development.

However, as self-evident as it might seem that it is from similarities in sound and meaning simultaneously that one finds the materials for genetic classification, there is a strand of thinking in keeping with the strong tendency towards formalism in linguistics to dismiss phonetic similarity. All that would count would be regularity of correspondence in abstraction from phonetic content. Of course, semantic similarity is not suppressed. For some, it is narrowed to semantic identity which in practice can only mean translation equivalence. The reason for this is that they wish to make it as difficult as possible to prove that anything is related to anything else.

I earlier thought that the importance of phonetic similarity was so obvious that it required no defense. We have seen in the previous section how differential similarities as in Welsh *tri*, Italian *tre* vs. Finnish *kolme* 'three' play a key conceptual role at the very initiation of comparison leading to classification. Those who oppose this obvious and sensible approach emphasize the frequency of accidental resemblances and the existence of phonetically unusual correspondences. Let us consider examples of each in turn.

Meillet (1958: 92) gives the following example. French *feu* has nothing in common with German *Feuer* 'fire', while the German form goes back to an original represented by words which are quite different, Greek *pur* and Armenian *hur*. What Meillet fails to consider is how we know that French *feu* is only accidentally similar to German *Feuer*.

We know it because if we compare French *feu* with the Spanish form *fuego* and Italian *fuoco*, we see that it goes back to a form in which the second consonant is a velar. Compare also French *peu* 'few' with Italian *poco*. Why do we compare French with Spanish and Italian in this and other instances? A glance at Table 7 in Greenberg (1987: 24) [p. 95-Ed.] will show why. From the very beginning, in word after word French aligns with Spanish, Italian and other Romance languages, while English lines up with German, Dutch, etc. Besides this, in Meillet's example, German *Feuer*, English *fire*, etc. are not all that different from *pur* and *hur* since $p > f > h$ is a very commonly encountered form of change and they agree in their final *r*.

A favorite in recent discussion has been the correspondences between Greek *dw* and Armenian *rk* as seen in the word for 'two' and a few other words. All this shows is that phonetically improbable resemblances may occur in correspondences. However, they are rare. No similar instance to *dw* ∼ *erk* has ever been found but *t* ∼ *d* or *t* ∼ *t* are commonplace.

More importantly, it has nothing to do with classifying Armenian as Indo-European. Scores of phonetically obvious resemblances between Armenian and other Indo-European languages exist and were noted earlier. Indeed, Hübschmann, the founder of modern Armenian studies who first established the affiliation of Armenian on a firm basis does not have this correspondence in his seminal paper on the position of Armenian (1875), what Meillet was later to call the *bel article* on this subject, and he never accepted this correspondence.

8.5 The biological analogy

Beginning in the nineteenth century with Lyell and Darwin on the biological side and Schleicher on the linguistic, the basically similar structure of biological and linguistic evolution has been noted. In both instances, differential similarities result in hierarchial groups which are explained by historical development from no longer existing ancestral forms, whose distance from the present mirrors the degrees of difference. Primates had a common ancestor later than mammals of which they are a branch just as Germanic languages have an ancestor, Proto-Germanic, which is later than Proto-Indo-European of which it is a changed form.

Yet, in biology as I noted (1987: 34) nothing is equatable with sound correspondence. The same observation was made by Dyen in the same year (1987: 708) when he stated that in biology there is no clear analogy to the law of regular phonetic change. However, he seems to view this as just one point, along with numerous other similarities and a few differences between linguistic and biological evolution without drawing any conclusions from it.

I believe that in the context of the present paper, since biological classification is not only possibly but is probably, if anything, more advanced than linguistic classification, what we see is that in the very initial steps of the protocomparative method, before regular correspondences enter in, we already have a valid genealogical classification.

The existence of regular sound changes enables linguists to reconstruct but reconstruction is not necessary in order to ensure the validity of the classification. Once more, biologists do not reconstruct and it is not a part of their taxonomic theory nor does this imply any inferiority in biological taxonomic

endeavors. I am confident that an application of the methods outlined here and elsewhere will ultimately lead to a taxonomy of languages not inferior to that of the biologists in their field.

References

Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache.* Frankfurt: Andreäische Buchhandlung.

Buck, Charles Darling. 1933. *Comparative Grammar of Greek and Latin.* Chicago: University of Chicago Press.

Callaghan, Catherine A. 1958. 'California Penutian: History and bibliography', *International Journal of American Linguistics* 24: 189–94.

——1991. 'Review of *Wintun Dictionary* (University of California Publications in Linguistics 95) by Harvey Pitkin', *International Journal of American Linguistics* 57: 129–31.

Campbell, Lyle and Mithun, Marianne. 1979. 'Introduction', in Lyle Campbell and Marianne Mithun (eds.), *The Languages of Native America.* Austin: University of Texas Press, 3–69.

Collinder, Björn. 1940. *Jukagirisch und Uralisch.* (Uppsala Universitets Årsskrift no. 8.) Uppsala: Almquist and Wiksell.

Delbrück, Berthold. 1884. *Einleitung in das Sprachstudium* (2 Aufl.). Leipzig: Breitkopf & Härtel.

Donner, Kai. 1901. 'Die uralaltaischen Sprachen', *Finnish-ugrische Forschungen* 1: 128–49.

Dyen, Isidore. 1987. 'Genetic classification in linguistics and biology', in George Cardona and Norman Zide (eds.), *Festschrift für Henry Hoenigswald.* Tübingen: Gunter Narr, 101–8.

Goddard, Ives. 1975. 'Algonkian, Wiyot and Yurok: Proving a distant relationship', in M. Dale Kinkade, Kenneth L. Hale, and Oswald Werner (eds.), *Linguistics and Anthropology.* Lisse: Peter de Ridder, 249–62.

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford: Stanford University Press.

Hübschmann, Heinrich. 1875. 'Über die Stellung des armenischen im Kreise der indogermanischen Sprachen', *Zeitschrift für vergleichende Sprachforschung* 23: 5–42.

McRobbie, Z. 1986. 'Review of English translation of Sámuel Gyarmathi's *Affinitas* by Hanzell', *Ural-Altaische Jahrbücher* 58: 159–60.

Meillet, Antoine. 1958. *Linguistique historique et linguistique générale.* Paris: Champion.

Newman, Paul. 1970. 'Historical sound laws in Hausa and Dera (Kanakura)', *Journal of West African Languages* 7: 39–53.

Paasonen, Heikki. 1907. 'Zur Frage von der urverwandschaft der finnish-ugrisch und indoeuropäischen Sprachen', *Finnish-ugrische Forschungen* 7: 13–30.

Pitkin, Harvey and Shipley, William. 1956. 'A comparative study of California Penutian', *International Journal of American Linguistics* 24: 174–88.

Ravila, Paavo. 1935. 'Die Stellung des lappischen innerhalb der finnish-ugrischen Sprachfamilie', *Finnish-ugrische Forschungen* 23: 20–65.

Watkins, Calvert. 1990. 'Etymologies, equations, and comparanda', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin: Mouton de Gruyter, 289–304.

This page intentionally left blank

9

Review of Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction, by Vladimir E. Orël and Olga V. Stolbova*

An etymological dictionary of the Hamito-Semitic languages has long been a desideratum. The only one available has been the long-outdated work of Marcel Cohen (1947). Since that time, the most momentous developments have been in west and east Africa. The addition of a major new branch, Chadic, consisting of over 100 languages in west Africa, has helped to cast new light on both Semitic problems and those of Hamito-Semitic (or as it is now more commonly called, 'Afroasiatic', or 'Afrasian'). The study of this branch has been zealously pursued by a number of specialists who have provided substantial new data and important new theoretical knowledge. One of the two authors of the present work has been active in Chadic studies. In addition, the Cushitic languages, a group itself of considerable time depth, if in fact it is a unitary group, have been subject likewise to intensive comparative and descriptive studies since the time of Cohen.

Proposals concerning the ancestral sound system of the overall family began with the pioneer work of Illich-Svytich (1967, 1971) in the context of a larger Nostratic entity, and other Nostraticists, e.g. Diakonoff and Bomhard, have continued this work. In such a broader context, of course, only etymologies shared by Proto-Afroasiatic (PAA) with other families have been considered. What has been lacking was a reasonably up-to-date etymological dictionary of Afroasiatic (Hamito-Semitic) as such. The present volume is handsomely printed, well-organized, and orthographically pleasing, and thus fulfills a real need in comparative linguistics. There has appeared in the same

year, 1995, another comparative dictionary, by Ehret (1995), an Africanist scholar at UCLA, a work with a markedly different approach and, at times, widely different etymological results. It is not my purpose here to review this other work, but it will be referred to occasionally in relation to some theoretical problems that are relevant to the work being reviewed.

The present work (hereinafter OS) comes from a different tradition. At least one of the authors, Stolbova, was part of a working team under the leadership of Diakonoff, which has been involved for a considerable length of time in compiling an Afrasian (Hamito-Semitic) comparative dictionary. An English version has been appearing since 1933 in a journal that is virtually unobtainable in the United States. I am indebted to Mlle Stolbova for making available to me the sections that have appeared for roots reconstructed with the initial non-nasal labials *p, ṗ, f*, and *b*. The etymologies of these forms differ in many instances from those in the Orël-Stolbova volume and even when they can be clearly equated from the fact that they have entries that are mostly or sometimes entirely identical, the reconstructed head form is rarely the same. The differences between these two sets of etymologies is relatively minor, however, compared to the etymologies in Ehret's dictionary.

There have been several frank statements by Diakonoff regarding the differences between OS and the Diakonoff group. Thus in 1993, Diakonoff notes how, during the apparently unavoidable delays in publishing the results, some of the original team 'have changed their views and are working on Afrasian etymologies independently' (1993: 5; citing here Militarev, Orël, and Stolbova). More recently, in a review of an unpublished work of Bomhard, Diakonoff notes that he does not believe in 'Orël's asymmetrical (!) original six vowel system' (1995: 219). The OS reconstructions indeed involve a six vowel system, *a, e, i, o, u, ü*, while Diakonoff adheres to a two vowel system, *a* and *ə*, in which *i* and *e* arise from *ə*; in the vicinity of palatal consonants, and *u* and *o* in the vicinity of velar consonants. This is correlated with another difference, namely, the positing of a labiovelar series, k^w, g^w, etc., which changes adjacent *ə* to *u* or *o*, a series absent in the reconstructed sound system of OS.

Now such differences, and others not stated by Diakonoff, are deducible from a comparison of the two sets of reconstructions. These are of great interest in themselves to comparative linguistics, especially in view of the emphasis laid by Russian linguists on reconstruction as a proof of genetic connectedness and the unwillingness of Nostraticists to admit languages as Nostratic until a protolanguage has first been reconstructed (e.g. for Ainu and Yukaghir)—even where differences are not much deeper than one might expect at the dialect level.

A fairly close examination of the etymologies in Diakonoff *et al.*, beginning with non-nasal labials and those in OS beginning with the same set of sounds, reveals a number of ways in which differing methods or assumptions can lead to differences in etymologies in the same linguistic stock, even where the total membership and basic divisions of the stock are not in question.

In considering rather briefly the sources of differences, the treatment will not be exhaustive and the order will be to some extent *ad hoc*. The purpose will not be so much to endorse one or the other treatment as to raise and briefly discuss the problems involved. The whole situation is, of course, not one contemplated in the usual presentation of the comparative method in textbooks of historical linguistics, where the assumption is made that there is just one version of the comparative method and that when correctly applied, it leads inevitably to the one correct result—an assumption belied by numerous facts of the kind involved in the present instance. Nor can this situation be attributed to differential competence. I believe no linguist who is acquainted with this family of languages will question the credentials of Diakonoff, Orël, or Stolbova as specialists in historical linguistics and in research on Afroasiatic languages.

One source of difference in the etymologies has to do with the protosound system. There are at least three major differences here between Diakonoff and OS. One was cited earlier with regard to the ancestral vowel system of PAA, for which OS posit six vowels, *a, e, i, o, u, ü*, and which is criticized, as we have seen, by Diakonoff as asymmetrical, while he assumes only two vowels, *a* and *ə*. Here OS in their introduction realize that there may be different viewpoints among specialists and state that 'our potential opponents are welcome to replace vowel signs with generalized V or * symbols [i.e. *a* and *ə*], thus arriving at a more usual variant of the Hamito-Semitic reconstruction' (p. xxi). Ehret also, it may be noted, has posited a vowel system, *a, e, i, o, u*, with original length distinctions for each of these. This is obviously closer to OS than to Diakonoff. How can such questions be decided? Diakonoff's exclamation point after 'asymmetrical' points to an obvious typological factor. It is particularly true for a reconstructed system, for which we have no direct attestation, that this factor becomes important. In fact, systems basically identical to that posited in OS do exist, for example, the Wu dialect of Chinese, which has *a, e, i, o, u*, and *ü* for its non-nasalized vowels (Chao 1970). Attested systems tend to show considerable symmetry, but also certain common (as well as certain rare) deviations.

A basically similar factor is involved in another point of dispute. In the labial series of OS, there is no reconstructed 'emphatic' *ṗ*. It is explained as a West Chadic innovation. Here Diakonoff, after some hesitation in earlier

writings, reconstructs *p as PAA, as has generally been the case with others. Now the system of obstruents for Proto-Semitic largely shows an organization into triads: unvoiced, voiced, and emphatic. It is a major contribution of Afroasiatic studies to have shown that this system goes back to PAA, and that the emphatics were, even in Proto-Semitic, glottalized ejective consonants, such as those found to this day in the Semitic languages of Ethiopia. For the labial set, however, there is clearly no glottalized member in Proto-Semitic, and the p of PAA, if it existed, was relatively rare. For example, in Diakonoff's etymologies with initial labials, as against forty-eight with p, there are only nine with $ṗ$. Some of these are found in OS under initial p, some under b, and some not at all.

There is here, once more, an empirical question largely distinct from what one might consider an 'ideological' question. Just as with the vowels, some will tend to opt for symmetrical systems and others will tolerate a degree of asymmetry if required by what they view as the facts. This is particularly true of rare phonemes. If etymologies are few, some or all may be rejected on various grounds, such as semantic plausibility, breadth of distribution within the language family, etc., or reinterpreted as involving other phonemes, sometimes by positing irregularities of phonetic development. In fact, in the real world, such irregularities do occur, as will be evident for Proto-Indo-European on just about any page of Pokorny's comparative dictionary (1959). Once again, OS endorse an asymmetrical solution while Diakonoff and others prefer symmetry at the cost of positing rare phonemes. Once more, also, there is typological evidence. In Greenberg (1970), a typological study of glottalic consonants, it is noted that in a language that has glottalic consonants (whether implosives, ejectives, or both), a gap at the bilabial point of articulation is found in a number of world areas. This does not, of course, prove that OS are right in this matter, but it does show that if a gap is to be found, it will tend to be at the bilabial point of articulation and certainly strengthens the case that the system they are positing is typologically plausible.

Another difference, this time outside of the area of phonology, is found in regard to derivation and the whole question of the nature of the root in Afroasiatic. It is now generally agreed—and here once more the broader Afroasiatic context provides important support—that the famous Semitic triliteralism is, in the main, historically secondary. In particular, original bilaterals, which survive frequently in common nouns, were extended by reduplication of the second consonant, by medial *w/y* (the so called 'hollow roots') or final *w/y* (the 'weak roots'), but above all by the suffixation (much less frequently the prefixation) of a consonant to which we cannot usually attribute a specific meaning. The situation is not unlike that found in

Indo-European with regard to 'root determinatives'. Although there can hardly be doubt regarding the phenomenon itself in Afroasiatic, clearly its extensive use carries with it some dangers. This is admitted by Diakonoff, when he notes that 'in the case of phonetically related groups of lexemes, we have sometimes united under the same reconstructed root groups of words connected but loosely in their semantics' (1993: 7). In their dictionary, OS proceed quite differently, and they articulate this difference by noting that 'the dictionary is organized on the *lexical* and not the radical principle' (p. xv). For Indo-Europeanists, it may be noted that Pokorny's comparative dictionary (1959) is radical, but Mann's (1984–7) is lexical. In a lexical dictionary, the usual method in compiling dictionaries of English, for example, we have separate entries for 'strong' and 'strength', although both have the same radical. The principle underlying the traditional dictionary of Arabic, however, and still largely maintained in bilingual dictionaries by Western scholars, is radical. The radical dictionary is not without compensating advantages, especially in showing ultimate etymological connections. Moreover, in practice, compromises between the two principles are possible. My present purpose is to call attention to the existence of two different types of approaches regarding what is an entry in both synchronic and comparative dictionaries and to a theoretical problem in lexicography.

Some further major sources of differences among etymological dictionaries compiled for the same linguistic stock may be briefly indicated here. One is what might be called the 'representation' problem. What must be the breadth and the nature of the distribution of an etymon within the overall family to justify its inclusion in an etymological dictionary? Here the assumed subgrouping within the family becomes a decisive factor. For example, within Afroasiatic, some scholars assert that traditional West Cushitic (Omotic) is so divergent that it forms a sixth branch. Others go even further in dividing Afroasiatic at the deepest level into Omotic versus non-Omotic. Does this mean that for an etymological item to be reconstructed as PAA it must occur in both Omotic and non-Omotic? As can be seen, where the division is into two branches, the problem becomes particularly acute. Surely, even if one accepts the special genetic position of Omotic, since it must have lost vocabulary items during the period from PAA to Proto-Omotic—items that could well have been retained in non-Omotic—many valid etymologies will be omitted in our set of reconstructions when they happen not to occur in Omotic; the trouble is that we do not know which ones. As the number of presumably comparable subgroups in a family increases, the larger the proportion of true protovocabulary that becomes discoverable by the comparative method. But, it will never be 100 per cent,

even though outside, more remotely related languages may reveal at least some of these items.

In regard to subgrouping, Ehret differs from both Diakonoff and OS in his subgrouping of Omotic, in that in his view, Omotic has a special position, and hence a number of the most familiar and widely accepted Afroasiatic etymologies are not contained in his volume because they do not occur in Omotic.

Particularly when a language has no obvious and close relatives, the restriction of external comparison to only items reconstructible by some representational criterion may cause a comparativist to exclude crucial items on the ground that they are not 'reconstructible' for the protolanguage. It may be noted that in Indo-European, which has been the model for historical comparative linguistics, such restrictions are not adhered to in practice. An example in point is Celtic, which falls clearly into just two branches, Gaelic and Brythonic. Strictly speaking, a form is only provably Proto-Celtic if it occurs in both of these branches. However, generally accepted Indo-European etymologies abound in which Celtic is represented by Gaelic, or by Brythonic, but not by both.

The chief reason, however, that etymological dictionaries of the same language family are never identical is semantic. For comparative purposes, semantics cannot be reduced to formal rules analogous to those of phonology in the absence of a parallel to sound laws in the semantic aspect of language. However, the problem cannot be avoided since every form that we cite in an etymology has a meaning. Some have sought safety in the requirement that compared forms have identical meanings. What we actually have is, of course, translation equivalence which is never perfect. In actual practice, of course, no one proceeds in this manner. All linguists have some knowledge of what semantic change is like and know that certain semantic equations are universally acceptable, e.g. 'moon' = 'month'. General acceptance is based on inherent plausibility, attestation from changes in languages with a written tradition, and the existence of languages in which both concepts are expressed by the same form. An example is the equation of 'wood' and 'tree', e.g. German *Baum* and English *beam*. Here the validity of the etymology is further strengthened by phonological evidence, e.g. German *Saum* = English *seam*, German *Traum* = English *dream*, etc. The plausibility of the 'tree' = 'wood' equation is further strengthened by the existence of numerous languages in which the same term is used for both concepts. Beyond such examples is a vast penumbra, ranging from the fairly plausible to what we may think of as at least possible. Here individual judgements may differ widely; what appears possible to one investigator will seem far-fetched to another.

It might be thought that all examples might be simply decided by regularity of sound correspondence, but it is not difficult to see that this can only be a factor, although often an important one. Thus the equation of Latin *sequor* 'I follow' with Gothican *saihwan*, (German *sehen*, English *see*) is universally accepted by Indo-Europeanists because of the regularity of the sound correspondence—although it would be difficult to find any example of this semantic equivalence anywhere else. On the other hand, there is surely a semantic factor also involved: if one were to show regular phonetic correspondence between the verb 'to follow' in one language and the noun for 'plum' in another, this would clearly not be an acceptable etymology. Moreover, the sound correspondences themselves do not arise in a semantics-free environment. They are first noted in forms with both semantic and phonetic plausibility; then the whole structure is built up from such beginnings.

A semantic factor further arises in instances that deviate from the sound laws but where the combined force of phonetic and semantic similarity makes the etymological identity universally acceptable. For example, the connection of Latin *nomen*, Russian *imja*, and English *name* is universally accepted although the usual vowel correspondences are violated and we must assume the dissimilatory loss of the initial nasal consonant in Slavic. Another example involves Greek *nuks*, Latin *nox*, English *night*. We would expect Greek *o*, not *u*. Brugmann sought to explain the Greek vowel as the result of the labializing influence of the Indo-European labiovelar k^w. However, this is not a regular change. Other attempts have been made, but no one rejects the etymology.

It should be clear from what is said here that the OS volume is a major contribution to Afroasiatic studies. One hopes that eventually the work of Diakonoff and his group will also be published in a single, comprehensive, and generally accessible form as a book or monograph.

In the meantime, OS will remain as a major contribution to comparative studies. A minor objection can be raised concerning the name 'Hamito-Semitic', with its very misleading linguistic and extralinguistic connotations. By now, both specialists and nonspecialists are well acquainted with 'Afroasiatic' or 'Afrasian', and one or the other should surely have been adopted in the present work.

References

Chao, Yuen Ren. 1970. 'The Changchow dialect', *Journal of the American Oriental Society* 90: 45–56.

Cohen, Marcel. 1947. *Essai comparatif sur le vocabulaire et la phonétique du chamito-semitique.* Paris: Honoré Champion.

Diakonoff, Igor M. 1993. 'Historical and comparative vocabulary of Afrasian', *St Petersburg Journal of African Studies* 2: 5–27.

——1995. 'Review of *Indo-European and the Nostratic Hypothesis*, by Allen S. Bomhard', *Mother Tongue* 1: 219–22.

Ehret, Christopher. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian).* (University of California Publications in Linguistics, 126.) Berkeley and Los Angeles: University of California Press.

Greenberg, Joseph H. 1970. 'Some generalizations concerning glottalic consonants, especially implosives', *International Journal of American Linguistics* 36: 123–45.

Illich-Svytich, Vladislav M. 1967. 'Materialy k sravnitel'nomu slovarju nostraticheskix jazykov', *Ètimologija* (Moscow) 1965: 321–96.

——1971. *Opyt sravnenija nostraticheskix jazykov*, vol. 1, ed. V. A. Dybo. Moscow: Nauka.

Mann, Stuart E. 1984–7. *An Indo-European Comparative Dictionary.* Hamburg: Helmut Buske.

Pokorney, Julius. 1959. *Indogermanisches Etymologisches Wörterbuch*, vol. 1. Bern: Francke.

10

Protolinguistic variation: a link between historical linguistics and sociolinguistics*

This study resumes and expands some of the topics discussed in Greenberg (1978) on the stages of the definite article and Greenberg (1981) which is devoted to two additional examples of Stage III articles, Nilo-Saharan *k*- and Penutian -*s*. In the present context, I will be especially concerned with an aspect of the Stage III article which only received incidental treatment in the earlier studies, namely, the fact that the distribution of the forms with and without the article shows a certain kind of randomness in relation to the languages and dialects of the linguistic stock in which it is found. I will also seek to show that this particular kind of randomness of distribution, to be defined and illustrated later, also exists in regard to other grammaticalized elements than definite articles in their later stages of development, and indeed is not confined to grammaticalized markers, but also extends to variants of morphophonemic origin and perhaps even to those arising from purely phonetic changes. Finally I will suggest that the sociolinguistic and linguistic situations which give rise to such variation are distinct from both the areal and the genetic types to which we are accustomed and thus constitute a third type of phenomenon with both historical and sociolinguistic implications.

In illustrating what is meant here by random variation, I will take as an example the Nilo-Saharan prefix *k*- which in my first study of languages in that family was called moveable *k*- (Greenberg 1966). The reasons for considering this the continuation of a Stage I article similar to the English definite article followed by a Stage II article (roughly speaking combining the characteristics of a definite and indefinite article) are set forth in Greenberg (1978, 1981) and are not repeated here. From the point of view of distribution over languages in a particular linguistic stock, what is here called random

variation has some, or usually all, of the following four characteristics as illustrated by Nilo-Saharan *k-*:

1. Random distribution of forms with and without *k-* over separate branches of Nilo-Saharan, e.g. Maba *eri*, Songhai *kuri*, Daza *gere* 'blood'. Here the languages cited belong to separate branches of Nilo-Saharan at the deepest level of subclassification. Likewise within the same branch of Nilo-Saharan similar phenomena are found. In Table 1 below some etymologies are to be found which are represented both in Nile Nubian and Bari which belong to separate subgroups of the Eastern Sudanic branch of Nilo-Saharan.

TABLE 1

Nile Nubian		Bari	
guar	'ant'	ki-gwur-te	'ant'
ur	'head'	ur-et	'crown'
aru	'rain'	k-are	'river'
gu-mur	'neck'	mur-ut	'neck'

It will be noted that all four possibilities are represented: languages with reflexes of *⋆k-* ('ant'), neither language ('head/crown'), and presence in either one without the other (e.g. 'rain/river' with *k-* in Bari but not in Nubian and 'neck' with *k-* in Nubian but not in Bari). This represents the usual situation and numerous additional examples could have been given.

2. Such random variation can also be found even among dialects of the same language. In Table 2 examples are adduced from Moru, a language of the Moru-Madi subbranch of the Central Saharan branch of Nilo-Saharan. Regarding the three dialects cited, Tucker (1940: 3) states 'The three dialects Miza, Kediru and Moroändri are so similar as to be almost identical.' Yet even at this low genetic level similar variations occur.

TABLE 2

Miza	Kedir	Moroändri	
k-umu	k-umu	umu	'fly'
k-ini	ini	ini	'skin'
k-ari	k-ari	k-ari	'leprosy'

3. There are even instances of free variation within the same dialect. For instance in the Keliko dialect of Moru *kari* and *ari* are given as variants of the same word for 'blood'. Of course, as with numerous statements regarding free variation, one does not know in the absence of sociolinguistic studies the

nature of this variation, whether between individuals, social strata, kin-groups, or geographically determined entities. Nor does one know whether the variants are differently distributed in accordance with such criteria in lexically distinct items.

4. Finally it should be noted that, usually at a very low genetic level, particular dialects or languages show a distinct or even exclusive preference for one variant over another. Thus in Table 2, Miza has *k-* forms in all three instances. An examination of the relatively brief comparative vocabulary in Tucker (1940) shows only rare instances in which Miza shows forms without *k-*.

Although the instances just cited show a kind of functionless choice of variants among related languages and dialects, this need not always be so. Four types of acquisition of new functions can be distinguished, once more illustrated from Nilo-Saharan *k-*. One of these, however, the second one listed, could not be found in this instance but will be illustrated later.

1. There are sporadic cases in which what would otherwise be homonymous forms are differentiated by the choice of one of the variants as against the other.

An example from dialects of Moru, the Central Sudanic language which is the source of the examples in the foregoing table, will illustrate this possibility. In both Miza and Kediru, the root *ari* without *k-* means 'bird', while *k-ari*, otherwise identical means 'blood'. It is interesting that in the third dialect of Table 2, Moroändri, the two are differentiated in a quite different fashion. In this dialect *ari* is the word for 'blood' while *ari-va* means 'bird'. The suffix *-va* in this latter form is a diminutive.

2. Sometimes both forms are retained but undergo a secondary semantic differentiation. As noted earlier, I have been unable to find an example of this involving Nilo-Saharan *k-*. An example in a quite different context is English *shade* and *shadow*, the first deriving from the nominative singular *scadu* and the latter from the oblique stem *scadw-* after the dissolution of the Old English case system.

3. There may be sporadic new grammatical functions. Thus in Kanuri, a language of the Central Saharan branch of Nilo-Saharan, in just one root, the form with *k-*prefix, *k-am*, is used as a singulative 'person', while the form without *k-*, *am,* is a collective meaning 'people'. The most closely related language to Kanuri, Daza, uses *am* as the general root for 'person' and has no form with *k-*. In Karimojong, a language of the Eastern Nilotic subbranch of Eastern Sudanic, examples of this kind are more numerous but the formation is still not productive, e.g. *(e)-ki-twani* 'a single scorpion'; *(ŋi)-twani*

'scorpions'. In these forms, the prefixes *(e)-* 'singular' and *(ŋi)-* 'plural' are new second-stage articles.

4. However, productive new grammatical functions are sometimes found. Since the *k-* prefix as a former article deriving from a demonstrative occurs only on nouns, it is reinterpreted as simply a mark of nominality and as such becomes a productive derivational element forming verbal nouns. An instance is Sara, a language of the Central Sudanic branch of Nilo-Saharan. As an example we may cite the verb root *usa* 'to eat' from which a verbal noun *k-usa* 'act of eating' is formed.

We see from the above example of reinterpretation, that one possibility is that the formative spreads and acquires new functions. Another possibility is the opposite, which we may call contractive. In such instances the element does not spread and acquire new functions (expansion with resemanticization) but survives in just a few examples in which it has been lexicalized, in the sense, that, from the synchronic point of view, it has been incorporated as an indistinguishable part of a lexeme.

However even in a purely synchronic description, it may be noted that there is, statistically, more than a chance correlation between certain grammatical or semantic classes, and certain sequences of phonemes which thus form a submorphemic entity.

These possibilities, namely of expansive resemanticization and contractive desemanticization can be illustrated from the Chibchan-Paezan suffix *-kwa*. The original meaning seems to have been 'egg, nut, or other similar round object'. It survives in a few instances as a lexical item with this meaning e.g. Terraba (Costa Rica) *gwa* 'egg', and Cuna (Panama) *kwa-(kwa)* 'nut'. In this latter case we find *-kwa* also as a suffix on round objects but also extended to a considerable part of the nominal vocabulary so that it is difficult to assign it a single, definite meaning synchronically.

As an example of expansion with resemanticization we may cite the instance of Millcayac, a Chibchan language of Argentina in which *-gue* is a productive derivational suffix forming verbal nouns, e.g. *cheri* 'to give'; *cheri-gwe* 'gift'. Thus it has ended in this case with a derivational function quite comparable to that of Nilo-Saharan *k-* in spite of its very different origin.

The second type of development that *-kwa* undergoes in some Chibchan languages is incorporation into a system of numeral classifiers and indicating small round objects. As noted in Greenberg (1972), the most common source of a general classifier is that for round objects. In some cases, after acquiring this function, the very fact that it can appear with all nouns, even if only in quantifying phrases, makes it once more a possible general marker for nouns.

However, given the syntactically limited use of numeral classifiers, it is the development which is less likely to occur, particularly if the original order is numeral–numeral classifier. Under these circumstances it may survive in isolated cases as a mark on a few numerals and can be considered from the synchronic point of view simply as part of the numeral. Still its presence in several numerals, and even more powerfully, the comparative evidence, will lead to the correct historical interpretation.

In the case of Chibchan *-kwa*, this occurs in some languages. For example in Kagaba, a Chibchan language of Colombia, it is found in just four numerals, *mai-gua* 'three'; *ku-gua* 'seven'; *abi-gua* 'eight'; *aita-gwa* 'nine'. With these we may compare Margua, another Chibchan language of Colombia, which has *mai* 'three' and *avi* 'eight'.

In both cases discussed above, Nilo-Saharan *k-* and Chibchan *-kwa*, we are dealing with grammaticalized elements in which the variation is between presence or absence of the item in question. It could have been further illustrated from instances like petrified honorifics or diminutives.

However, the variational phenomena earlier enumerated are not confined to grammaticalized morphemes which alternate between presence and absence. Similar distributional properties across language and dialects, and in some instances, similar examples of semanticization are to be found in the case of morphophonemic alternants. Since in such cases, the item in question did not previously have a meaning, we may talk about semanticization rather than resemanticization. Moreover, the variants, typically two in number, both have overt phonological expression in contrast to the examples treated above.

An important source of such variation is the breakdown of vowel harmony systems. There are two main types of development. One is through merger of pairs of alternants resulting typically in so-called neutral vowels as in the instance of Mongolian *i* which functions both as back and front vowel. The usual assumption is that this is the result of a change by which its former back partner *y* becomes *i*, thus eliminating the alternation. It is easy to see how further mergers may finally destroy a vowel harmony system completely and this has indeed happened in Kerek and Aliutor, both dialects or perhaps separate languages closely related to Koryak, a Chukotian language. In Koryak itself as well as in Chukchee and the more distantly related Kamchadal, a system of high-low harmony still functions. Another course of events, however, gives rise to the kind of variability that we have already seen in the case of grammaticalized elements. An example is that of the East Mongolian languages, all of which are spoken in China. It is clear that Proto-Mongol in regard to vowel harmony was essentially like Classical Mongolian

and the present day Western Mongolian languages such as Khalka, Buriat, and Kalmuck. The back-front harmony of these languages is stem-driven in that the vowel of the stem remains constant, and the derivational and inflectional affixes, which all follow the stem, vary in regard to backness or frontness depending on the stem.

In the Eastern Mongolian languages, the stem vowels remain basically in their inherited form, but the vowels of affixes tend to have a single variant, each one being an independent case. Thus in a particular language, the front variant can be found in one affix, but the back in another, or they may be in free variation, a variation which thus has nothing to do with the vowel of the stem. The breakdown of harmony may here be attributed to contact with Chinese. In Table 3, a number of inflectional morphemes which originally had the vowel *a* with stems in back vowels and *e* in stems with front vowels is shown for the four Eastern Mongolian languages Baoan, Dagur, Dunsian, and Monguor.

TABLE 3

	Baoan	Dagur	Dunsian	Monguor
Plural	le	—	la	—
Ablative	se	se	se	dza
Locative	re	aare	—	—
Instrumental	gale	gala	—	—
Comitative	—	—	le	la
Causative	ge	gaa, gee	ga	ga, ge
Past Participle	sang	sen	sen	dzan

As can be seen we have a kind of cross-linguistic and internal language variability in principle basically similar to that which was encountered in regard to Nilo-Saharan *k-* and Chibchan *-kwa* even though the variants are of phonological rather than morphological origin.

A very similar situation obtains in regard to the so-called Iranicized Uzbek dialects in which, presumably under Iranian influence, the Turkic vowel harmony system, which except for the existence of distinct high back and front vowels has essentially the same structure as Mongolian, has broken down in affixes with results that are basically similar to those found in Eastern Mongolian languages. In neither of these instances is there any semanticization of variant forms. That morphophonemic variants can be utilized to express grammatical distinctions is, of course, well known from the example of German umlaut in which noun pluralization, the expression of the subjunctive in verbs, and that of the comparative and superlative of adjectives are

expressed by umlauted vowels although generally as a subsidiary mark along with affixes of the usual kind.

An interesting example of the breakdown of a vowel harmony system based on height, in a system which still functions in a few limited aspects of the grammar, e.g. the numeral classifier system, and the third person singular prefixed pronominal object of the verb, occurs in Gilyak.

The Gilyak system of vowel harmony is shown in Table 4.

TABLE 4

High	i	y	u
Low	e	a	o

If we compare the two main dialect areas, that of the Amur basin in Siberia, and that of Northern Sakhalin, we see along with numerous instances in which both dialects have generalized the low variant, or both have retained the high variant, a considerable number of cases such as those in Table 5 in which one dialect has chosen the low and the other the high variant.

TABLE 5

Amur	Northern Sakhalin	
yl	al	'mouth'
park	pyrk	'only'
mut	mot	'pillow'
nik	nek	'recently'

In general the Amur dialect prefers the high variant and Northern Sakhalin the low variant but there are a fair number of exceptions. In addition there are instances in which both variants are found in both dialects but have differentiated their meanings, e.g. *vi-*, 'go, walk', *ve-* 'run (of animals)'; *lax* 'cloud', *lyx* 'rain'. In one instance we even have an incipient grammatical-ization, *nog* 'to be fragrant' (intr.) and *nugnug* 'to smell' (tr.). In all the forms just cited the two dialect areas agree.

In addition to morphological elements and morphophonemic alternants, it seems likely that variants developing out of word sandhi can give rise to a similar pattern of cross-linguistic and intralinguistic variation. A well-known instance is the so-called *s*-moveable of Indo-European which appears preceding roots in unvoiced stops, *r, l, m*, and *n*. An example is Latin *tegere* 'to cover' as against Greek *stégos* 'roof' which also has a variant *tégos* in the same language.

Can a similar pattern of distribution result from variants produced by sound change? The usual pattern is either clearly genetic or areal. Thus the Indo-Aryan change of *e > a, merging with original *a which is found only in this branch of Indo-European is clearly genetic. A classic instance of an early areal feature is found in the case of 'incomplete satemization'.

The fronting of front velars to sibilants found in the satem branches of Indo-European, namely Indo-Aryan, Balto-Slavic, and Albanian is only complete in Indo-Aryan and probably Armenian. In Balto-Slavic there are instances of velar reflexes in particular forms, in regard to which the languages differ from each other and even show variation within the same language. Thus corresponding to Sanskrit *sru-* 'to hear' we find Old Church Slavic *slu-ti* as expected but in Baltic Lithuanian *klausy-ti*, Lettish *klausi-t*, and Albanian *guhem* 'I hear'.

Corresponding to Sanskrit *aśman* 'stone' we find forms with velars in both Baltic and Slavic: Old Church Slavic *kamy*, Lithuanian *akmuo* and Latvian *akmens*. However, in Lithuanian there is the doublet *asmuo* with the differentiated meaning 'edge, blade'. In the word for 'dog', Sanskrit *śvan* corresponds to Lithuanian *szuo* as expected but within Latvian we find the surprising variation *suns* 'dog' but *kuna* 'bitch'. Sanskrit *śmasru* 'beard' is connected etymologically to Lithuanian *smakra*, Latvian *smakrs* 'chin', and Albanian *mjakrë* 'beard'.

These and other examples show some degree of randomness in the distribution of the palatal and sibilant but on the whole exhibit an areal pattern which, as noted above, suggests an eastern origin for the sound change with imperfect propagation westward among the Indo-European dialects.

An instance which does seem to show the type of distribution seen above for morphological and morphophonemic variants are the reflexes of the reconstructed syllabic *r* of Proto-Indo-European, in Greek *ra* or *ar* and Germanic *ru* or *ur*. Brugmann in the second edition of his comparative grammar (1897: 4) notes that there is no satisfactory solution to this alternation. In his later summary grammar (1902: 131) he says 'Probably a Proto-Indo-European difference in pronunciation is the reason.' This variation in regard to Greek is illustrated here by a few examples. Athenian *kardía* 'heart' corresponds to Ionian and Homeric *kradía*. Within the Athenian dialect itself we find *kárta* 'very' but *kratús* 'strong' and the variant past passive participles *dartós*, *dratós* 'flayed'. In Pindar who wrote in the Theban Aeolic dialect we even find *thrasùs kardía* 'bold heart' in which the reflexes *ra* and *ar* are found in the same phrase. The choice was probably because of the meter but that both variants were available to him is significant.

An interesting case of secondary semantic differentiation is the existence of the two forms *thrásos* and *thársos* for 'courage' in Athenian and the Standard Koine. Aristotle (Eudemian Ethics 1234b. 12), after his discussion of the golden mean as lying between a particular vice and its corresponding virtue, distinguishes another case, namely that the excess of a good quality is likewise condemned and is to be contrasted with its possession to a moderate and fitting degree. His example is *thársos* 'courage' versus *thrásos* 'foolhardiness'. In a similar vein Ammonianus, a lexicographer of the second century AD, says that *thársos* is said of human beings, that is 'reasonable courage', as against *thrásos*, the unreasoning courage of animals.

I believe that we are to envisage the third alternative as distinct from the genetic and the areal but describable from the sociolinguistic point of view as follows. The protolinguistic community showed variations reflecting changes which were just in progress as it began to break up. They were distributed idiosyncratically across small groups and even individuals. Within each group that later became a separate language there was a specific distribution subject, of course, to later analogical changes and often with an inherited preference for one variant or the other. We may compare these to the 'founder' groups of population geneticists. No particular subgroup represents a perfect sample of the original population in language just as in regard to gene frequencies.

The ensuing results take the form of a random distribution, continuing often for a surprising length of time, such as those that we have found in the examples illustrated in this study. Such ongoing changes in the protolanguage only affect a small part of the total linguistic structures involved but it may be present in any aspect of the language.

Examples such as those discussed are then not amenable to either genetic or areal explanation. They present a problem to the comparativist seeking to reconstruct a total and uniform ancestral linguistic system. If, however, we take seriously the facts about linguistic variation observed and studied by sociolinguists, we will not find such phenomena surprising. It is the thesis of this paper that the kind of random distribution of competing forms discussed here and which could be illustrated by many more examples, is reasonably accounted for by the sociolinguistic factors just mentioned.

References

Brugmann, Karl. 1897. *Grundriss einer vergleichenden Grammatik der indogermanischen Sprachen* (2nd edn.), vol. 1. Strassburg: Trübner.

Brugmann, Karl. 1902. *Kurze vergleichende Grammatik der indogermanischen Sprachen*, vol. 1. Strassburg: Trübner.

Greenberg, Joseph H. 1966. *The Languages of Africa*. Bloomington: University of Indiana Press.

—— 1972. 'Numeral classifiers and substantival number: problems in the genesis of a linguistic type'. *Working Papers in Language Universals* 9: 1–40.

—— 1978. 'How does a language acquire gender markers?' in Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.), *Universals of Human Language, vol. 3: Word Structure*. Stanford: Stanford University Press, 47–82.

—— 1981. 'Nilo-Saharan movable *k-* as a Stage III article (with a Penutian parallel)', *Journal of African Languages and Linguistics* 3: 105–12.

Tucker, Archibald N. 1940. *The Eastern Sudanic Languages*. London, New York: Oxford University Press.

11

Indo-Europeanist practice and American Indianist theory in linguistic classification*

There exists among American Indianists and in linguistics in general no coherent theory regarding the genetic classification of languages. The most common formulation which might seem to be such is that there must be regular sound correspondences between languages for them to be related. The problem as to whether there is an empirically operational method to give a yes or no answer as to the presence or absence of sound correspondences is discussed at length in Greenberg 1987 (especially 1–37) [§6-Ed.] which resumes and expands earlier discussions, particularly Greenberg (1957: 35–45) [§2-Ed.]. But even if it worked, and the objections to it on this ground are shared by critics as diverse as Meillet and Ives Goddard, it is the answer to the wrong question, namely as to whether languages are related.

Suppose regular correspondences really gave an unequivocal answer; then we could prove that Swedish is related to Albanian, that Norwegian is related to Bengali, and that Bulgarian is related to Armenian since they are all Indo-European languages. However, we would still not have a classification. The importance of classification is that it specifies units at various levels which are meaningful and fruitful objects of comparison. From such comparison we derive insights concerning linguistic change which can be generalized by comparing historically independent instances, leading to the generalizations of diachronic typology, important hypotheses regarding human cultural nonlinguistic history, and also the basic background for the study of language contact. That, for example, the suffixed definite article of Bulgarian is a significant areal feature is highlighted by its nonexistence in other forms of

Slavic and by the knowledge of its concrete diachronic source derived from the comparative study of the Slavic languages which constitute a natural taxonomic unit.

A large part of this paper is devoted to the history of linguistics, especially that of Indo-European linguistics, in order to show that the notion of sound correspondences as proving relationship common in textbooks of historical linguistics at the present time is contradicted by the fact that classifications of Indo-European and other families which are widely or unanimously accepted were made long before even the notion of regular sound correspondences existed or the correlated notion of the exceptionless nature of the sound changes which produce them had become part of the conceptual armory of linguists.

The paper may be of further interest in that it advances a thesis regarding the relation between the historical succession of ideas and their logical relations to each other which constitutes a parallel to the manner of drawing conclusions about possible changes in diachronic typology from logical implicational relations in synchronic typology. In synchronic typology, the existence of an implication of the type A . . . B leads to the diachronic conclusion that A cannot arise unless B previously exists, for if A did arise without this precondition we would have A without B, a presumably nonexistent type which violates the implicational relationship. So in intellectual history, an *implicans* cannot exist without its *implicatum* being present, usually previously but in some instances due to the genius of particular investigators, simultaneously.

For the history of linguistic theory, we may exemplify it in the topic which interests us here in the following manner. The notion that a group of languages should be systematically compared to reconstruct their ancestral form, implies that we already have recognized what languages are to be compared in this manner before we attempt a reconstruction; in other words, reconstruction implies classification and not vice versa. So we find, in fact, that the notion of a class of Indo-European languages whose resemblances as a historically significant fact, at least as early as Jones's famous pronouncement of 1786, preceded the first reconstructions by Schleicher in the mid-nineteenth century by at least fifty years. As we shall see, there are other conceptual stages involved in the development of historical linguistics besides the ones just cited.

We will begin with the famous and oft-quoted statement of Sir William Jones in 1786, the bicentenary of which was recently celebrated in a meeting at Rice University. It will be useful to reproduce and analyze from certain points of view this pronouncement.

The Sanskrit language, whatever may be its antiquity, is of a wonderful structure; more perfect than Greek, more copious than Latin, more exquisitely refined than either; yet bearing to both of them, both in the roots of verbs and forms of grammar, a stronger affinity than could have been produced by accident; so strong that no philologer could examine the Sanskrit, Greek, and Latin without believing them to have sprung from some common source, which perhaps no longer exists. There is a similar reason though not quite so forcible, for supposing that both the Celtic and Gothic had the same origin as Sanskrit.

There are several points to be noted about this statement. One is that Jones mentions a whole series of languages. Having, as we know, a broad acquaintance with languages including Arabic, and probably Hebrew, along with his knowledge of the classical languages which was part of the intellectual equipment of educated people in his day, and of Germanic (then called Gothic), the addition of Sanskrit to his repertory as the result of his stay in India where the statement was made, enabled him to see valid groupings based on differential resemblances. In later work he outlined the Semitic and Finno-Ugric families. It is also clear that his mention of verb roots and forms of grammar referred to the resemblance involving form and meaning simultaneously in lexical and grammatical morphemes respectively thus tacitly excluding typological commonalities.

That he mentions verbs is no doubt because he and others of this period were particularly struck with the resemblance of the verbal endings, e.g. the *-m*, *-s*, and *-t* of the first, second and third persons respectively which are identical in Sanskrit and Latin, while those of the noun were less evident. Bopp's work, which is unanimously recognized as initiating the comparative method and which was published in 1816, had as its title 'Concerning the conjugational system of the Sanskrit language in comparison with that of the Greek, Latin, Persian, and Germanic languages'.

In other words, I am asserting that results, universally viewed as valid, were attained by assessing nontypological resemblances as significant in a number of languages against a tacit background of other languages which were not Indo-European and of whose knowledge in the case of Jones we have valid historical evidence. Bopp, also, in his pioneer work of 1816 at one point cites forms from Arabic, showing his acquaintance with non-Indo-European languages.

The linguist of this initial period who most specifically mentions what we would today call multilateral comparison was Rask. Diderichsen (1974: 297) paraphrases Rask as follows: 'the more languages and dialects you take into the comparison, the more gaps you are able to fill by intermediate forms.' Diderichsen then goes on to quote a letter of Rask probably to be dated in

1809 (*Sammlede Afhandlinger* 1: 15) in which he states 'that I have discovered such a fundamental coherence between so distant languages (Greek, Latin, Gothic, Icelandic, German) led me to investigate so many tongues as time would allow.' It should be added that Rask noted that Samoyed, Eskimo, and Aleut had an identical pattern for singular: dual: plural—namely, zero: *k*: *t*—and appreciated its historical significance within what I would now call Eurasiatic.

Let us return to Jones, however, to consider one further point. In a sense, his most radical new step, since in fact people had noted language groupings before, was in his statement regarding Sanskrit, Latin, and Greek, that they had 'sprung from some common source, which perhaps no longer exists.' Note his hesitancy which is echoed in Bopp's epoch-making work of 1816 in which he still hesitates between Sanskrit as the source language and a common ancestor earlier than and not to be identified with any of the languages compared (Bopp 1816: 9).

Soon after, hesitation on this matter evaporates and with it linguistics develops the equivalent of a branching evolution explanation of linguistic similarities and differences, the biological parallel to which did not receive general recognition until the appearance of *Origin of Species* in 1859.

That not only can correct classifications be made but even specific etymological equations, without the recognition of the regularity of sound change or the notion of specific sound correspondences, is shown in Bopp's work. In it there can be found a whole series of what one is tempted to call inspired guesses which as far as I can see are all still generally accepted equations. For example, he finds the Sanskrit verb stem *as-* 'to be' in Latin *eram* 'I was' (for *esam* he says) just as he does in the third person singular present *es-t* = Sanskrit *asti*. He sees that the Latin suppletive stems seen in *est* 'he/she is', but *fu-it* 'he/she/it has been' are to be equated with the two different Sanskrit verbs 'to be', *as* and *bhū*. He also equates the Greek vowel prefix 'augment' of past tenses e.g. *e-pheron* with the Sanskrit *a-* of *a-bharam*.

In doing all of this, he never asks if the Sanskrit *bh* of *bhū* 'to be' is represented by Latin *f* in other instances or in fact which is the earlier sound. He cites equations in which *s* between vowels in Greek is lost but is not at all disturbed by the future *eso-mai* in which it is retained, and he equates the augments of Greek and Sanskrit *e* and *a* simply as short vowels without further specification.

The notion of regular sound change as embodied in Grimm's first law became generally known from the second edition of his *Deutsche Grammatik* (vol. 1, 1822). Grimm had predecessors including especially Rask (cf. Jankowsky 1972: 72–6) but the law goes under Grimm's name for good and

sufficient reasons. Grimm himself was aware of the numerous exceptions and showed no concern about them. In Bopp's later works, in which he extended the scope of his comparison from verb conjugation to all facets of language structure and ultimately added all of the languages known in his time which have been accepted as Indo-European, he makes use of sound correspondences but only loosely and is untroubled by exceptions.

The same is true of Pott, the founder of Indo-European systematic etymology whose *Etymologische Forschungen* first appeared in a two-volume edition (1833–6) and then in much expanded form in a second edition in six volumes (1859–76) and of whom we shall have more to say later.

The next major step was taken by August Schleicher who dominated comparative linguistics in mid-century. This was to make the family-tree metaphor involved in comparative linguistics explicit and to reconstruct the forms of the *Ursprache* which, as is well known, Schleicher did so confidently that he wrote a short fable in Proto-Indo-European.

Although Schleicher reconstructed both Proto-Indo-European phonology and morphology, his posited forms are today universally rejected. If we read his Indo-European fable, we immediately see the reason why. It reads almost like Sanskrit. To illustrate the points at issue, let us consider Schleicher's reconstruction of the Indo-European vowel system as containing three basic vowels as in Sanskrit, namely *a, i, u* and which he equated with the Gothic three-vowel system. It was noted by Scherer in 1864 that European languages fairly consistently had *e*, in other forms *o*, and still others *a*, corresponding to Sanskrit *a*. Scherer's conclusion was that the European languages here shared an innovation which showed they were a single branch as against Indo-Aryan, but he could give no condition for split of *a* into *a, e, o*. By 1878, a number of investigators independently discovered the palatal law of Sanskrit, namely that within Sanskrit itself *ch* occurred before *a* when the European languages had *e*. Therefore, the merger of *e* and *a* (the hypothesis of the merger of *o* and *a* soon followed) must have occurred in Sanskrit, while other languages like Greek showed the original vocalism. The internal evidence of Sanskrit was concordant with the external correspondence between Sanskrit nonpalatal consonants before Greek *a*, and palatals before *e*. Many instances, of course, had to be explained by analogic change which increased paradigmatic unity.

The previous year (1877) saw a discovery which had an even greater impact and which again was made about the same time by several scholars. Verner's publication was the earliest so it still bears the name of Verner's Law. This disposed of almost all the exceptions to Grimm's Law of Germanic consonant shift by bringing in as a condition the position of the consonant in relation to

the pitch accent as it appeared in Sanskrit and Greek but which no longer existed to our knowledge in the earliest written records of Germanic.

These two discoveries and others lent credence to the notion first advanced by Leskien in 1876 that 'sound laws have no exceptions'. A group of linguists, which included Leskien, Brugmann, Osthoff, Delbrück, and Paul, known as the Neogrammarians (*Junggrammatiker*), advanced this thesis vigorously. There were numerous replies and counterreplies especially during 1885–6 (cf. Wilbur 1977) often called the Lautgesetz controversy. From the point of view of the present paper, two things in particular should be noted. One was that in addition to analogy as an explanation of exceptions, a whole series of minor sporadic processes had to be admitted e.g. the irregularity of Proto-Romance *grevis* 'heavy' instead of *gravis* (as in Classical Latin) because of the influence of the vowel of the antonym *levis* 'light'. The whole fight became a controversy about the semantics of the word 'exception'. The Neogrammarians said that such examples did not fall under the sound law and were therefore not true exceptions. Ultimately the Neogrammarians had to soften their statements and the doctrine then amounted to a pragmative directive, an excellent one in fact, namely to look for explanations to exceptions. Also as one can discover from a reading of Collinge (1985), numerous sound laws were advanced whose validity was attacked by other scholars and which, in large part, remain unresolved to the present day.

The second thing to be noted is that if one reads the extensive documentation in Wilbur 1977, nowhere does it occur to anyone that if sound laws are not exceptionless, one cannot prove the relationship of the Indo-European languages. It is rather that this relationship is the premise on which all comparative Indo-European work is based. Concerning the great etymologist Pott, Bloomfield (1933: 15) states without reserve: 'Our modern etymologies in the Indo-European languages are due largly to the researches of Pott.' Yet this same Pott, by then a venerated figure, in the last year of his life (1887) launched a blistering and sarcastic attack on the Neogrammarians. For him, Schleicher's three vowel system *a, i, u* was still good enough, yet his labors are the foundation stone of present-day Indo-European etymology.

As to how one actually classifies languages, both the Neogrammarians and their opponents say almost nothing. The only statement that I find in this regard is that of Delbrück, the collaborator with Brugmann on the *Grundriss* and the author of the classic two volumes on comparative syntax contained in it. In a general work designed to set forth basic Neogrammarian views on historical linguistics, he states simply (1904: 121–2):

My starting point is that specific result of comparative linguistics that is not in doubt and cannot be in doubt. It was proved [*erwiesen*] by Bopp and others that the

so-called Indo-European languages were related. The proof [*Beweis*] was produced by juxtaposing [*Nebeneinanderstellung*] 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.

Nowhere in this passage does he use terms like 'sound law' or 'sound correspondence' or even *Vergleich* 'comparison' which would imply something more than mere *Nebeneinanderstellung* 'juxtaposition', here the key word. Nor is reconstruction mentioned.

It is the great achievement of the Neogrammarians that both their discoveries regarding Indo-European with respect to change and reconstruction and their method are still in all essentials the basis of comparative linguistics.

There followed comparative grammars in the Neogrammarian tradition such as those of Meyer-Lübke in Romance, Prokosch in Germanic, and Pedersen in Celtic. All of these used as their starting point reconstructed Proto-Indo-European based on the work of Brugmann and his associates. How, for example, could one write a comparative grammar of Germanic and disregard the explanatory force of Verner's law? Yet the evidence for it all lay outside of Germanic itself in other branches of Indo-European. This is all contrary to the contemporary view of many that one must first reconstruct the protolanguages of the smaller groups and then compare these to arrive at a reconstruction which will 'prove' the validity of the larger group.

Another major development was the application by linguists in the Neogrammarian tradition of their methods to non-Indo-European groups of languages, e.g. Dempwolff in Austronesian, Meinhof in Bantu, Brockelmann in Semitic, and Collinder in Uralic. One will search in vain in these writers for a single statement indicating that they believed they were proving the validity of the linguistic stock to which they applied the comparative method of reconstruction, usually to a small number of languages if the family was very large (e.g. Austronesian).[1] As with the Indo-Europeanists, they rather looked on the validity of the stock they were studying as the underlying assumption on which all their work was based. In line with later thinking, they all believed that sound laws had exceptions and some even believed, e.g. Meinhof, that the protolinguistic reconstruction was merely a convenient fiction.

Until Collinder (1960), there was no reconstruction of Uralic which included the vowels. Uralic is a family whose existence was discovered even

---

[1] For further details concerning the relationship between sound correspondences, regular sound change, reconstruction, and language classification in the work of Meinhof, Dempwolff, and others who did reconstructions for non-Indo-European languages, see Greenberg (1987: 31–2) [§6-Ed.].

before Indo-European and whose validity not even the most sceptically minded doubts. In this work he states (p. 405) that 'it is a matter of course that in many instances the reconstruction of a PU (Proto-Uralic) or PFU (Proto-Finno-Ugric) word is more uncertain than the etymology on which it is based.'

As historical linguistics grew and became more specialized, the majority group (Indo-Europeanists) simply took for granted their initial premise and never sought to analyze the processes by which the original Indo-European or similar hypotheses arose. When confronted with new languages discovered in inscriptions, manuscripts (e.g. Tokharian), or cuneiform tablets, Indo-Europeanists reacted in a sensible way and never sought to prove the affiliation of new languages by deducing their sound system from the by-then standard Brugmannian reconstruction.

If one had, about 1900, told Indo-Europeanists that a new group of Indo-European languages would be discovered in Asia Minor with records beginning as early as the sixteenth century BC, they would have been unanimous in predicting that in both sound system and grammatical structure it would be similar to such languages as Vedic Sanskrit and Classical Greek and given the still earlier date of first attestation even closer to the Brugmannian reconstructed Proto-Indo-European than these languages.

However, a number of languages were discovered written in cuneiform script whose Indo-European affiliation is doubted by no one. The earliest deciphered and the one with the largest body of textual material is Hittite. One would think that Indo-Europeanists would have recognized and sought to prove their Indo-European character by sound correspondences and by deriving Hittite forms from the generally accepted form of the reconstructed protolanguage. However, as we know, these languages were drastically different from the other earliest known Indo-European languages. So far from providing a further confirmation of the by-then standard reconstruction, the results of what might be called Anatolian shock, a shock from which Indo-Europeanists have as yet not fully recovered, were to force us to revise in truly essential ways our theories about Indo-European, a revision which certainly has not yet produced unanimous agreement. The Indo-Europeanists quite simply agreed to accept these languages as Indo-European in spite of all this. The kinds of arguments presented by Hrozný, who first convinced the learned world of the Indo-European character of Hittite, were expressed in his work of 1917 in which he stated (p. vii):

Everyone who wishes to interpret the Boghazkoi texts from the moment of their publication will, like the author, come to the same conclusion on the basis of instances like the fact that *wadar* means 'water', that its genitive is not *wadaras*

but remarkably enough *wedenas*, that the Hittites have a participle in -*nt*, that 'what' (masc.) is *kuis* and in the neuter *kuid*, that 'I' is *ug* (cf. Latin *ego*), 'to me' *ammug* (cf. Greek *emoige*), 'thou' *zig* (cf. Greek *suge*), 'to thee' *tug* (Gothic *thuk*, etc.), that the Hittite present is inflected *jami, jasi, jazi, jaweni, jatteni, janzi* etc. etc.

Note that Hrozný does not cite sound correspondences and that the resemblances cited are with various Indo-European languages or with none in particular as in regard to the verbal paradigms. It is, in other words, against a broad background of knowledge of Indo-European languages and what is widespread and diagnostic of them as a group and not by a hypothesis of relationship to or a comparison with a single language that Hrozný presented the thesis now universally accepted by Indo-Europeanists. What is more, the Hittite writing system, borrowed from Akkadian, a Semitic language (and that in turn from Sumerian), presents many problems. It has syllabic signs for $C + e$ in only a limited number of cases, some believe it had an *o* which could not be written in the cuneiform syllabary; it indicated consonant gemination sporadically, that is, the same form may appear with or without it in the same word in different texts. Sturtevant, in fact, devised a rule named after him, accepted by some and not by others, that where a geminate consonant is ever written in any form of a word, it indicates an unvoiced stop; but where never written, it is voiced. In spite of the fact that the Hittites were poor transcribers of their own language and that numerous basic words are written as ideograms, its Indo-European nature is not in doubt.

An even more astonishing case is Lydian, an Anatolian language known from a much later period than Hittite and accepted by Indo-Europeanists as forming a separate branch of Anatolian. It is accepted as such in the eminently orthodox comparative Indo-European etymological dictionary of Pokorny, which lists exactly four Lydian words in its index! This is partly because words common to other Anatolian languages but not found elsewhere in Indo-European are not included and because grammatical formatives are not contained in Pokorny's dictionary. Basically Lydian is considered Indo-European because it is Anatolian. One of the signs of the Greek-derived alphabet is transcribed as *p* by some writers and *q*, indicating probably a labiovelar *kw* by others. There are two signs for *l*, one being conventionally transcribed by the Greek letter *lambda*. Both seem to correspond to Indo-European *l* but the reason for the appearance of one or the other is subject to no known rule.

As a final case in point, let us consider Venetic, known from less than 200 mostly very short inscriptions from the area near Venice. Many of these are single letters or consist of the letters of the alphabet in sequence. The others

are very short and monotonous in subject. There are here also some signs used inconsistently and with disputed phonetic values.

   Like Lydian, it is universally recognized as Indo-European. Unlike Lydian, the problem is to which subgroup of Indo-European it belongs: Italic, Illyrian, or perhaps a branch of its own. The standard work on Venetic is that of Beeler. The reasons he gives for considering Venetic an Indo-European language are the following (1949: 13):

> (a) The contrast between the inflectional system of two series of names, one in -*os*, -*oi* and -*on* (like the nominative, dative and accusative singulars, respectively, of the Indo-European ŏ stems) and the other with -*a*, -*as*, and -*ai*- (like the nominative, genitive, and dative singulars of the IE ā stems);
>
> (b) The verbal ending *to*, presumably that of the third person singular of the secondary indicative middle, Greek -το, Sanskrit -*ta*;
>
> (c) A large number of derivative suffixes, e.g. -*i̯o*-, -*no*-, -*so*-, -*tor*-, which can be abundantly paralleled in the languages of the Indo-European family;
>
> (d) Many striking lexical correspondences, such as ·*e·χo* (= Latin *ego*), *me χo* (= Gothic *mik*), *zoto* (= Greek εδοτο), *lo·u·zera·i·* (= Latin *Libera*);
>
> (e) The characteristically Indo-European nature of the vowel alternation in *vho·u·χo·n·tah* and *vhuχiia*.

Once again, and this could be repeated for every language whose Indo-European affiliation has been recognized in the post-Brugmann era, there is no mention of sound correspondences, but rather references to concrete sound–meaning resemblances in forms diagnostic of Indo-European as a whole, considered against a general background of knowledge of the family rather than by comparison with some single member to prove relationship.

   In a recent publication regarding Amazonian languages, Kaufman makes such statements as the following (1990: 18): 'I believe that between 500 and 600 items are necessary . . . about 100 points of grammar of the type presented by affixes and particles should be among the items compared.' Whence these large and arbitrary numbers? Were they to be taken seriously, as we have just seen, the Anatolian and other languages known from limited inscriptional material could not be considered. Yet the study of the Anatolian languages in particular has become the very cornerstone of Indo-European studies. Kaufman states, 'Outlines of the grammars of all languages should be known' (1990: 24). In the case of Venetic, no first or second person verb forms have been recorded because of the nature of the inscriptions.

   We have seen that in actual practice no Indo-Europeanist used sound correspondences to prove the Indo-European affiliation of newly discovered languages, nor in the extended discussion of sound laws in the 1870–1880s did

the notion that sound correspondences were relevant in any way to the establishment of the Indo-European family arise. Statements by Neo-grammarians about how to do classification are virtually never found in the literature of this period. The citation from Delbrück is the only one I could find. I think this is so for two reasons. By then historical linguists were all specialists in some well-established stock, mostly Indo-European, and took the results of classification for granted. Further, as we have seen in the statement of Delbrück cited above, the procedure seemed so simple and obvious that apparently it needed no extended discussion or the development of a body of theory.

What has come about more recently is that regular sound changes, an important but by no means exclusive weapon in the armory of reconstruction, but one which only comes into play in an essential way after one has determined which languages to compare, have become misinterpreted as 'proof' of relationship. Obviously one cannot apply the comparative method until one knows which languages to compare. This point is clearly stated in Newman (1970: 39):

The proof of genetic relationships does not depend on the demonstration of historical sound laws. Rather the discovery of sound laws and the reconstructions of linguistic history normally emerge from the careful comparison of languages already presumed to be related.

The two stages are distinct although obviously related. Sound correspondences do not spring out armed like Athena from the head of Zeus. They start from the comparison of forms of similar sound and meaning noted in the initial stage. This is followed by a period of refinement, which seems invariably to add new etymologies and almost never invalidate the initial ones. Some irregularities are never satisfactorily explained. For example, in Greek there are often initial vowels preceding a sonant, which do not appear in other Indo-European languages e.g. the initial vowels of *a-leipho* 'I smear' = Hittite *lip-*; *e-leutheros* 'free' = Latin *liber*; and *onoma* 'name' = Latin *nomen*. These initial vowels created difficulties from the beginning of systematic Indo-European studies but did not ever result in the rejection of the Greek forms as valid etyma, much less the Indo-European affiliation of Greek. With the coming of laryngeal theory Rix sought to explain *e-*, *a-*, and *o-* respectively as reflexes of an *e-*, *a-*, and *o*-colored vowel following lost original laryngeals H_1, H_2, and H_3. However, there is often no independent evidence for these laryngeals and Rix's law, as it is called, is rejected by some Indo-Europeanists. Whether one considers the initial classification as part of the comparative method, that is, as its first stage or not is simply a matter of definition. It seems more consistent to include it.

At any rate, it somehow became common to confuse the later stages with this first stage to the extent that we find in the second edition of Anttila's well-known textbook of historical linguistics the following statement (1989: 318): 'The basic criterion is simply sound correspondences. Languages that fit into regular sound correspondence belong to the same family.... sound correspondences provide genetic classification....'

It would require further historical research to find out how the confusion of the later stages with the initial stage occurred, but I strongly suspect that Herman Moeller's attempts beginning in the first decade of the twentieth century to show the genetic relationship of Indo-European and Semitic played a key role here. The notion that Indo-European and Semitic, without reference to the other branches of Afroasiatic, and in disregard of the greater resemblance of Indo-European to Uralic and other languages of northern Asia, forms a valid genetic group is obviously a linguistically unnatural hypothesis motivated by nonlinguistic considerations.[2] Moeller, seeing very little that was obvious, resorted to reconstruction and sound correspondences to prove the relationship. The Indo-Europeanists' general response was negative. As specialists, they wished to pursue their studies in isolation. Whence the dogma of the immaculate conception of Indo-European and in conformity with this of many other families.

An obvious example of the prevalent confusion of the initial stages of classification and the later ones of the systematic study of sound correspondences and reconstruction is that of the Armenian development of Proto-Indo-European initial *dw-* to *erk* in the words for 'two', 'fear', and 'long', cited by Thomason (1990). This example is simply irrelevant as the problem under discussion is classification. How do we know that Armenian is Indo-European given the previous recognition of a group of Indo-European languages without the affiliation of Armenian as a distinct subgroup? The decisive event was the publication of what Meillet was to call the 'beautiful article' (*bel article*) of Hübschmann in 1875 which showed that Armenian was an independent branch of Indo-European and not a dialect of Iranian as had been previouly thought. Meillet indeed called Hübschmann 'the indisputable master of studies of Armenian linguistics' (1936: 7).

If all the correspondences between Armenian and the other Indo-European languages were of the type cited by Thomason, how could the relationship ever have been discovered? In fact in his decisive article of 1875 this

---

[2] A leading Nostraticist, Starostin, has noted that Afroasiatic is only more remotely related to Indo-European than the rest of Nostratic. He states (1989: 43) that he prefers for the present to exclude Afroasiatic material from his Nostratic comparisons.

correspondence is never mentioned and no etymology is given for 'two' and the other words mentioned above. The correspondence, called by Godel 'amazing' was discovered by Meillet in 1894 and was never accepted by Hübschmann. It is discussed and rejected by him in his Armenian grammar of 1897 where he considers the sound change not *erk < dw* but *k < d*, as others have done, with the *er* of *erku* as the result of contamination with the initial of the word for 'three' *er-ek'*. This interpretation has been accepted by others e.g. Winter (1962) but Hübschmann rejects it even in this form. At present it is generally accepted in the formulation of Meillet but clearly Armenian is an independent branch of Indo-European whether one accepts it or not and it therefore has nothing to do with classification.

The existence of a number of unobvious correspondences has indeed made the study of Armenian historical phonology notoriously difficult. Still there are a large number of obvious similarities between Armenian and the other Indo-European languages. Let those who think that phonetic similarities are irrelevant, and who do not understand that the discovery of widespread similarities is the starting point of classification, produce a list of what are by many called disdainfully 'look-alikes' similar to the one which follows, comparing Armenian to Afroasiatic or some other stock.

In what follows I cite similarities which, in many cases are close to identities between Armenian and other Indo-European languages. To simulate as closely as possible my own method of multilateral comparison in areas where earlier written records are typically absent, I have confined myself to contemporary Indo-European languages.

1. *ak-n* 'eye', pl. *ač-k'*. Italian *occhio* (pronounced *okkyo*); Bulgarian *oko*; Russian *ochi* 'eyes'.
2. *alam.* 'I grind, crush'. Greek *aletho* 'I grind'.
3. *amis* 'month'. Italian *mese*; Welsh *mis*; Nepali *mas.*
4. *anjuk* 'narrow'. Icelandic *öngvar* 'straits', *öng-vegi* 'narrow way'; Polish *wāski* 'narrow'.
5. *anwan* 'name'. Welsh *anw*; Breton *ano.*
6. *anurj* 'dream'. Greek *onir-on.*
7. *armuk-n* 'elbow'. Ossete *arm* 'cupped hand'; English *arm.*
8. *astl* 'star'. Italian *stella*; German *Stern*; Dameli (Indo-Iranian) *ishtari.*
9. *ayl* 'other'. Welsh *ail*; Greek *all-os.*
10. *beran* 'mouth'. Lithuanian *burna*; Irish *bearn* 'fissure'.
11. *ber-em* 'I bring'. English *bear*; Russian *ber-u* 'I take'; Serbo-Croatian *ber-em* 'I gather'; Panjabi *bar-na* 'to bear'; Icelandic *bera* 'carry, support'.

12. *barjr* 'high', *berj* 'height'. Persian *burz* 'mountain'; German *Berg* 'mountain'.
13. *buc-an-em* 'I nourish, feed', *buc-i* (aorist). Oriya *bhoja* 'food, meal'; Bengali 'cooked or served food'.
14. *caneay* 'knew'. Lithuanian *zinoti*; Russian *znaju* 'I know'.
15. *cer* 'old, old man'. Ossete *zäronk* 'old'; Persian *zar* 'old'.
16. *cin* 'birth'. Hindi *jan-na* 'to bear, be born'.
17. *cun-r* 'knee', pl. *cunk-k*. Persian *zanu*; French *genou*; Pahari (Aryan) *janu*.
18. *cax* 'branch'. Lithuanian *szaka*; Persian *shax* 'twig, branch, antlers'.
19. *du* 'thou'. German *du*; Italian, Spanish *tu*; Oriya *tu*; Ossetic *du*.
20. *d-ne-m* 'I put, place' (aorist *e-di*). Lithuanian *de-ti* 'put', Russian *de-t'* 'to put', *de-nu* 'I will put' (perfective with future meaning).
21. *duř-n* 'door'. Russian *dver'*; Demeli (Indo-Aryan) *dar*; Persian *dar*; English *door*.
22. *dustr* 'daughter'. German *Tochter*; Russian *doč'* (gen. sg. *dočeri*); Bulgarian *dəšterja*.
23. *eln* 'deer'. Bulgarian *elen*; Welsh *elain*.
24. *em* 'I am'. Persian *am*; Albanian *jam*; English *am*.
25. *erek* 'evening'. Icelandic *rökkur* 'darkness, twilight'.
26. *es* 'I'. Latvian *es*; Lithuanian *ash*; Bulgaian *az*.
27. *hin* 'old'. Breton *hen*; Lithuanian *sen-as*; Latvian *sen-s*.
28. *im* 'my'. Spanish *mi*; Russian *moj*, etc. etc. Compare also *m* Armenian first person singular and plural of verb in the present and *me-k* 'we'.
29. *inn* 'nine'. Greek *ennea*.
30. *jer* 'warm weather', *jerm* 'hot'. Lithuanian *garme* 'heat'; Russian *gor-et* 'to burn'.
31. *jil* 'tendon, rope'. Russian *zhila* 'tendon, sinew'.
32. *jmeŕn* 'winter'. Russian *zima*; Lithuanian *zhema*.
33. *kin* 'woman', pl. *kanay-k*. Icelandic *kona*; Greek *gineka* (modern Demotic form derived from Classical singular oblique and plural stem).
34. *kov* 'cow'. German *Kuh*; English *cow*; Phalura (Dardic Indo-Iranian) *gu*.
35. *lizem* 'I lick'. Persian *lisidan* 'to lick'; Russian *lizat* 'to lick'; Lithuanian *liesz-ti* 'to lick'.
36. *lkʻanem* 'I leave', (aorist *lkʻi*). Lithuanian *lik-ti* 'to leave'.
37. *loys* 'light', *lucanem* 'I kindle'. Italian *luce* 'light'; Russian *luč* 'ray'; Welsh *llug* 'shining, brilliance'; Icelandic *lysa* 'shine'; Danish *lys* 'light'.
38. *luanam* 'I wash, bathe' (aorist *lua-ci*). Italian *lava-re* 'to wash'; Greek *lu-ome* 'I wash'.

39. *lucanem* 'I release, loosen' (aorist *luci*). English *to loose*; German *lösen* 'to loose'.
40. *manuk* 'child', *man-r* (gen. sg. *man-u* 'small, thin'). Lithuanian *menk-as* 'few, unimportant'.
41. *meg* 'fog'. Russian *mgla*; Lithuanian *migla* 'cloud'; Oriya *miha* 'rain cloud'.
42. *melr* 'honey'. Albanian *mjal*; Welsh *mel*; Greek *meli*.
43. *meranim* 'I die' (aorist *meray*). Persian *mirad* 'dies'; Lithuanian *mir-ti* 'to die'; Russian *u-miraju* 'I die'; Italian *morire* 'to die'.
44. *mi* 'one'. Greek *mia* (fem).
45. *mis* 'flesh'. Bulgarian *meso* 'meat, flesh'; Latvian *mieso*; Marathi *mas*; Kashmiri *maz, mas*.
46. *mit* 'meaning, sense'. Welsh *meddwl* 'to think, thought'.
47. *mizem* 'I urinate', *mez* 'urine'. Latvian *mieznu, miezu* 'I urinate'; Serbo-Croatian *mizam* 'I urinate'; Kurdish *miz* 'urine'; Ossete *miz-an* 'to urinate'.
48. *mnam* 'I remain'. Persian *mandan* 'to remain'; Greek *meno* 'I remain'; Kurdish *man* 'remain'; Italian *ri-manere* 'to remain'.
49. *mor-an-am* 'I forget'. Lithuanian *mirsti* 'to forget'.
50. *mux* 'smoke'. Irish *much* (with *s* moveable English *smoke* etc.)
51. *nist* 'seat'. English *nest*; Welsh *nist*.
52. *oskr* 'bone'. Cornish *ascorn* 'leg'; Welsh *asgwrn* 'leg'.
53. *ost* 'branch'. German *Ast*.
54. *-s.* 'second person singular of verb in the present and other tenses'. The citation of resemblant forms in other Indo-European languages is here superfluous.
55. *sar* 'top, peak'. Persian *sar* 'head'; Pashto *shir* 'head'; Oriya *sira* 'head'.
56. *sar-n* (gen. sg. *sar-in*) 'ice'. Russian *seren* 'frozen snow'; Latvian *sern-s* 'frost'.
57. *sirt* 'heart'. Russian *serdce*; Serbo-Croatian *srce* (gen. pl. *srdaca*); Lithuanian *sird-is*; Latvian *sird-s*.
58. *shun* 'dog'. Lithuanian *shuo* (gen. sg. *shun-s*); Hindi *suna*.
59. *tam* 'I give'. Russian *dam* 'I will give' (perfective present with future meaning); Italian *da-re* 'to give'; Hindi *de-na* 'to give'.
60. *tasn* 'ten'. French *dis* [*dix*]; Russian *desjat*; Oriya *dasa*; Bihari *das*.
61. *tur* 'gift' (cf. 59). Greek *dor-on*; Russian *dar*.
62. *us-anim* 'I learn' (*us-ay*, aorist); Russian *uči-t'* 'to teach'.
63. *utem* 'I eat'. Lithuanian *edu*; Russian *jedim* 'we eat'; English *eat*.
64. *ut* 'eight'. Italian *otto*; Kashmiri *ə:th*; English *eight*.
65. *varem* 'I kindle', *varim* 'I burn'. Lithuanian *vir-ti* 'to cook'; Latvian *var-it* 'to cook'; Russian *var-it*, 'to cook'.

Almost all of these etymologies have, of course, additional reflexes in other Indo-European languages, many of which are almost as similar to the Armenian forms as those just cited. I have noted only two accidental 'look-alikes', Armenian *her* 'hair' and *tap* 'heat' the latter generally agreed not to be related to Hindi *tap-na* 'to warm onself', Russian *tep-lyj* 'hot', etc.

In a recent statement (Bateman *et al.* 1990; the linguistic section was presumably written by Goddard), after first enumerating common origin, borrowing, accident, and sound symbolism as sources of resemblances, without justification and without looking at the facts about languages, the conclusion is drawn that since there are alternative explanations to the genetic, resemblances of these other kinds become so overwhelming in number as soon as one passes from the most obvious low-level groupings, that genetic groupings of a deeper sort are impossible to discover. An actual examination of the data, as in the case of Armenian, shows that this is a myth resting on monumentally incautious statements made by linguists who have not bothered to look at the empirical data on a broad scale.

Also lacking is attention to, or understanding of, the probabilistic considerations that underlie all empirical science. No distinction is made between the few stray resemblances which characterize comparisons between unrelated or distantly related languages when they are taken in isolation as against the massive resemblances characteristic of languages within the same genetic grouping. Neither is there recognition that phonetic change normally results in phonetically similar sounds. There are perhaps hundreds of attested instances of the type $p > f$ while $dw > erk$ has, as far as I know, no parallel.

Moreover there is no understanding, as indicated by the $erk < dw$ example, that the whole enterprise of comparative linguistics does not give, as it were, a single sudden and perfect result, but rather is itself a process in which certain conclusions both logically and in the development of historical linguistics itself precede others. It is unfortunate that historical linguistic textbooks, almost invariably written by Indo-Europeanists, give finished results in apparently completely established protosystems of phonology and grammar without any indication of the historical processes by which they were attained. This suggests also that they are immutable, but as we have already seen from our sketch of the history of Indo-European studies, they are but successive approximations, often superseded by new and at least partially different systems on the basis of new data and our advancing understanding of both synchronic and diachronic typology. Even 'look-alikes', weeded out in earlier study of a linguistic stock may be rehabilitated in later theories. For example, a favorite example of how mere surface similarity is misleading has been the resemblance of English *have* and German *haben* to Latin *habere* 'to

have', whereas the Germanic forms are rather related to Latin *capere* 'to seize'. However, in the new glottalic theory accepted by a large number of Indo-Europeanists, and ultimately resting on typological grounds, aspiration is a variable property of the root and variation between voiced and unvoiced aspirated roots is allowed. The Germanic forms are now assigned to the same basic root as Latin *habere* and what is now its variant *capere* (Gamkrelidze and Ivanov 1984: 146).

It might be thought that even the foregoing list of Armenian 'look-alikes' is invariably tested by regular sound correspondences and that those which do not conform are forthwith rejected. Actually there are remarkable deviations which have never been satisfactorily explained and yet the etymological connections are not rejected by anyone. Thus *es* 'I' should come out *ec* and *du* 'thou' should be *tu*. Although no really satisfactory solution has ever been suggested, they have never been rejected. An important factor is that *es* in the nominative alternates with forms based on *m* in most of the oblique cases and this is a general Indo-European irregularity, cf. *I* and *me*. In the case of *du* we have two major markers of second person in Indo-European, *t* in the second singular independent pronoun and in the verb plural affix as against *s* in the second singular verbal affix. In addition the independent pronoun base is followed by *u/w* so the agreement in this pattern by Armenian obviously outweighs the phonological deviations.

From Indo-Europeanists we now turn to the Americanist approach to classification, one derived from the mainstream of later Indo-Europeanist thought on the subject. The most detailed recent review of principles and specific classifications (without, however, the inclusion of South America in spite of the inclusive title) is Campbell and Mithun (1979). The editorial introduction contains a classification of North and Central America that is 'more conservative than any since Powell' (p. 19). The methodological parts of the introduction leave one puzzled. Thus Levine's critique of Sapir's Na-Dene with its inclusion of Haida is endorsed. 'Haida, an isolate, is demonstrably not related (Krauss this volume)' (p. 39). Turning to Krauss, we find a further definitively negative statement. 'We owe it to Levine in a recent paper for debunking once and for all the claim that Haida has been demonstrated to be genetically related to Tlingit' (Krauss 1979: 841). Further, Levine apparently also has Goddard's approval (e.g. Goddard 1986: 191) where we are told that Levine (1979) is just the sort of study we need 'by critically examining the specific claims and proposals made by Sapir'.

Levine examines the ninety-eight etymologies in Sapir (1915), and sets up eight essentially irrelevant criteria for rejecting etymologies based for the most part on exactness and recurrence of sound correspondence. Campbell's,

Mithun's, Krauss's and Goddard's endorsement of Levine's article suggest that they adhere to the widely held view that relationships are proved by regularity of sound correspondence. It was my first intention to provide evidence for the validity of Na-Dene by presenting the supporting lexical and grammatical data. Sapir did not take Eyak into consideration and I was able to consult the excellent dictionary of Krauss. Moreover, numerous new etymologies had been advanced in the important work of Pinnow. I worked independently for a period in the 1970s, adding new etymologies. I then began to discover Pinnow's work which basically coincided with mine but was more ample, given his specialized knowledge of these languages. Since he had not convinced Americanists, I felt more evidence falling on deaf ears would be ineffective.

Therefore I decided to take a different tack for two reasons. One was that those who endorsed Levine's article as decisively negative had not even taken the elementary step of asking how many etymologies survived Levine's attack. Levine had eliminated the same etymologies again and again by different criteria. The other reason was that Levine's criteria were so restrictive, and what is more, so irrelevant when confronted with what we know about the process of linguistic change, that it is doubtful whether Indo-European would survive a similar attack. Of Sapir's three-way etymologies involving Athabaskan, Tlingit, and Haida, seventeen withstood all of Levine's attacks. I then searched Walde and Pokorny's standard Indo-European etymological dictionary for three-way etymologies involving Celtic, Albanian, and Armenian. I found thirty-seven as against thirty for Athabaskan, Tlingit, and Haida, hardly surprising in view of the fact that Indo-European has been cultivated by innumerable specialists for well over a hundred years. An application of Levine's criteria to the Indo-European etymologies leaves only six etymologies intact. I allowed vowel Ablaut variants, the two forms of *r* in Armenian, whose variation has never been completely explained and the sporadic appearance of initial *h* in the same language, likewise unexplained. Applied strictly, Levine's criteria would probably destroy every Indo-European etymology involving these three groups.

From this it is clear that if by chance the various branches of Indo-European had been spoken in native California and scholars like Kroeber, Dixon, and Sapir had suggested that they form a stock such as Penutian, its validity would have been rejected. To take but one example among many, Hirt, who wrote the first comprehensive grammar of comparative Indo-European after the famous *Grundriss* of Brugmann and Delbrück, notes in the initial section of his volume on the Indo-European vowel system that the first reaction is one of bewilderment (1927: 5). For example, Sanskrit *a* can on

occasion correspond to any of the five Latin vowels *a, e, i, o,* and *u.* One can imagine how most Americanists would have reacted to this and would have been led by these and innumerable other irregularities and complexities to reject Indo-European as a valid stock.

The foregoing discussion has been predicated on the notion that most Amerindian specialists require that relationships be proved by the application of the comparative method based on regular sound correspondences. This view would certainly be supported by the various attempts to relate languages found in the pages of the *International Journal of American Linguistics* which have invariably cited sets of sound correspondences and a reconstructed set of protophonemes preceded by asterisks.

However, we are in for a surprise. In spite of their endorsement of Levine's article, Campbell and Mithun in their initial chapter, refer us to Goddard (1975) for further enlightenment regarding proof of genetic relationships. In terms reminiscent of Newman's statement cited earlier, Goddard is quoted as saying 'In general, the establishing of phonological correspondences goes on within a family of languages known to be related. . . . ' Moreover, according to Goddard as quoted by Campbell and Mithun, loanwords can show regular correspondences also, the instance given being French loanwords in English. In fact, the same example is given in Greenberg (1957) [§2-Ed.] and, for aught I know, by earlier writers. This is accompanied by the statement that 'in fact it is virtually impossible to prove a distant genetic relationship on the basis of lexical comparisons alone.' But as I have shown (Greenberg 1987: 24) [§6-Ed.], the comparison of basic vocabulary on a broad scale leads immediately to correct results when applied in Europe. Surely other parts of the world must be essentially similar in this respect. For example the relationship of Tibetan to Karen in Burma is a very distant one, but Sino-Tibetan appears as a valid entity in broad vocabulary comparisons. Moreover, there is no difficulty in this area of distinguishing groups like Austroasiatic from Sino-Tibetan in this manner even though both families have a minimum of grammatical apparatus that can be used for comparison.

In fact, Goddard (in Bateman *et al.* 1990: 181) endorses the obvious, by accepting the table of European languages mentioned above and noting that it correctly classifies the languages of Europe and even their subgroupings but he regards these groupings as particularly obvious. This is similar to Thomason's view that no doubt I classified African languages correctly but that they are easier to classify than American Indian languages. As Newman states (1995) it seems easy after it is done. Still Kaufmann (1990) states that the Hokan languages are no more divergent than Indo-European. So why shouldn't the same methods work here also?

A further and even less defensible method, namely the carrying out of a reconstruction as proof of relationship is advocated by Kaufman (1990). It will be discussed later as part of the discussion of his paper regarding the genetic status of Subtiaba.

From the foregoing discussion it might appear that no significant resemblances among the 200 or so stocks in the Americas exist. On the contrary, the presence of such widespread resemblances has come to be recognized in the doctrine of Pan-Americanisms. The first reference I find to this, though without use of the term is in Campbell and Mithun (1979: 54). 'Similarly widespread forms in many American Indian languages do little to support a suspected closer relationship between smaller subsets of these language which may come to be compared.' The internal contradiction involved in this approach is obvious on the surface. How can there be the 'closer relationship' referred to unless there are wider relationships?

The first use of the term Pan-Americanism for these sorts of resemblances appears to be in Campbell and Kaufman (1980: 153). In criticizing the Brown and Witkowski article on Mixe-Zoque, they talk of 'widespread forms, Pan-Americanisms'. Campbell and Kaufman assert that any attempt to relate Mayan to Zoque must exclude fourteen of the proposed etymologies, and this in an article limited to roots containing velars. If, as is likely on a rough estimate, one-third of Mayan roots contain velars, there must be about forty Pan-Americanisms shared by just two languages of the same subgroup of Penutian, so their number must be truly enormous.

But how are they to be explained? Bright (1984: 15) also discusses them and says it is not necessary to resort to a genetic explanation. That numerous similar forms extending over two continents, and each with its own distribution can have language contact as their total explanation defies common sense and historically documented evidence from other parts of the world. Even in the Balkans, an area which has been the subject of extensive study precisely because it is an area of intensive language contact, resemblances due to contact, are, on the whole, easily detectable. There is no difficulty in recognizing Serbo-Croatian, Bulgarian, and Macedonian as Slavic Indo-European languages, Romanian as a Romance language of the Italic branch of the same family, and Albanian and Greek as further branches of Indo-European.

No doubt to avoid an explanation by language contact over the length and breadth of two continents, Bright (1984: 25) has even suggested that he would not 'oppose a hypothesis that American Indian languages must have had relations of multilingualism and extreme language contact, perhaps in an age when they were crossing the Bering Strait from Siberia to Alaska.' This hardly requires comment.

From the previous discussion it would appear that those who adhere to a doctrine of Pan-Americanisms either wish to explain them by borrowing or perhaps leave the question of their explanation indeterminate pending future research. However in a publication which appeared after the present conference, Kaufman (1990: 26) says regarding Pan-Americanisms that there are a great number of similarities among American Indian languages and that these similarities are not due to borrowing. In an immediately preceding section he notes that I and Swadesh have asserted that all the languages of the New World except Na-Dene and Eskimo-Aleut are genetically related to each other. He then goes on to say that other scholars including Mary Haas, Victor Golla, and himself have opined that this may be true. If the exception of Na-Dene and Eskimo-Aleut carries over, as seems to be likely from the context, I fail to see that much is left in dispute except the no doubt important problem of the subgrouping of Amerind.

Regarding the history of the notion of a threefold classification I wish to set straight the historical record at least in accordance with my knowledge of it. I made this proposal in a paper given in 1956 at the Congress of Anthropological and Ethnological Sciences. As is usual with large congresses there is a considerable time interval between the holding of such a conference and the publication of its papers, which in this case appeared in 1960 [§4-Ed.]. In 1959, clearly unaware of my paper delivered in 1956, Lamb proposed exactly the same threefold division of the languages of North America. Swadesh, in a *Festschrift* for Paul Radin published in 1960, also advanced the same threefold classification. It is clear that Swadesh, like Lamb, was unaware of my paper. The idea was thus 'in the air' during this period of roughly four years 1956–1960. If one adds to this the opinions of Haas, Golla (who apparently endorses it in print: Golla 1987) and Kaufman (as just noted) the extent of agreement on this point becomes impressive.

Among Pan-Americanisms the first person singular pronoun *n* and the second person singular *m* have acquired a certain celebrity, both at the Conference* itself and in the scientific periodicals and the popular press. In the course of this discussion the basic facts have been to some extent obscured, yet they are essentially clear. Bright, emphatically not my supporter (e.g. Bright 1988), mentions them as his only example of Pan-Americanisms. Payne (1990) also, who seems to support the usual theory of Pan-Americanisms mentions them almost casually as instances. In *Language in the Americas* (1987: 48–55) it takes seven pages to enumerate

---

* The conference at which this paper was originally presented-Ed.

their occurrences in the Americas. They had previously been noticed by Trombetti, Sapir, and Swadesh. They have been independently reconstructed for groups such as Penutian and Uto-Aztecan in ignorance of their widespread occurrence elsewhere. When Michelson attacked Sapir for reconstructing these two forms for Algic, a grouping now universally accepted, he quoted a large number of instances in other North American Indian languages. These occurrences, according to him, helped disprove Algic because everyone knew that these other languages were unrelated.

I believe it was only after the full impact of this piece of evidence (and there are numerous others) had been realized as simply incompatible with the notion of 200 odd stocks that we begin to get the desperate attempts to find alternative nongenetic explanations. The borrowing of a first or second person pronoun is an utterly rare event; that it should have occurred numerous times over a vast area is simply ruled out. If relegated to an earlier period in Siberia, as suggested by Bright, one wonders why present-day Siberian languages which almost all have the pattern m, first person and t, second person and where n/m does not appear did not also borrow the Amerind pronouns, or vice versa, in some instances. One can only attribute this to some foreknowledge among these groups that certain ones would cross over to the Americas and others not.

If we exclude genetic and contact explanations then they have to be explained as accidents (clearly an untenable position) or as sound symbolism, some inherent tendency of languages to have nasals in first and second person pronouns. This attempt has indeed been made by Campbell who even at the Conference attributed such a preponderance of nasals to the phonetic nature of infant sucking reflexes!

Along with this has come, in some quarters, the rejection of the genetic explanation as such and a revival of the notion of 'mixed languages' along with the rejection of the family tree metaphor, a kind of analysis which I believe, with suitable modifications from wave-like phenomena in early periods of differentiation of a stock, is perfectly viable. I believe that the genetic question can always be answered.

An extended discussion of this question would require another paper of a scope comparable to the present one and is not attempted here. However, one should note that the doubts about Altaic in Doerfer's recent book have been influential here. His theory, nowhere explicitly stated in his book, seems to be that through numerous episodes of extended contacts, among varying languages and dating even from the period of the presumed protolanguages of Turkic, Mongolian and Tungusic, these language groups supposedly with

no demonstrable relationship to each other have come to resemble each other more and more.

This theory is, in a sense, the counterpart of Trubetskoy's concerning Indo-European, one universally rejected by Indo-Europeanists (for a discussion, see Thieme 1953), that the resemblances among Indo-European languages are to be explained from contacts among a series of originally distinct and unrelated languages. How these languages arose is not explained.

It is not commonly noted, but should be obvious, that such theories are the complete obverse of the wave theory first propounded by Johannes Schmidt in 1872. Here one starts with a hypothetical unity within which innovations arise randomly at certain points diffusing over varying distances. The ultimate result is a complete absence of well-defined breaks so that the family tree notion breaks down. Unfortunately, people will not hold still. There are migrations, besides which internal geographical and ethnopolitical boundaries tend to produce the isogloss bundles which become the foundation of dialect and ultimately of language differences. Schmidt himself treats only intermediate groupings within Indo-European e.g. whether there is an Italo-Celtic unity. He was aware that say, between Germanic and Slavic there were no intermediate languages. A language is unambiguously the one or the other. He sought to mitigate this by saying there were once intermediate dialects which died out. Since his time, numerous new Indo-European languages have been discovered but they turn out to be just as distinct as the branches discovered in the nineteenth century. The discussions in Leskien (1876) and Brugmann (1884) are still worth reading in their sensible reconciliation of wave with family tree theory.

As for Doerfer, he has developed a vastly elaborate set of hypotheses regarding contacts between Mongolian on the one hand and Turkic and Tungusic on the other, such that he explains almost everything. That there are extensive borrowings, for example, in Yakut (a Turkic language) and Manchu (a Tungusic language) from Mongolian, is well known and evident in the comparative work of earlier scholars like Poppe, Ramstedt, and Menges.

In his book he does not say explicitly that the Altaic languages are not related, but he is clearly hostile to the notion. In fact, there are a number of etymologies which he does not account for as well as various agreements in morphological markers. Nor does he mention the Sapir-like threefold agreement in an irregular alternation between the direct and oblique stems in the first person singular pronouns of Chuvash, Mongolian, and Tungusic. Since Chuvash constitutes a separate branch as against the rest of Turkic its testimony is of particular importance. In the main body of Turkic, the

difference has been levelled analogically. The forms are set forth in Table 1, in which the oblique stem is illustrated by the genitive. This by itself is enough to show that the Altaic languages are related.

TABLE 1

	Chuvash	Mongolian	Tungusic
Direct	epə	bi	bi
Genitive	man-ən, man	min-u	min-i

Several times in the course of the foregoing exposition I have quoted such expressions as 'demonstrate' or 'prove', always in quotes. Such expressions are common in the present-day literature regarding genetic classification of American Indian languages, but also in other parts of the world. These terms are only appropriate in mathematics or logic, whose propositions are true by definition. In the empirical sciences such expressions are out of place. Even in an advanced science such as physics, its practitioners are in general agreement that they are dealing with probabilities, although of an extremely high order. A theory that explains the relevant facts much better than any other is accepted and is strengthened by its fruitfulness in producing still further results.

In linguistics, there is frequently a search for something like absolute certainty which is simply out of place. In historical linguistics this shows itself in at least two ways. One is that relationship is shown with complete certainty by regularity of sound correspondences. I believe that this matter has already been dealt with adequately by me and by others. The second is that there is a completely certain way of deciding all etymologies. Reconstructed forms are viewed very much as premises in an argument and sound laws resemble principles of deduction. If the existing forms cannot be deduced in this manner the etymology is rejected. Examination of any comparative dictionary of Indo-European will show that even after almost two centuries, there are numerous unsolved problems. Besides all this, semantic change turns out to be virtually impossible to formalize. This led Guthrie to demand identity of meaning in etymologies and has led Goddard to demand almost as much.

Our problem in classification is a historical one and in historical problems evidence is weighed, not counted. The significance of any piece of evidence has three aspects, weight, independence, and relevance. Weight is not easily given quantitative expression but it is clear that certain kinds of resemblances are inherently more significant than others ('weightier'). Thus a resemblance involving a single consonant has less weight than one having two consonants, and one involving recurrent resemblances more than one that does not;

a resemblance that has an obvious sound symbolic element is less weighty than one that does not. Most weighty of all are agreements in alternations because they involve the simultaneous relations of two (or occasionally more) sounds with corresponding morphological categories. If the phonetic elements do not have a common phonological source but are arbitrary (i.e. suppletive), deriving from historically different morphemes, their weight is very great indeed. For example the agreement within Germanic among English. German and the Scandinavian languages seen in English *good: better: best* is extremely powerful evidence for a historical connection. That these agreements could have arisen accidentally is truly infinitesimal.

The second factor is independence. The agreement in a number of separate items, especially if some of them are of great weight, adds to, or more accurately, multiplies the probability of each item as not being accidental. I will illustrate independence of evidence with regard to writing systems partly to emphasize that historical inference in all fields is basically similar and to help demystify the linguistic case which has come to have an esoteric aspect which intimidates the nonspecialist. If we compare two alphabets like the Greek and the Phoenician, we may note three independent facets. One of these is that graphic symbols of a particular shape should have a certain phonetic value. A second is that these symbols should have a certain order in the alphabet. The third is that each symbol should have a particular name. Now that, just to take the first four letters: that two historically independent alphabets should have *a*, *b*, *g* and *d* as their values in the same order and that the names should be in Greek *alpha*, *beta*, *gamma*, *delta* corresponding to, using Hebrew in place of the closely related Phoenician language, *aleph*, *beth*, *gimmel*, and *daled* respectively, is utterly beyond chance.

The third factor is relevance and in historical classification this has to do with distribution across languages in relation to the distribution of other independent elements with a similar distribution, the existence of which is the basis for a hierarchical genetic classification. A particular resemblance, depending on its distribution across languages, may be relevant to different levels of classification.

The three factors of power, independence, and relevance can be illustrated from two linguistic examples, the discovery that Hittite was Indo-European, and the pronominal evidence for the relationship of the Algic languages with each other and their status as a valid genetic group within Amerind at some level.

It was noted earlier that it was Hrozný (1917) which proved decisive in the acceptance by Indo-Europeanists for the affiliation of Hittite. His work was based on the enormous fund of cuneiform tablets found in the course of

excavations initiated by Winckler at Boghaz-Köy in Turkey, the site of the Hittite capital. Approximately nine-tenths of these tablets are in Hittite written in a cuneiform syllabary which also contained ideographic signs. This form of writing was virtually identical to that of the long since deciphered Semitic language Akkadian.

However the story does not really begin with Hrozný's book or his preliminary article of 1915 based on the Boghaz-Köy tablets. Documents in the Hittite language were found earlier in the excavation of Ikhnaton's tomb at Tell Amarna in Egypt. A body of royal correspondence was found there in cuneiform script mostly in Akkadian, the international diplomatic language of the time. However, besides these there was a letter in Hurrian and two letters in Hittite consisting of correspondence with the land of Arzawa, a Hittite province. That these letters involved correspondence between the king of Egypt and the king of the Hittites was soon recognized and is not in dispute.

In 1902, Knudzton published these letters, together with extended comments on the language of the letters by two other Norwegian linguists, Bugge and Torp. The title of Knudzton's book on the Arzawa letters was 'The two Arzawa letters, the oldest sources in an Indo-European language'. On what did Knudzton base these conclusions, now known to be correct? The formulaic beginning of such letters between kings was well known and found in letters written in Akkadian. 'It is well with my houses, my wives, my children, my nobles, etc. May it be well with your houses, your wives, your children, your nobles, etc.' The words for house, wife etc. are all in ideograms in the first Arzawa letter and in the usual order. The cuneiform text runs as follows. 'Well houses-*mi*, wives-*mi*, etc., *es-tu* well houses-*ti*, wives-*ti*, etc.' The word for 'well' is spelled out in Akkadian. As pointed out by Torp (1902: 108), 'The conclusion that we have to do with an Indo-European language is obvious from Knudzton's discovery of *es-tu*, imperative third singular of the verb 'to be' and of *-mi* and *-ti* as enclitic possessives of the first and second person.' Regarding *es-tu*, it is to be analyzed as consisting of the verb stem *es-* 'to be' and *-tu*, imperative third person singular, all of this in exact agreement with Latin *esto* and Greek *esto* with the same meaning.

In addition, Knudzton had discovered an accusative singular *-n* < **-m*) as in Greek, a genitive singular *-as* (cf. *es/os* of the consonant stems in Proto-Indo-European) and *-n* of the first person singular of the past (cf. once more Greek *n* < *m*). Torp also conjectured a second person singular imperative marker *-k* as in Lithuanian. Bugge interpreted *du-qa* as the second person singular independent pronoun with the same strengthening particle as in Gothic *thu-k* and Greek *su-ge*.

All of the items mentioned in the previous paragraph are at present accepted and incorporated into grammars of Hittite. There were also, of course, a few erroneous interpretations. However, in the very year that Knudzton's volume appeared, Hirt, a leading Indo-Europeanist, vigorously attacked the thesis that Hittite was Indo-European. He admits that *estu* must mean 'may it be' and that -*mi* and -*ti* are correctly interpreted. However, the last two he maintains point rather to Hittite being a Finno-Ugric language; the first receives no explanation and he disregards the other inflections discovered by Knudzton and his fellow Norwegians. The leading Orientalist of the day, Winckler, lent the weight of his authority to Hirt's negative criticism without discussing the linguistic evidence but by the assertion that on non-linguistic grounds an Indo-European language could not have been spoken in Anatolia at that time. Weber, in his commentary on the Knudzton edition of the Tell Amarna tablets tells us that Knudzton himself under the weight of these apparently authoritative criticisms began to doubt his own thesis (Weber 1915: 1074). Ironically, 1915 was the very year in which Hrozný, on the basis of the vast new material from Boghaz-Köy published his preliminary paper on the Indo-European affiliation of Hittite. He gave due credit to Knudzton. Friedrich in his well-known work on the decipherment of scripts and discovery of new languages passes over Knudzton's brilliant discovery in presumably embarrassed silence and begins his account with Hrozný.

The significance of this episode in connection with the present discussion is in its illustration of the third principle of historical evidence, namely relevance. It was not just that Knudzton, Torp, and Bugge asserted correctly that Hittite was Indo-European and on the basis of fuller evidence turned out to be right. The three items asserted by Torp to be sufficient, *estu,* -*mi*, and -*ti* are indeed decisive. However they have different relevance. The pronominal elements -*mi* and -*ti* are found not only in Indo-European and Finno-Ugric but in other languages of Northern Asia which belong to the Eurasiatic family. Hirt's argument that these pronominal markers show that Hittite might be Finno-Ugric rather than Indo-European is at least partially reminiscent of Michelson's argument against Sapir that Sapir's hypothesis that *n* and *m* as first and second person pronouns in Algic is invalid because these forms are also found in a large number of other languages of North America. The form *estu*, however, has a different relevance, namely that it *is* diagnostic of Indo-European.

It is only by a broad approach to a large area that one can, by a genetic classification of the reasonably well-documented languages, discover what is diagnostic of valid genetic units at various levels. Given such background evidence, even a small amount of diagnostic evidence can be decisive. Torp's

statement was not only entirely correct, but it was no accident that this was so.

In a recent review of South American classification Kaufman says regarding Taiririu, a language for which only four words are recorded, that even Greenberg was not bold enough to classify it. The reason was not that there were only four words, but that they were not the *right* four words. In cases like this, in which a language can be classified on the basis of a few words, it is misleading to say that the evidence consists of only these few words. The evidence, often massive, from the language group to which it can be connected is also relevant. Hence new Indo-European languages have been securely identified on the basis of quantitatively slim but weighty, independent, and relevant evidence, as was seen earlier in the case of Lydian.

In addition to the factors just discussed regarding the weight of evidence in historical inquiries, the notion of fruitfulness was mentioned, though very much in passing. It was in fact illustrated in the history of the discovery that Hittite was Indo-European. This hypothesis was originally advanced, as we saw, by Knudzton on the basis of a very small amount of extremely cogent evidence and was later completely confirmed by a vastly greater amount of evidence. I shall call this property of correct hypotheses fruitfulness.

It can be specifically illustrated, in addition to the Hittite example just discussed, by the history of the Algic hypothesis, namely that Algonquian is related to two languages of California, Wiyot and Yurok. Put within a more general framework it is the hypothesis that these three form a genetically valid grouping at some level within Amerind. As is well known this thesis was first advanced by Sapir in 1915 and encountered vigorous rejection by the well-known Algonquian specialist Michelson. The effect of this attack was to make Sapir's thesis a highly controversial one in the general opinion of Americanists. In my class lectures of the period before 1952 I stated it in roughly these terms without having actually read Sapir's article or looked at the evidence myself. When I finally did, I was frankly astonished that it should be controversial. In the context of my African classification I would have placed the Algic languages at some relatively low subgrouping level. Therefore in Greenberg (1953) [§1-Ed.], which includes a worldwide review of classification problems, I stated that the close affinity of Algonquian to Wiyot and Yurok was obvious.

In an article in 1958 which is generally regarded as having ended the controversy, Haas put some emphasis on the notion that the highly speculative article by Sapir, based on the sketches of Wiyot and Yurok in Kroeber (1911) had received full confirmation as a result of the new data which had become available through the grammars of Wiyot by Reichard (1925) and of

Yurok by Robins (1958) and the as yet unpublished material on Wiyot by Teeter. She cited my opinion that the validity of the Algic hypothesis was obvious but was evidently puzzled by it. In my view the original article by Sapir was sufficient. It was not just the identity of the four singular pronominal prefixes *n/k/w/m* for the first, second, third, and impersonal in all three groups, cited a number of times subsequently by Goddard as decisive, but also the existence of two distinct plural suffixes *nan, for the first person and *waw for the second (these reconstructions are not intended to be exact), a highly unusual typological feature accompanied by phonetic agreement between Algonquian and Wiyot. There is also an agreement between the same languages in intercalating a -*t*- between the personal prefixes and the vowel of initial vocalic stems. Naturally, being correct, Sapir's thesis also turned out to be fruitful in that still further facts, described later, provided additional support for the original thesis.

The body of this paper has been devoted to the theoretical problems relating to the methodology of classification. Two of the papers presented at the Conference, however, those of Everett and Kaufman, professed to test subgrouping hypotheses regarding parts of Amerind and I propose to discuss both of these papers here. Everett suggests that if a vast computer program feeding data from all the languages of the Americas were undertaken, such a program might well come out with results quite like that of my threefold classification, but that these results would be of little interest to a field worker interested in a restricted group of languages. He believes that the real action, so to speak, is at the lower subgrouping levels. I consider the major groupings I have proposed for the Americas, namely Amerind, to be surer than any of the subgroupings though I believe that these in general are also well founded. A well-known example is Austronesian. It has never been doubted as a valid genetic unit, but even today [1990-Ed.] there are sharp differences among scholars regarding the subgrouping of the family.

Be that as it may, let us turn to Everett (1990). His paper falls into two parts. The first is concerned with the validity of the grouping I call Arawakan which I classify as a member of the Equatorial subgroup of Amerind. The second part is devoted to refuting my contention that Mura-Matanawi is a member of the Paezan subgroup of Chibchan-Paezan in my overall classification of Amerind. In both instances he comes to negative conclusions.

Under Arawakan I include five groups: (1) Maipuran (Arawakan proper); (2) Arawa; (3) Chapacura; (4) Uro-Chipaya; (5) Guamo. I express some reservations about Guamo, an extinct language known from only two mission vocabularies from rather similar but distinct dialects. Everett only asks whether Chapacura and Arawa, two groups on which he has done field work

are related. It is therefore a bilateral comparison which omits Maipuran, the most extensive language group in South America. He brings some new field data from some languages of the Arawa and Chapacura groups. The material is interesting but in his comparison he seems to disregard all previous work on the languages of these groups as well as on the three others I include. He notes that Kaufman brought to his attention the fundamental work on Chapacuran of Crequi-Montfort and Rivet (1913), but there is no indication that he used it. He is unaware that the gender marking pattern *u* masc./*i* fem. found in his Arawa material is identical with the common and diagnostic gender marking pattern of Maipuran (cf. Matteson, 1972: 127). Even disregarding this earlier work there is no reference to the Summer Institute of Linguistics volume containing an extensive chapter on Arawakan (Matteson 1972) in which Arawan is treated as merely a subgroup of Maipuran. This opinion was first voiced by Ehrenreich in 1897 and is heavily documented in Rivet and Tastevin (1938–9). I was actually more conservative than these writers in that, as is evident from my three notebooks on Arawakan, Arawan is rather coordinate with Maipuran in the Arawakan subgroup of Equatorial than itself a subgroup of Maipuran. Everett merely does a binary comparison of two of the five groups, omitting even the vast Maipuran stock, and only cites typological evidence, and concludes that Chapacuran is not related to Arawan because of 'fundamental differences in the grammar'. By the same line of reasoning one might well conclude that English is not related to Hindi. Even Campbell and Mithun, who cannot be accused of being prejudiced in my favor, recognize that I made a fundamental contribution by separating genetic criteria (sound and meaning simultaneously) from typological (sound only or meaning only), citing Greenberg (1963) [§5-Ed.] to this effect in their introduction to *The Languages of Native America* (Campbell and Mithun 1979: 51). Everett does not consider any of the numerous etymologies involving languages of the Arawakan group in my book.

Everett's treatment of the Paezan affinities of Mura-Matanawi suffers from even graver defects than does his discussion of Arawakan. According to him 19 of the 33 forms cited for Mura-Matanawi in my book are wrong. However, after checking my notebooks and then the original sources, I found that virtually all the forms I cited are found there. Let us consider the first fifteen. Of these 1, 2, 3, 10, 12, and 15 are correct according to Everett. Number 4 is Matanawi *torupi* 'shadow'. Here as in other instances he eliminates forms as incorrect with the observation that it is not a possible or likely Muran word. It is highly suspicious that he only cites specific forms from Pirahã, the dialect of Mura that he has worked on. There is just one published source for Matanawi that I am aware of, namely Nimuendaju (1923). It would seem that

Everett is guessing that a form cannot exist in Matanawi unless a similar form is found in Pirahã. This is as though someone who was an expert on Rumanian when confronted with a form in French which did not agree with Rumanian said that the French form was not a possible Romance word. Actually *torupi* is found in Nimuendaju (1923) as the fifth entry in the left column of page 169. He also says that the sixth entry *upi, opi* 'brother' is incorrect. It is found in the same source on p. 168, right column, line 14. The next item is my Mura citation *pe* 'cloud'. Everett says it means 'water' but in the dialect of Mura cited by Nimuendaju, p. 160, line 2 from the bottom, it occurs with the Portuguese gloss *nuvem* 'cloud'. According to Everett no. 8 *abe* 'let's go', cited by me as Mura is incorrect. However it occurs on page 165, column A, line 8 with the correct Portuguese gloss. The ninth item is Matanawi *manyo* 'deer', also said to be an error. It is found on page 169, line 10 from the bottom. The tenth entry is Matanawi *isi* which he says rather means 'juice'. It occurs with the Portuguese gloss 'drink!' (imperative) on page 171, line 7 from bottom. The eleventh item is Mura *piase* 'summer'. Again my citation is correct but doubtless Everett is correct in saying that this is a phrase meaning 'water is low' and in fact *yasi* 'dry' occurs in my notebook but I failed to make the analysis. Number 13 is an item not found in my book namely *kao* 'far'. He suggests it and then says it is wrong. I find it in my notebook and also rejected it. I should certainly get credit for refraining from a wrong etymology. Number 14 *jaa* 'feather' according to Everett is not a possible Muran word. It occurs on p. 169, column B, with the Portuguese gloss *penna* as a Matanawi word. We see then that in every instance I faithfully reproduced what was in my sources but that in one instance I failed to analyze a complex form. Further, *pe* 'cloud' is in all probability the same word as that for 'water' in another etymology. The word for 'water' often means 'rain, sky' and from this the transition to 'cloud' is easy.

It seems clear that Everett like some other Americanists ignores or rejects earlier sources. Is this justified? Historical linguistics is after all a historical discipline. What would we think of a historian who willfully ignored sources because they were imperfect? As historians know, all sources are imperfect. It is just that some are more imperfect than others. In the case of Nimuendaju we must give him credit for being the first, on the basis of his field researches, to assert that Mura and Matanawi are related (1923: 220). This relationship is accepted by Everett. Nimuendaju's material could not have been all that poor if it enabled him to reach this correct conclusion.

The basic assertions of Kaufman (1990) are that in classifying Subtiaba with Hokan in agreement with Sapir I erred in that Subtiaba-Tlapanec is rather an Oto-Manguean language and further that Oto-Manguean as a whole is

related to Hokan and that therefore there is no basis for my Central Amerind grouping which consists of Uto-Aztecan, Kiowa-Tanoan, and Oto-Mangue. Kaufman insists that protolanguages be used in comparisons. It was precisely in my Central group that I used protolanguages, Whorf and Trager (1937) for Azteco-Tanoan, Miller (1967) for Uto-Aztecan, and Rensch (1976) for Oto-Manguean. However, according to Kaufman Rensch showed in an unpublished manuscript of 1983, which was obviously unavailable to me, that the Oto-Manguean reconstructions were seriously flawed and since they are only valid randomly my comparisons of Oto-Manguean, Uto-Aztecan, and Tanoan rely on essentially chance resemblances for Oto-Manguean. These are very strong words. If I cannot rely on a reconstruction carried out as a published dissertation at a major University, Pennsylvania, in which the only indebtedness mentioned is to an eminent Indo-Europeanist, Hoenigswald, what am I to trust? In fact, as evident in the discussion in my book, I had reservations about some aspects of the reconstruction (1987: 123–4). Still, it could hardly be valid only randomly, as Kaufman asserts. However, the most startling observation is still to come. According to Kaufman, the same Rensch in 1977 showed Subtiaba-Tlapanec to be Oto-Manguean. I was well aware of and had read this study and was not impressed; indeed Kaufman admits that this proposal has not aroused enthusiasm. Naturally enough Rensch, in seeking to show that Subtiaba-Tlapanec is Oto-Manguean, uses the reconstructions from his published dissertation. In other words, these only randomly correct reconstructions when used by Rensch himself can prove that Subtiaba-Tlapanec is Oto-Manguean but when used by Greenberg in his section on the Central subgoup of Amerind they undermine the whole concept of a Central group of Amerind!

The work of Rensch on the Oto-Manguean affiliations of Subtiaba-Tlapanec was published with a twin study by Oltrogge in which Subtiaba was compared to two other Hokan languages of Central America, Jicaque and Tequistlateco. This study gives 130 etymologies either between Subtiaba and Jicaque or Subtiaba and Tequistlateco and reconstructs a protolanguage for the three of them. According to Oltrogge (1977: 38): 'That Jicaque is genetically related to both Subtiaba and Tequistlateco is clearly established by the reconstruction of phonetically reasonable proto systems based on recurring sound correspondences. . . .' Both Rensch and Oltrogge realize that they have arrived at an impasse. Using what they take to be standard methods, Subtiaba has been proven to be both Hokan and Oto-Mangue. The only way out that they can see is to suggest the hypothesis that Hokan and Oto-Mangue are related. It is this suggestion that Kaufman adopts by assigning Subtiaba to Oto-Mangue in agreement with Rensch, while

explaining the resemblances to Hokan by the more distant relationship between the two large groups. I of course agree that Hokan is related to Oto-Mangue in that both are Amerind and Kaufman admits that he has not yet investigated the possible connection of Oto-Mangue with Uto-Aztecan and Tanoan.

It is clear that by asserting the relationship of Hokan and Oto-Mangue Kaufman has taken a major step towards Amerind. To judge from the introduction to Campbell and Mithun these two scholars do not even consider Hokan to be an established grouping. There is also a very considerable discrepancy between Kaufman's conclusions about Hokan and Oto-Manguean and his recent discussion of the linguistic situation in South America for which he posits 118 stocks although with some suggestions regarding minor consolidations. The reason for this discrepancy is obvious. In Central and North America his work is parasitic on that of presently unfashionable figures such as Kroeber, Dixon, and Sapir, while there were no such predecessors in South America.

From the foregoing review it should be evident that the notion that I have abandoned the comparative method in favor of some new and strange invention of my own is without foundation. When I do comparative linguistics, as I have done since my 1948 article on the tonal system of Proto-Bantu (Greenberg 1948), I do it just like anyone else. My criticisms are not directed at the comparative method as such. There is no other way of doing comparative linguistics. All the points I have made are to be found in the writings of comparativists of impeccable credentials. What I would like to emphasize is that the picture presented by Campbell *et al.* and repeated in the popular press that historical linguistics has an utterly rigorous method, however slow, which reconstructs linguistic history step by step with complete precision is sheer myth, as a look at any etymological dictionary or comparative grammar of Indo-European or some other major stock will show. As it is, the achievements of linguistics, including that of its historical branch are indeed impressive and doubtless superior to that attained in the study of any other human cultural activity. There is no need to distort its results so that they appear to have a logical precision which in the nature of things they cannot possibly have. Even so, as we have seen, they can only be applied to a stock whose membership has already been ascertained. That the force of this obvious consideration has begun to be appreciated can be seen in the statements of Thomason (1990) that it is a good way to generate hypotheses, and of Kaufmann (1990: 24) and Liedtke (1989) that it should be combined with bilateral comparisons. Such bilateral comparisons have no purpose, except where, on the basis of a wider comparison we can see that

there are branches with a special relationship, e.g. Balto-Slavic (in Indo-European). With roughly fifteen branches there would be 105 two-way comparisons. Needless to say this absurd procedure has never been applied in Indo-European studies.

References

Anttila, Raimo. 1989. *Historical and Comparative Linguistics* (2nd rev. edn.). Amsterdam: Benjamins.

Bateman, Richard, Goddard, Ives, O'Grady, Richard, Funk, V. A., Mooi, Rich, Kress, W. John, and Cannell, Peter. 1990. 'Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language', *Current Anthropology* 31: 1–24, 177–83.

Beeler, Madison. 1949. *The Venetic Language.* Berkeley: University of California Press.

Bloomfield, Leonard. 1933. *Language.* New York: Holt.

Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, germanischen, persischen und germanischen Sprachen.* Frankfurt: Andreäische Buchhandlung.

Bright, William. 1984. *American Indian Linguistics and Literature.* Berlin: Mouton.

——1988. 'Review of Greenberg (1987)', *American Reference Book Annual* 88: 440.

Brown, Cecil and Witkowski, Stanley. 1979. 'Aspects of the phonological history of Mayan-Zoquean', *International Journal of American Linguistics* 45: 35–47.

Brugmann, Carl. 1884. 'Zur Frage nach den Verwandschaftsverhältnissen der Indo-Germanischen Sprachen', *Internationale Zeitschrift fuer allgemeine Sprachwissenschaft* I: 226–56.

Bugge, Sophus. 1902. 'Bemerkungen zu den Arzawa-Briefen', in Knudzton 1902, 57–107.

Campbell, Lyle and Kaufman, Terrence. 1980. 'On Mesoamerican linguistics', *American Anthropologist* 82: 850–7.

Campbell, Lyle and Mithun, Marianne (eds.). 1979. *The Languages of Native America.* Austin: University of Texas Press.

Collinder, Bjorn. 1960. *Comparative Grammar of the Uralic Languages.* Stockholm: Almqvist and Wiksell.

Collinge, N. E. 1985. *The Laws of Indo-European.* Amsterdam: Benjamins.

Crequi-Montfort, G. de and Rivet, Paul. 1913. 'Linguistique bolivienne. La famille linguistique Chapacura', *Journal de la Société des Américanistes de Paris* 10: 119–71.

Darwin, Charles. 1871. *The Descent of Man, and Selection in Relation to Sex* (1st edn.). London: J. Murray.

Delbrück, Berthold. 1904. *Einleitung in das Studium der indogermanischer Sprachen* (4th and renamed edn. of *Einleitung in das Sprachstudium*, 1880.) Leipzig: Breitkopf & Härtel.

Diderichsen, Paul. 1974. 'The foundations of comparative linguists: revolution or continuation?' in Dell Hymes (ed.), *Studies in the History of Linguistics*. Bloomington: University of Indiana Press, 277–306.

Doerfer, Gerhard. 1985. *Mongolo-Tungusica*. Wiesbaden: Harrassowitz.

Ehrenreich, Paul. 1897. 'Materialien zur Sprachenkunde Brasiliens III. Vokabeln von Puru-Stämmen', *Zeitschrift für Ethnologie* 29: 59–71.

Everett, Daniel L. 1990. 'A structural comparison of Arawan and Chapacuran: any evidence for a genetic classification?' Paper presented at the Conference on Language and Prehistory in the Americas, Boulder, Colorado, 1990.

Friedrich, Johannes. 1957. *Extinct Languages*. New York: Philosophical Library.

Gamkrelidze, Tomas and V. Ivanov, Vyacheslav. 1984. *Indoevropejskij jazyk i indoevropejtsy*. Tbilisi: Izdatelstvo Tbilisskogo Universiteta.

Goddard, Ives. 1975. 'Algonkian, Wiyot and Yurok: proving a distant relationship', in M. Dale Kinkade, Kenneth L. Hale, and Oswald Werner (eds.), *Linguistics and Anthropology*. Lisse: Peter de Ridder, 249–62.

—— 1986. 'Sapir's comparative method', in William Cowan, Michael K. Foster, and E. F. Konrad Koerner (eds.), *New Perspectives in Language, Culture and Personality*. Amsterdam: John Benjamins, 191–210.

Golla, Victor. 1987. 'Comments in *Current Anthropology* book review of Greenberg (1987)', *Current Anthropology* 28: 657–9.

Greenberg, Joseph H. 1948. 'The tonal system of Proto-Bantu', *Word* 4: 196–208.

—— 1987. *Language in the Americas*. Stanford: Stanford University Press.

Grimm, Jacob. 1822. *Deutsche Grammatik* (2nd edn.), vol. 1. Göttingen: Dieterichsche Buchhandlung.

Haas, Mary. 1958. 'Algonquian-Ritwan: the end of a controversy', *International Journal of American Linguistics* 24: 54–73.

Hirt, Hermann. 1927. *Indogermanische Grammatik, vol. 4: Der indogermanische Vokalismus*. Heidelberg: Winter.

Hrozný, Friedrich. 1915. 'Die Lösung des hethitischen Problems', *Mittheilungen der Deutschen Orient-Gesellschaft* 56: 17–50.

—— 1917. *Die Sprache der Hethiter*. Leipzig: Hinrichs.

Hübschmann, Heinrich. 1875. 'Über die Stellung des armenischen im Kreise der indogermanischen Sprachen', *Zeitschrift für Vergleichende Sprachforschung* 23: 5–42.

—— 1897. *Armenische Grammatik*. Leipzig: Breitkopf & Härtel.

Jankowsky, Kurt R. 1972. *The Neogrammarians*. The Hague: Mouton.

Kaufman, Terrence. 1990. 'Language history in South America: what we know and how to know more', in Doris Payne (ed.), *Amazonian Languages*. Austin: University of Texas Press, 13–74.

—— 1990. 'Tlapaneko-Subtiaba, OtoMangue, and Hoka: where Greenberg went wrong'. Paper presented at the Conference on Language and Prehistory in the Americas, Boulder, Colorado, 1990.

Knudzton, J. A. 1902. *Die zwei Arzawa-Briefe, die älteste Urkunden in Indogermanischer Sprache*. Leipzig: Hinrichs.

Knudzton, J. A. (ed.) 1908–15. *Die el-Amarna-Tafeln, mit Einleitung und Erläuterungen*, 2 vols. Anmerkungen und Register bearb. von Otto Weber und Erich Ebelung. Leipzig: Hinrichs.

Krauss, Michael E. 1979. 'Na-Dene and Eskimo-Aleut', in Lyle Campbell and Marianne Mithun (eds.), *The Languages of Native America*. Austin: University of Texas Press, 803–901.

Kroeber, Alfred L. 1911. *The Languages of the Coast of California North of San Francisco*, University of California Publications in American Archaeology and Ethnology, 9.

Lamb, Sydney M. 1959. 'Some proposals for linguistic taxonomy', *Anthropological Linguistics* 1/2: 33–49.

Leskien, August. 1876. *Die Deklination im Slavisch-Litauischen und Germanischen*. Leipzig: S. Hirzel.

Levine, Robert D. 1979. 'Haida and Na-Dene: a new look at the evidence', *International Journal of American Linguistics* 45: 157–70.

Liedtke, Stefan. 1989. 'Review of Greenberg (1987)', *Anthropos* 84: 283–5.

Matteson, Esther. 1972. 'Proto Arawakan', in Esther Matteson *et al., Comparative Studies in Amerindian Languages*. (Janua Linguarum, series practica vol. 127.) The Hague: Mouton, 160–242.

Meillet, Antoine. 1936. *Esquisse d'une grammaire comparée de l'armenien classique* (2nd rev. edn.). Vienna: Imprimerie des PP. Mekhitaristes.

Miller, Wick. 1967. *Uto-Aztecan Cognate Sets*. (Janua Linguarum, vol. 48.) The Hague: Mouton.

Newman, Paul. 1970. 'Historical sound laws in Dera (Kanakuru)', *Journal of African Languages* 7: 39–51.

——1995. *On Being Right: Greenberg's African Linguistic Classification and the Methodological Principles which Underlie it*. Bloomington, Indiana: Institute for the Study of Nigerian Languages and Cultures, African Studies Program, Indiana University.

Nimuendaju, Kurt. 1923. 'Documents sur quelques langues peu connues de l'Amazone', *Journal de la Société des Américanistse de Paris* n.s. 15: 215–22.

Oltrogge, David. 1977. 'Proto Jicaque-Subtiaba-Tequistlateco: a comparative reconstruction', in David Oltrogge and Calvin R. Rensch (eds.), *Two Studies in Middle American Comparative Linguistics*. (Summer Institute of Linguistics, Publication no. 55.) Arlington, Texas: Summer Institute of Linguistics, 1–52.

Payne, David L. 1990. 'Some widespread grammatical forms in South American languages', in Doris L. Payne. (ed.), *Amazonian Languages*. Austin: University of Texas Press, 75–88.

Pott, August Friedrich. 1833–6. *Etymologische Forschungen auf dem Gebiete der indogermanischen Sprachen*, 2 vols. Lemgo: Meyer.

——1859–76. *Etymologische Forschungen auf dem Gebiete der indogermanischen Sprachen*, 6 vols. Lemgo: Detmold, Meyer.

Rask, Rasmus K. 1834. *Samlede tildels forhen utrykte afhandlinger*, vol. 1. Copenhagen: Popp.

Reichard, Gladys. 1925. 'A Wiyot grammar and texts', *University of California Publications in American Archaeology and Ethnology* 22: 1–125.

Rensch, Calvin R. 1976. *Comparative Otomanguean Phonology*. Bloomington: University of Indiana Press.

——1977. 'Classification of the Oto-Manguean languages and the position of Tlapanec', in David Oltrogge and Calvin R. Rensch (eds.), *Two Studies in Middle American Comparative Linguistics*. (Summer Institute of Linguistics, Publication no. 55.) Arlington, Texas: Summer Institute of Linguistics, 53–108.

Rivet, Paul and Tastevin, Constant. 1938–9. 'Les langues Arawak du Purus et du Jurua (groupe Araua)', *Journal de la Société des Américanistes de Paris* 30: 71–114, 235–88; 31: 223–48; 32: 1–55.

Robins, R. H. 1958. *The Yurok Language: Grammar, Texts and Lexicon*. Berkeley and Los Angeles: University of California Press.

Sapir, Edward. 1913. 'Wiyot and Yurok, Algonkin languages of California', *American Anthropologist* 15: 617–46.

——1915. 'The Na-Dene languuges: a preliminary report', *American Anthropologist* 17: 534–58.

——1925. 'The Hokan affinity of Subtiaba in Nicaragua', *American Anthropologist* 27: 402–35.

Starostin, Sergei. 1989. 'Nostratic and Sino-Caucasian', in Vitaly Shevoroshkin (ed.), *Explorations in Macrofamilies*, Bochum: Brockmeyer, 42–6.

Swadesh, Morris. 1960. 'On interhemisphere linguistic connections', in Stanley Diamond (ed.), *Culture in History: Essays in Honor of Paul Radin*. New York: Columbia University Press, 894–924.

Thieme, Paul. 1953. 'Die Heimat der indogermanische Grundsprache', *Akademie der Wissenschaften und der Literatur. Abhandlungen* no. 11. Wiesbaden: Steiner.

Thomason, Sarah G. 1990. 'Hypothesis generation vs. hypothesis testing: a comparison of Greenberg's classifications in Africa and in the Americas'. Paper presented at the Conference on Language and Prehistory in the Americas, Boulder, Colorado, 1990.

Torp, Alf. 1902. 'Bemerkungen zu den Arzawa-Briefen', in Knudzton 1902, 108–24.

Weber, Otto. 1915. 'Anmerkungen', in Knudzton (vol. 2) 1915: 1009–1357.

Whorf, Benjamin and Trager, George L. 1937. 'The relationship of Uto-Aztecan and Tanoan', *American Anthropologist* 39: 609–24.

Wilbur, Terence H. (ed.). 1977. *The Lautgesetz Controversy: A Documentation (1885–1886)*. Amsterdam: John Benjamins.

Winter, Werner. 1962. 'Problems of Armenian phonology: III', *Language* 38: 254–62.

This page intentionally left blank

Part III
Indo-Pacific, Amerind, Eurasiatic

This page intentionally left blank

12

The Indo-Pacific hypothesis*

The evidence presented here is intended to demonstrate that the bulk of non-Austronesian [non-AN] languages of Oceania from the Andaman Islands on the west in the Bay of Bengal to Tasmania in the southeast forms a single group of genetically related languages for which the name Indo-Pacific is proposed.[1] The major exception to this generalization is constituted by the indigenous languages of Australia nearly all of which are generally accepted as related to each other. This Australian stock does not seem to align genetically with Indo-Pacific and is excluded from it, at least tentatively. With regard to the remaining languages, membership in the overall family can be asserted with varying degrees of confidence, as will be indicated at appropriate points in the discussion. This uncertainty stems largely from the significant gaps in the linguistic documentation, since as has been pointed by Grace (1968), the study of non-Austronesian languages has, at least until very recently, lagged far behind that of the Austronesian languages of the same area.

In 1958 the present writer prepared an unpublished report on the classification of the non-Austronesian languages of the Pacific for the Wenner-Gren Foundation (Greenberg 1958). This was a preliminary attempt to set forth language groupings viewed at least tentatively as independent. It was chiefly concerned with New Guinea where the largest number of non-AN languages outside of Australia are to be found and which exhibits by far the greatest linguistic diversity. Nine major groups in New Guinea and an additional five on other islands were delineated.

In 1960, in a paper presented to the Association for Asian Studies at its New York meeting, evidence was presented chiefly in the form of etymologies (Greenberg 1960) to show that all of the languages previously described as

independent groupings formed a single family in which the previously defined groupings entered as subgroups. In the present paper these subgroupings, with certain significant modifications, are retained. Evidence is presented for at least some of these subgroupings as well as for the validity of Indo-Pacific as a whole.

Since 1960 there has been a rather considerable increase in the amount and quality of research on the non-AN languages. In particular far more grammatical information is now available so that the nonlexical evidence, at that time virtually confined to the pronouns, can now be significantly expanded. In this period also, particularly for New Guinea, the notion of Papuan as merely a name for genetically diverse languages with only the negative characteristic of not being Austronesian, has largely been superseded by the notion that a very large proportion of these languages are indeed related. Further, as noted by Grace (1968) the subgroupings suggested by the present writer in 1958 and 1960, except for Cowan's West Papuan Phylum, show 'little disagreement with those reported by others'. The question of Cowan's West Papuan Phylum will be discussed later. It should be noted that Cowan deserves signal credit for the first serious attempts to relate New Guinea non-AN languages to those outside of the island.

Since 1960, I have assembled virtually exhaustive vocabulary information from published sources and a few unpublished ones enumerated in footnote 1 and in the second section of the bibliography. These were entered in twelve notebooks each of which contains places for upwards of 350 entries, many of which however are obviously not available in the existing sources. Each notebook contains entries from sixty up to eighty language sources and is organized roughly by groupings of languages whose closer relationship to each other is evident on inspection. In addition detailed notes on grammar were copied in three additional notebooks and reclassified by historically significant rubrics in still another notebook.

In order to avoid unnecessary duplication in the bibliography accompanying this article, Klieneberger (1957) is taken as basic and all sources found in Klieneberger are cited here with the letter K followed by the numbering in that work. Hence the bibliography consists entirely of items either overlooked by Klieneberger, stemming from geographical areas not covered in his bibliography, or published subsequently to it. Unpublished sources are cited by identifying letter and are enumerated in the second section of the accompanying bibliography.

Languages are usually referred to here by the name as it appears in the original source. It is realized that different sources often contain data under divergent names for what are merely dialect variations of the same language

or even the same dialect. Hence this exposition will give a misleading picture of the amount of actual linguistic variation within Indo-Pacific. However this seemed preferable to an attempt to distinguish language from dialect by the use of standard language names. This does not seem for the most part warranted on the basis of our present knowledge. The present procedure also has the advantage that assertions about languages can be immediately traced to the linguistic source which furnishes the information upon which such statements are based.

Even before consideration of the specific language groupings, a word should be said about the problem of detecting and excluding from the evidence cited those forms which are borrowings from AN languages. The existence of two language families in contact over such a vast area is indeed a unique situation. There is also the special influence of Malay as a widespread lingua franca. As a result of this, distant non-AN languages can show startling similarities resulting from independent borrowings from Malay or locally contiguous AN languages. Hence all etymologies were checked with Dempwolff's Proto-AN reconstructions and with Malay. This is, however, not sufficient since there are many locally distributed etymons in AN languages not included in Dempwolff. Hence vocabularies of approximately fifty AN languages similar in scope to those for non-AN languages and concentrating on those languages found in proximity to Indo-Pacific languages were also compiled. This is clearly not adequate and I cannot guarantee that AN loans have been entirely excluded from the present study. Of course, forms of limited distribution within AN as a whole are sometimes borrowings from non-AN languages. What is needed is a series of intensive studies of individual situations. To have done this would have required an effort at least as great as the one involved in the present study as a whole. Hence efforts in this direction were limited in the fashion just described.

Not all such mutual borrowings stem from AN in general or Malay in particular. For example, in the first unrevised set of Indo-Pacific etymologies there was a striking set of resemblances in a word meaning 'axe' or 'knife' involving three consonants and in languages geographically distant from each other. Thus, two geographically distant languages in New Guinea, Demta in West Irian and Sungumana in the Madang area of Australian New Guinea, have *sapere* 'knife' and *sapor* 'axe' respectively. However, Alor *sapada* 'knife' gives away the story, namely that we are dealing in all probability with loans from Portuguese *espada*.

In what follows, I will first consider languages outside of New Guinea in an approximate west to east order and then those of New Guinea itself. Evidence both lexical and grammatical will be adduced in support of a number of the

broader subgroups proposed. This subgrouping is not exhaustive and is in some respects at least quite tentative as will be indicated where appropriate. Such a degree of uncertainty is only reasonable at this stage. After all, in AN studies which are incomparably further advanced there is still no overall subgrouping which has met with general acceptance. After the discussion of individual subgroupings the case for the validity of Indo-Pacific as a whole, both grammatical and lexical, will be presented.

The westernmost group of languages to be considered is that of the Andaman Islands. In an earlier review of classification problems (Greenberg 1953) [§1-Ed.], it was noted that, unlike the languages of the neighboring Nicobar Islands, those of the Andamans did not affiliate with Austroasiatic or indeed with any other language family of the mainland of southern Asia. They were therefore assigned independent status.

Within the Andaman Islands are two separate groups of languages. The first of these covers almost all of Great Andaman, that is, the North, Middle and all except the southern extremity of South Andaman. It is relatively well documented in Portman (1887 and 1898) and other sources. It consists of a northern and a southern dialect cluster. Both the distinction between the two and their close relation are obvious and have been noted by previous observers. The sources of citation for these languages in the etymologies at the end of this article are Portman 1887 which gives data on Biada and Bogijiab of the northern and Kede and Chariar of the southern group, and Portman 1898 which includes Beada and Bale of the northern and Puchikwar, Juwoi, and Kol of the southern. Of these names Beada and Biada are mere nomenclatural variants.

On the southern tip of South Andaman in Great Andaman and on Little Andaman to the south, Jarawa and Önge are spoken respectively. These are merely dialectally different. They differ profoundly from the main group of languages covering most of Great Andaman. They are very meagerly documented in the Önge of Portman (1887). A. R. Brown (1914) gives some examples from a work of Bonig published in 1903 in Fort Blair which was not obtainable. He also shows that the early vocabulary of Colebrook (1795) refers to this language. As pointed out by Brown there are very few vocabulary resemblances between this language and those of Great Andaman and the only real point of contact is typological, the existence of a remarkable system of nominal classification based on parts of the body. A few citations from Önge have been included in the general Indo-Pacific vocabulary, but both its special relationship to the languages of the rest of the Andamans and its assignment to Indo-Pacific must be considered highly provisional.

The next group to be considered consists of certain languages on the islands of Timor and Alor and the small island of Kissar near the eastern extremity of Timor.

At least two non-AN languages are spoken on Portuguese Timor and were first described by Capell (K.359). These are Makasai on the eastern end of Timor and Bunak in the central portion. Berthe (1959) also presents data on Bunak under the name Buna. The ethnographer Josselin de Jong (1937) describes the language of Oirata known to be a Timorese settlement on the nearby small island of Kissar. Another presumably AN language is also spoken on Kissar. Oirata turns out to be non-AN and quite similar to Makasai. Cowan (1963 and 1965*a*) has advanced the thesis that these three languages, Bunak, Makasai, and Oirata, form a group within his larger West Papuan Phylum which also includes the languages of Northern Halmahera and a large group in the western portion of West Irian.

It has evidently not been noted up to now that at least one language of Alor, an island not far distant from Timor, is non-AN. This is Abui for which the work of two ethnographers, Nicolspeyer (n. d.) and Du Bois (1944), provide linguistic data. The former source gives fairly extensive interlinear texts in Dutch and Abui while the latter gives the Abui equivalents of a list of stimulus words for psychological tests as well as other scattered vocabulary. This language is given as AN in the well-known language atlas of Salzner (1960).

The internal subgrouping of the Timor-Alor group is clearly as follows:

1. Abui
2. A. Bunak
   B. 1. Makasai
       2. Oirata

The following are grammatical formatives common to two or more Timor-Alor languages. As with other subgroups and for lexical as well as grammatical morphemes, some items of broader distribution within Indo-Pacific as a whole have been included along with discussion of that fact. The reason for doing this is twofold. We wish the forms cited to be a first approximation to the common genetic linguistic inheritance of the subgroup and this will obviously include such presumably older elements. In addition, the specific variant of the more widespread element may show common peculiarities and thus have a positive bearing on the establishment of the subgroup. Thus, in Indo-European, English *four*, German *vier*, Italian *quattro* and Spanish *cuatro* are all cognate, but because of a series of shared sound changes and other common

peculiarities such forms have an obvious relevance to the distinctness of Germanic and Italic as subgroups of Indo-European.

Turning now to the grammatical data for the Timor-Alor languages we find the following:[2]

1. First person singular pronoun: Oirata *an, an*(*ri*), Makasai *ani*, Bunak *ne*(*to*), *n-* (possessive), Abui *ne*(*do*), *n-* (possessive).

All the languages have forms in *n*. This however is the most widespread type of first singular pronoun in Indo-Pacific in general. Bunak and Abui have a suffix *-to* and *-do* respectively added to other pronouns in their independent form.

2. Second person singular pronoun: Oirata *e:*(*ri*), Makasai *ai*, Bunak *e*(*to*), Abui *e*(*do*). Here again the resemblances in all four languages are clear. As pointed out by Cowan (1965*a*), such second person singular pronominal forms in *e* are found in certain languages of his West Papuan Phylum in West Irian. Otherwise they are rare or nonexistent in Indo-Pacific.

3. Bunak and Makasai share a third person pronominal element in *g-*. Makasai has *gi* as its third person singular pronoun. In Bunak as described by Capell there is a third person singular possessive pronoun *gia* and a plural *gi*. In Berthe's account the possessive and pronominal object prefix for the third person for both numbers is *g-*.

4. As first person plural exclusive pronouns Makasai has *ini*, Oirata *in*(*te*), *ina, in-* 'possessive', and Bunak *nei*. For Abui, Nicolspeyer gives two first person plural pronominal elements *ni-* and *pi-* without any rule for distinguishing their use. However it is highly likely that *ni-* is really exclusive and *pi-* inclusive as indicated by the close agreement of both with the pronouns of the other Timor-Alor languages.

5. In the first person inclusive plural we find Oirata *ap*(*te*), *ap-* 'possessive', Makasai *fi*, and Abui *pi-*.

A first person plural pronoun *ni*, usually without the inclusive/exclusive distinction is extremely widespread in Indo-Pacific. The languages of

---

[2] Ideally perhaps one would cite all forms in the orthography of the original source but in that case an additional apparatus of interpretation would have to be added since, e.g. *j* stands for a semivowel in Dutch and German sources but usually for a palato-alveolar affricate in English and American sources. I have therefore normalized the transcriptions. In symbolization I have sought to avoid diacritics since these are a constant source of error in the final version. I have generally used digraphs where at all possible. When the two symbols are to be taken in their separate values a dash intervenes. For example, *ng* is [ŋ] but *n-g* is [ŋg]. The following details should be noted. The glottal stop is indicated by *q*, a palatal semivowel by *y*, and voiceless and voiced palato-alveolar affricated by *c* and *j* respectively. The velar lateral which occurs in some New Guinea languages is symbolized by *kl* or *gl*, voiceless and voiced respectively. A bracketed nasal before a voiced stop indicates automatic (non-phonemic) prenasalization.

northern Halmahera agree with Timor-Alor in having a *p-* pronoun for first person plural inclusive. However only the Timor-Alor languages have the pattern *n-* exclusive/*p-* inclusive.

6. For the second person plural pronoun there is agreement among all the languages except Abui. The forms are Oirata *i:(te)*, *i:*, Makasai *i*, Bunak *ei*. Forms resembling these are rare and sporadic elsewhere so that we again have a characteristically Timor-Alor pronominal morpheme.

7. Abui *do* 'this, he, she, they' is cognate with Bunak *doe* 'this'.

8. Both these languages have an identical formation involving the foregoing element plus a locative suffix: Abui *do-n*, Bunak *doe-ni* 'here'.

9. Bunak and Abui share another demonstrative element *h* seen in the Bunak *himo* third person singular animate pronoun, *homo* third person singular inanimate and demonstrative, and *halaqi* 'those, they'. The same morpheme probably occurs as *h-* prefixed to verb to indicate an inanimate object. This is probably cognate with Abui *ha, he, ho* largely in free variation in Nicolspeyer's material. All these forms are described as third person singular possessive prefixes and as indicating the object of a verb. The latter usage is very close to that of Bunak *h-*.

10. Oirata *ita-*, Makasai *ta-* reciprocal verb prefix, and Bunak *t-* reciprocal verb and possessive prefix are probably related. This formation probably exists in Abui also as can be deduced from the pair *aruda* 'to waken' *taruda* 'to awake'.

Lexical evidence for the validity of the Timor-Alor grouping is set forth in the following set of etymologies. Oirata, Makasai, Bunak, and Abui are cited in that order with abbreviations O, M, B, and A respectively. Widespread etyma also found in the general Indo-Pacific list at the end of this paper are preceded by an asterisk. For matters pertaining to transcription and related problems see the accompanying footnote 2.

1. to ask O. *uste*, M. *seti*
2. to bathe B. *wer*, A. *welang* 'to wash'. Compare M. *bani* 'wash'.
3. behind O. *ura, uraa*, A. *werka*
4. belly O. *atu*, M. *atu(bere)*. *bere* means 'large' in Makasai. For its use as a suffix cf. 'star'.
5. bird O. *asa* 'fowl', M. *asa*. Compare B. *tie*.
6. to bite O. *tia* 'bite, eat', M. *tio*
7. bitter O. *malare* 'angry, sour, hot, bitter', A. *malohi*
8. brother (older) O. *ka, kaka* 'brother', M. *kaka* 'older brother, older male cousin', B. *kaa* 'older brother'
9. child O. *modo*, M. *mata*

10. cold B. *hurat*, A. *palata*
*11. come O. *maqu*, M. *maqu*, B. *man*, A. *me, mei, mirani*
12. to cover B. *pilik*, A. *banuike*
*13. to cry O. *oore*, M. *iar*, B. *ɔl*
14. to cut O. *utu*, A. *uti*
15. day O. *wadu*, M. *watu*, A. *wari* 'sun, day'. Compare B. *hɔt.*
*16. to descend O. *ipa* 'fall down or off', B. *pai* 'arrive', A. *pe, pei* 'descend, go' (cf. Indo-Pacific 'to fall')
17. to die O. *umu*, M. *umu*, A. *amui(mana)* 'the dead'. *mana* is a plural suffix.
18. to dig M. *toi*, A. *tehi*
19. dry O. *tata*, A. *tir*
20. earth₁ O. *miki* 'mud' B. *mɔk* (Capell), *muk* (Berthe), A. *mok*
21. earth₂ O. *no*, A. *anai*
*22. to eat O. *una, nawa*, M. *nawa*, A. *nei*
23. excrement O. *atu*, A. *asi*
*24. eye O. *ina*, M. *nana* 'eye', *ena* 'to see' (cf. Indo-Pacific 'to see')
25. face O. *panu*, A. *pong*
26. to fall O. *tapa*, B. *tɔp*
27. to fear O. *mane*, A. *mielang* 'afraid'
*28. female O. *panar* 'female of animal', B. *fana* (Capell) 'wife', *pana* (Berthe) 'woman'
*29. fire O. *ada*, M. *ata*, B. *ɔtɔ*, A. *ara*
30. fish O. *ahi*, M. *afi*, B. *afu*
31. flower B. *(ɔtel) (g)ubuk* i.e. (tree its) flower, A. *biek*
32. fly (n.) O. *uhur*, A. *fulfai*
33. foot M. *iti*, B. *(g)idi* (Capell), *(g)iri* '(his) foot', A. *iti* 'thigh'
34. fruit M. *isu*, A. *iti*
35. garden M. *ama*, B. *maa*
*36. to give O. *ina*, M. *(g)ini* 'to give (it)'
37. to go O. *mara*, B. *mal* 'to go, to go out'
38. good O. *roo*, M. *rau*, B. *lɔi*
*39. hair O. *wata*, M. *asa*, B. *ada*
*40. hand O. *tana*, M. *tana*, B. *dɔn*, A. *tang*
41. to hear O. *ware, wali* 'ear', M. *wali*, B. *fali, faling*
*42. to hit O. *wewese*, M. *basi*, B. *pas*
43. to hold M. *nai*, B. *(h)one* 'to hold (it)'
44. hole O. *uru, uruh* 'cave, hole', B. *uul* 'to open', A. *ul* 'gate, hole'
45. inside₁ B. *mil* 'interior', A. *min*
46. inside₂ O. *mudu*, M. *-mutu* 'in'

47. to kill O. *uda*, M. (*g*)*uta* 'to kill (him)'
48. to know B. *tada* (Capell), *tara* (Berthe), A. *tire* 'investigate, look at'
49. leaf O. *asah*, M. *asa*
*50. long O. *lokide* 'high', B. *legul*, A. *lohu, lougda*
51. to lose O. *mule* 'disappear, go astray', M. *mɔlu*
52. to make M. (*g*)*ini* 'to make (it)', B. (*h*)*ini* 'to make (it)'
*53. man₁ O. *nami* 'husband', A. *neng* (Du Bois 'man', Nicolspeyer 'male')
54. man₂ M. *anu*, B. *ɛn*
55. moon O. *uru*, M. *uru*, B. *hul*
56. mouse O. *ura*, A. *rui* 'rat'
*57. name O. *neene*, M. *nai*, A. *hanin*
58. neck O. *mani*, M. *manikai*
*59. nose O. *munikain*, A. *min*
60. old₁ M. *matu*, B. *matas*
61. old₂ O. *laite*, M. *laqida* 'old man'
62. to put O. *rau* 'lay down, set up', M. *rai*, B. *lai*
63. rain O. *aya*, M. *ai*, A. *ye* 'water'
64. to roast O. *raqaye*, A. *rahai*
65. root O. *ai* 'tendon, sinew, vein', A. *ai*
66. rope O. *taru* 'rope, vein', A. *tila* 'string'
*67. sibling (older) B. *nana*, 'older sister', A. *nano*
68. side O. *weqe* 'side, rim, waist', B. (*hɔt*)*wa* 'sky' = '(sun) region'
69. to sing O. *leule* 'song', M. *leu*
70. to sit O. *mire, imire*, M. *miit*, B. *mit*, A. *miti*
71. to sleep O. *yeete* 'to lie down, to sleep', M. *taqe*, A. *iti, ta, tade*
72. small O. *kiikilai* 'small, young', M. *kaqulai*, B. *kɔl*
*73. to smell O. *muee*, M. *qamuhu* 'to stink'
74. to stand O. *nate*, A. *natei, natia* 'to put up a cult statue'
*75. star O. *ihi*, M. *fi*(*bere*), B. *bi*. M. *bere* means 'large'
76. steal M. *mani*, B. *bini*
77. stone₁ O. *here*, B. *hɔl*
78. stone₂ M. *afa*, B. *op* 'mountain', A. *vi*
79. tail O. *ula*(*pula*), M. *ula*. O. *pula* means 'sprout, sprig'
80. to throw O. *lene*, M. *lian*
81. top O. *ia, ia-a* 'the top of', B. (*doo*)*qa* 'roof' = '(house) top'
82. tree O. *ete*, M. *at*, B. *tel*, A. *tei* 'bush'
83. to walk M. *wale*, A. *filia*
84. war O. *talu* 'to resist, vanquish', M. *sala*, B. *tala*, A. *taloi, taluka* 'to fight'
85. water O. *ira*, M. *ira*, B. *il*

86. what? O. *ina*, M. *naqi*, B. *en*, A. *nala*
*87. when? O. *tun* 'time, period', B. *twen*, A. *tena*
88. where? B. *teo*, A. *tehi*
89. to wish O. *kaarak*, M. *kaarak*
90. woman O. *tuhur, tuhuru*, M. *tufurai*
91. to wrap up M. *bɔka*, B. *(g)ɔbɔk* 'to wrap (it) up', A. *buk* 'to tie'
92. year B. *to*, A. *tung*.

A number of recurrent correspondences may be noted in the foregoing material. For example Oirata *h* commonly corresponds to *f* in other languages (30, 32, 75, 90) except in final position where it corresponds to zero (44, 49). Abui *h* is found where there is zero in other languages (18, 57, 88), Abui *r* after a vowel usually corresponds to *t* in Makasai and Oirata and to *d* in Bunak (15, 19, 29, 49), Bunak *l* corresponds to the *r* of other languages (13, 37, 38, 55, 62, 77, 85), etc.

The next set of non-AN languages to be considered are those of North Halmahera not far from the western extremity of New Guinea. These languages form an obvious group of closely related languages so that no demonstration is necessary.

The basic cleavage within the group is between the southern languages Ternate and Tidor and the remainder. The languages in roughly geographical order beginning from the north and with closely related languages listed together are as follows: Loda (Hueting 1908; van Baarda 1904; Fortgens 1928), Tobelo (Hueting 1936; Fortgens 1928; Ellen 1916*a* and 1916*b*; van der Roest 1905), Tabaru (Fortgens 1928), Isam (Hueting 1908), Pagu (Ellen 1916*a*; Fortgens 1928), Tololiku (Hueting 1908), Wai or Waioli (Hueting 1908; Fortgens 1928), Sau (Fortgens 1928), Modole (Ellen 1916*b*; Fortgens 1928), Galela (Hueting 1908), Ternate (Hueting 1908; Fortgens 1928, 1930) and Tidor (Hueting 1908).

To the north of the eastern part of New Guinea, non-AN languages are spoken on New Britain and New Ireland. Languages of New Britain for which sources were available to me are Taulil (Laufer K.1032; Futscher 1959), Butam (Futscher 1959), Baining (Rascher K.966; Futscher 1959 for the northern dialect), Sulka (Mueller K.1028) and Uasi (V). Of these Butam and Taulil are hardly more than dialectally different and Taulil-Butam, Sulka, and Baining may be considered a genetic subgroup within Indo-Pacific. Uasi on the other hand seems to be at least as close to the non-AN languages of Bougainville to the southeast. The languages thus far mentioned are all spoken in the central or eastern part of New Britain. Capell (1954: 39) mentions a language Idne in the extreme western part of the island for which no sources are available.

In spite of the separate and equivocal status of Uasi, all the non-AN languages of New Britain will be treated as a subgroup in the Indo-Pacific etymologies in the concluding section of this paper.

There is known to be at least one non-AN language on New Ireland immediately to the north and east of New Britain. It is called Panaras by Capell, and the only material in print is the translation of the phrase 'our Father who art in heaven' (Capell 1954). In Chinnery (n. d.), an anthropological account of Northern New Ireland, kinship systems are recorded and two of these, Limalua and Letatan are non-AN. Their geographical position agrees well with that of Capell's Panaras but the word for 'father' does not in either case resemble that in the phrase given by Capell. Limalua and Letatan are closely related to each other.

To the southeast of New Britain and Ireland, the island of Bougainville, the northernmost major island of the Solomons, is largely inhabited by speakers of non-AN languages. Allen and Hard (n. d.) give a classification of the non-AN languages of Bougainville based on a glottochronological study, but without any accompanying data. Their most general result is that there are two stocks on Bougainville, Kunua-Keriaka-Rotokas-Eivo and Nasioi-Nagovisi-Siwai-Buin. Of the former group, the only published material is on Konua (Mueller K.981), evidently the same language as the Kunua of Allen and Hard. For the latter group there are data from Siwai, also called Motuna (Macadam K.1007), Telei (Grisward K.1035; Thurnwald K.1036; Wheeler K.1037), Baitsi (Macadam K.968), Nasioi (Macadam K.1009; Rausch K.1010 and K.1011; Ogan 1966), Koromira (Rausch K.1010) and Koianu (Macadam K.980). Of this second group Siwai, Telei, and Baitsi belong together as against Nasioi, Koromira, and Koianu. My own results agree with those of Allen and Hard insofar as data are available. Thus, of the languages just enumerated Konua is the most divergent. At the same time it has enough in common with the remaining languages of Bougainville to justify setting up the non-AN languages of the island as a distinct subgroup of Indo-Pacific.

Farther south in the Georgia archipelago are four non-AN languages on separate islands which seem to constitute another subgroup. These are Bilua on Vella la Vella (Ray K.1162; Ribbe K.1063; Lanyon-Orgill K.1057; I), Baniata on Rendova (Ray K.1069; Waterhouse K.1070; Lanyon-Orgill K.1057), Laumbe (Ray K.1096; I) and Savo (Codrington K.360; Ray K.1132; I; G). This group will be referred to as the Central Solomon languages.

Davenport (1962) has called attention to the non-AN character of the non-Polynesian languages of the Santa Cruz islands to the east of the group just enumerated. He cites glottochronological evidence to divide these into the language of the Reef on the one hand and those of Ndeni, the chief island

of Santa Cruz. These latter are closely related to each other. Davenport has made available copies of the TRIP vocabularies of the Reef language and of Banua, Nea, Noole, and Nabalue. Examination of these materials leads me to agree with Davenport both with regard to the non-AN affiliations of these languages and the subgrouping into Reef and Ndeni. It also appears probable that the Santa Cruz group has a special relationship to the Central Solomon languages. The larger grouping consisting of Santa Cruz and Central Solomons is called here the Central Melanesian group. The term Melanesian is of course generally applied to the AN languages of this area but in the present instance the word Melanesian is intended in its geographical sense. Evidence will be adduced here in support of both the Central Solomon and Central Melanesian group as a whole.

The problem of AN loans as a source of resemblances among non-AN languages is a particularly vexing one in regard to the Central Solomon languages which have been under strong Austronesian influence.

An example is Bilua *kola*, Savo *kalakala* both meaning 'liver'. Since the nearby Melanesian languages Bugotu and Kirunggela have *kola* and *kolae* respectively this item is not cited in the word list below. In other instances forms have been included when on balance it seemed that a reasonable case could be made for a non-AN origin in spite of similarities with AN languages.

The most powerful indication of the unity of the Central Solomon group comes from the pronominal system. A previously strong case is made even more impressive from data in the unpublished Ivens papers. A series of sentences translated from Mota into Savo and Mota into Laumbe give a more consistent and complete picture of the pronominal system than has been available in print. In particular it shows that the Savo inclusive and exclusive first person plural pronouns had been reversed in the published sources. They are now brought into line with other languages of the group. It further appears that Savo has an inclusive-exclusive distinction in the dual also and these forms resemble the corresponding ones in Bilua. Unfortunately nothing is available regarding the pronouns of Baniata except for a probable second person singular form *no* deduced from one of Waterhouse's sentences. The following are the chief points of resemblance:

1. First person singular. Savo *angi*, Laumbe *ngai, nga* (object prefix), Bilua *anga*
2. Second person singular. Savo *no*, Laumbe *ngo* (bound subject, object and possessive), Bilua *ngo*, Baniata *no*
3. Third person singular. Laumbe *o* (possessive prefix), Bilua *o* (subject prefix)

4. First person dual exclusive. Savo *angge*, Bilua *engge*
5. First person plural inclusive. Savo *mai*, Laumbe *mee, me* (object prefix), Bilua (*ani*)*me*, cf. (*ani*)*vo* first person plural exclusive.
6. First person plural exclusive. Savo *ave*, Laumbe *ee*, Bilua (*ani*)*vo*
7. Second person plural. Savo *me*, Laumbe *imi, me* (bound form), Bilua *me*
8. Third person plural (1). Savo *ze*, Bilua *se*
9. Third person plural (2). Savo *mi* (object suffix), Laumbe *ma, mi* (bound forms).

The Santa Cruz languages show several agreements with the pronouns of the Central Solomons group. Banua and Noole first person singular (*ni*)*nge* and (*ni*)*ngə* respectively may be compared to item 1 above. Reef *imi*, Nea and Noole (*ni*)*mi* second person plural agree with number 7 above. *ni* is a prefix common to all the pronouns in the Ndeni languages. Finally Reef *ingo* resembles Bilua *enge*. Both are first person plural exclusive pronouns.

The following proposed etymologies involve either agreements among the languages of the Central Solomons or between any of these and the Santa Cruz group. The numerous forms restricted to the Santa Cruz group are not included. Language names are abbreviated as follows: Bi. (Bilua), Ba. (Baniata), L. (Laumbe), S. (Savo), R. (Reef), Bn. (Banua), A. (Nea), O. (Noole), B. (Nabalue). Central Solomon citations are separated from Santa Cruz by a colon. Almost all Santa Cruz nouns occur with an *n* prefix which is not separated here from the stem.

1. beach Bi. *zavanga*, Ba. *sava*
2. belly L. *vala*, S. [*m*]*boli* 'guts', O. *bəlu*
3. black S. [*m*]*boraga* (Capell), [*m*]*borayai* (Grace): A. *blu*, O. *iblu*
4. body S. *minjila*, Ba. *monggu*
5. brother Bi. *visi*, L. *vaes*
6. child Ba. (*orou*)*ve* 'fruit' = '(tree's) child', L. *vo*: A., O. *obue*
7. to cook L. *ofai*: R. *ebi*, A. *ubi*, Bn., O., B. *bi*
8. to cry S. *ngei*: R. *engi*
9. dark Bi. *kutumana*, Ba. *kutu*
10. to die Bi. *vou*, S. *ave*: Bn., A., O. *bə*, B. *iba*
11. to eat Bi. *vuato*, Ba. *oua*, La. *eui*: R. *nuue* 'food', Bn. *novu* 'food', A., O. *neiua* 'food'
12. egg L. *keru*, S. *kolei*
13. to enter L. *ele* 'to land', S. *alei*
14. eye Bi. *vilu*, Ba. *bero*

15. face L. *lemi* 'eye', S. *la*[*m*]*bi*
16. to fall S. *taui*: Bn. *tao*, A. *itao*, B. *ido*
17. fire Bi. *uga*: B. *aku*
18. five Bi. *sii*(*ke*), L. *si-ie*. For *ke* as suffix note Bilua *zouke* 'three', *arike* 'four'.
19. four Ba. *avo*: R. *uve*
20. fruit Bi. *epa*, Ba. *po*: R. *nvvpa* 'flower'
21. to go S. [*m*]*bo*, [*m*]*bua*: O., B. *və* 'walk'
22. good Bi. *uriawu*, S. *uria*
* 23. head S. [*m*]*batu*, L. *vatu*. Cf. however Tambatamba, Kirunggela (Choiseul) *botu*
24. to know Bi. *niania*, Ba. *nia*
25. left (hand) Bi. *made*, S. *male*: R. *modi-i*
26. to lie down L. *fufui*, S. *pau*: A., O. *iubu*
27. to live Bi. *elo*: R., O. *Ilv*, Bn., A. *lv* B. *Ili*
28. man Bi. *maba*, S. *mapa*
29. many Bi. *kulio*: Bn., A. *kulu*, O., B. *Iklu*
30. to push S. *eri*: Bn. *aro*, O., B. *aluo*
31. rain Ba. *uzia* 'sky', S. *uza*
32. road Bi. *keve*, S. *keve*
33. root Bi. *kataikili*, L. *kala*: R. *ni-ikile*
34. saliva Ba. *so*, L. *suav*, S. *suasua*
35. shadow Ba. *uzi* 'spirit', S. *vuzi*
36. to sit Bi. *papu*, L. *fifi*, S. *epia*
* 37. to sleep Ba. *isia*, S. *izi*
38. small Bi. *siloa, sila*, L. *talam, tula*
39. Smoke₁ Bi. *iolu*, L. *hou*
40. Smoke₂ S. *akuaku*: O., B. *oko*
41. star₁ L. *simu*, S. *simusimu*
42. star₂ Bi. *pino*, Ba. *finofino*
43. sun Ba. *ui*, L. *kui*, S. *kuli*
44. sweet Ba. *imo*, S. *imo*
45. that Bi. *ko*, L. *hea*: A. *kɔ*, B. *eka*
46. tongue S. *lapi*: R. *nælibi*, Bn. *nalap*, A. *nalepu*, O., B. *nalepn*. But cf. Nggela *lapi*
47. tooth Bi. *neo*, Ba. *nani*, S. *ngale*: Bn. *nangi*, A., O. *ningu*, B. *nungu*
48. tree Bi. *ore*, Ba. *orou*, L. *houla*, S. *kola*
49. two Bn. *eri*, L. (*kana*)*al* 'twenty' (cf. *kanongam* 'ten', *kanelaol* 'twelve'): R. *lil*, Bn., A. *li*, O. (*a*)*li* (cf. (*a*)*tu* 'three', (*a*)*p e* for *a*- as a prefix), B. (*tu*)*li* (cf. (*tu*)*te* 'one', (*tu*)*tu* 'three' (*tu*)*po* 'four')

    50. water Ba. *fio*, S. *piva*
  * 51. wing Ba. *gofona*, S. *gavara*
    52. woman Bi. *reko*, S. *a[n]daki, adoki*

The last non-New Guinea language to be considered is Tasmanian. All the existing material has been collated in Schmidt (1952). Schmidt distinguishes five dialects North East, Middle East, South East, West, and North. The Northern dialect is the most divergent. The present writer agrees with Schmidt's divisions and Tasmanian is cited in the Indo-Pacific word list in accordance with this dialect division.

Having considered the groupings of Indo-Pacific languages outside of New Guinea, we now turn to the very complex situation in New Guinea and small neighboring islands. Seven large groupings of New Guinea non-AN languages are distinguished here. They are discussed in a roughly west-east order as follows: Western, Northern, Southwestern, Southern, Central, Northeastern and Eastern. There are a small number of minor groups and individual languages, often poorly documented, which do not appear to affiliate on present evidence with any of the major groups.

The Western group consists basically of the New Guinea subdivision of the languages which Cowan has outlined and expanded in successive publications as his West Papuan Phylum. In this he includes also the non-AN languages of Timor and Halmahera which have already received separate discussion. The 'second large grouping' of Cowan (1957*b*) consists of the western portion of my Northern group (called Hollandia in my unpublished report of 1958). It is gratifying that there is independent confirmation and exact agreement on this boundary between Western and Northern inasmuch as my own conclusions were reached prior to the appearance of Cowan (1957*b*). Cowan (1960) incorporated into his West Papuan Phylum the Kapaur group, one of those which Anceaux had discussed in his publication on the languages of the Bomberai peninsula (Anceaux 1958) and which I had already included in this grouping in my 1958 report. The Western subfamily should also be expanded to include another one of Anceaux's Bomberai language groups, namely that consisting of Faranjao, Mairasi and Etna Bay. The unpublished material in Anceaux's notebooks which were made available to me provide material beyond that accessible in print which shows this clearly. Some of this evidence will be presented here.

A listing and subclassification of these languages in close agreement with Cowan follows. Material on this entire group was available to me from Anceaux's notebooks. Therefore this is not stated separately for each language and only other sources are mentioned here.

(1) Meninggo (also called Meax), Mansibaber (Wirz K.477; Cowan K.432, 1957*c*), Manikion (van der Sande K.476), Mantion (Cowan K.432, 1957*c*), Amberbaken (Cowan K.432, 1957*c*), Faranjao, Mairasi (Anceaux 1958), Etna Bay (Anceaux 1958), Mogetemin, Konda (Cowan K.432, 1957*c*), Jahadian (Cowan K.432, 1957*c*), Puragi (Cowan K.432, 1957*c*), Kampong-Baru (Cowan K.432, 1957*c*), Solowat, Itigo, Bira, Najarago, Tarof, Barau (Anceaux 1958), Arandai, Mor (Anceaux 1958).

(2) Madik (Cowan 1957*a*, 1957*c*), Karon (Laglaize K.470; Cowan 1957*a*, 1957*c*), Waipu, Moi (Cowan 1957*a*, 1957*c*), Seget, Kalabra (Cowan 1957*a*, 1957*c*), Moraid (Cowan 1957*a*, 1957*c*), Tehit, Waliem.

(3) Ajamaru (Cowan 1957*a*, 1957*c*), Asli-Sidi, Maibrat, Aitinjo.

(4) Kapaur (also called Iha) (Anceaux 1958), Baham (also called Patimuni) (Anceaux 1958), Karas (Anceaux 1958).

It will be noted that Faranjao, Mairasi, and Etna Bay have been included in the first subgroup. These languages seem to be most closely related to Manikion and Mantion from which they are separated by a solid block of AN languages to the northwest.

The following are some of the etymologies which indicate the affiliation of the Etna Bay group to the Western New Guinea family in general and the first of the subgroups enumerated here in particular. Entries from the Etna Bay group are given first and separated by entries from the other languages by a colon.

1. belly Faranjao (*ne*)*we* '(my) belly': Mansibaber (*m*)*ifa* 'body', Amberbaken *ifu*. (Mansibaber has an *m-* prefix on almost all nouns.)
2. black Etna Bay (*ka*)*saɸa*: Maibrat *safi*, Ajamaru, Aitinjo, Mogetemin *safe*. That Etna Bay *ka-* is a prefix as can be seen from (*ka*)*kase* 'red', (*ka*)*sumu* 'heavy', etc.
3. cold Mairasi *argiri*, Faranjao *arjeri*: Manikion *tukurid*, Amberbaken *toguri*. (*r* in Etna Bay languages frequently corresponds to *t* or *d* in other languages.)
4. to come Faranjao *mbuo*: Ajamaru, Seget, Madik, Waipu, etc. *ma*, Jahadian *mo*. (Etna Bay *mb* corresponds to *m* in other languages.)
5. day Mairasi *arawo*: Moi *dewe*, Kampong-Baru *tio*
6. to die Etna Bay *are*: Amberbaken *eut*
7. dog Etna Bay *ansi*, Mairasi *asi*: Mansibaber (*m*)*es* Manikion, Mantion (*m*)*ihi*
8. dry Mairasi *ɸoa*: Mantion *efi*, Amberbaken *ap*
9. ear Mairasi *βira*, Etna Bay *eɸura*: Kalabra (*te*)*fit* '(my) ear'
10. to eat Mairasi, Faranjao *oro*: Manikion (*b*)*it* ('eat!' *b* is second person singular pronoun), Waipu (*n*)*at* 'I eat', etc.

11. fire Etna Bay *iboro*, Mairasi, Faranjao *iworo*: Madik *but*, Karon *bot*
12. to fly Etna Bay *ɸi*: Meninggo *ofo*, Mansibaber *ofu*
13. foot Mairasi (*ne*)*qoro* '(my) foot', Faranjao (*ne*)*oro* '(my) foot': Baham *kweit*, Karas *ko-or*, Mansibaber (*m*)*ukueda*, etc. Mairasi *q*, (glottal stop), Faranjao zero, correspond to Etna Bay *k* and presumably derive from earlier *k*.
14. to go away Mairasi *itai* 'go away!': Manikion (*b*)*eta*, Mantion (*b*)*ita* '(you) go away!'
15. hand Mairasi (*ne*)*rowo* '(my) hand', Faranjao (*ne*)*roβo* '(my) hand': Kalabra (*ne*)*tefo* '(my) hand'
16. lip Mairasi, Faranjao (*ne*)*mbiara* '(my) lip': Amberbaken *imet*; Konda (*pa*)-*meri* (*pa* means 'mouth')
17. moon Mairasi, Faranjao *usira*: Meninggo (*m*)*est*, Mansibaber (*m*)*est*, (*m*)*esta*, Tehit *sirou*, Konda *suri*, Jahadian *huro*
18. rain Etna Bay *ghamu*, Faranjao *ijamu*, Mairasi *yamu*: Ajamaru, Aitinjo, etc. *om*
19. root Mairasi *ɸara*: Kapaur, Baham *pur*, Moraid *pere*
20. shoulder Etna Bay *umburu*: Seget *madu*, Jahadian (*ne*)*madi* '(my) shoulder'
21. skin Faranjao *oreiya*: Puragi *atai*, Kampong-Baru (*na*)*re* '(my) skin'
22. snake Mairasi, Faranjao *ambere*: Inanwatan *meto, medo*, Solowat, Itigo, Bira *medo*
23. to stand Mairasi *isai*: Manikion *esa*
24. sun Etna Bay, Mairasi, Faranjao *tende*: Seget *tale*, Kalabra *til*, Barau *itane*, etc.
25. sweet Etna Bay *ivi*, Mairasi *iβi*, Faranjao *iwi*: Mor *iaba*
26. tail Etna Bay (*na*)*suru*: Manikion, Mantion (*me*)*sera*
27. wet Mairasi *atero*, Faranjao *atoro*: Amberbaken *dot*

There are also important pronominal resemblances. The second person singular absolute pronouns Etna Bay *ne*(*mi*), Mairasi, Faranjao *ne*(*me*), possessives Mairasi *ne-*, Etna Bay *ni*(*mbana*) may be compared to Amberbaken, Madik, Karon *nan*, Waipu, Moi, Waliem *nin*, Ajamaru *nyio*, Aitinjo *nio*, etc. Such second person singular pronouns in *n* are uncommon in New Guinea outside of the Western family. This is also true of the third person singular and plural pronouns in *n*. Third person singular Etna Bay *ne*(*mi*), Mairasi *nini*, Faranjao *nai* resembles Karon *ne*, Manikion *ena*. The third plural pronouns Etna Bay *ne*(*nggi*), Mairasi *niqi*, appear to go with Asli-Sidi *nio*, Ajamaru, Aitinjo *ana*, etc. There is also recorded a second person possessive prefix *ka-* in Faranjao which agrees with Kapaur *ko* 'thou'.

The next major group to be considered is the North New Guinea stock. The western languages coincide with those in Cowan's 'second large grouping' (Cowan 1957*b*). However Cowan's group stops at the political border between former Dutch New Guinea and the Australian mandated territory to the east. This is hardly likely to be a linguistic boundary and indeed Cowan records the judgement of Capell (1954: 11) that the Vanimo of Australian New Guinea is the same language as Sko in West Irian. The Northern group extends far to the east where it encounters a sharp linguistic boundary with the Northeastern subfamily, a boundary already recognized by Capell in principle. The Ndu-Kwoma subgroup of the Northern family was included too late to be incorporated in the following treatment. These and other languages not assigned here to the seven major New Guinea subdivisions will be treated later.

In the following listing of languages, the Northern stock is divided into five major subfamilies in roughly west to east order. For three languages, Siaute, Nori, and Apris, the evidence, while sufficient to assign them to the family as a whole, is not adequate for subgrouping. For all of the languages in the first two groups there was material in Anceaux's notebooks and once more this fact has not been listed separately.

(1) Sentani group: Sentani (Bink K.508; Moolenburgh K.511; van der Sande K.512; Cowan K.509, 1965*b*), Nafri, Tanahmerah II (in distinction from another Tanahmerah to be mentioned later), Nimboran (Schneider K.484; Cowan K.430; Anceaux 1965), Gresik (Cowan K.430), Kemtuk (Cowan K.430), Kuangsu (Cowan K.430), Waibron Bano-Demenggong (Cowan K.430), Demta (Cowan K.430), Mekei (Cowan K.430), Boven Tor (Cowan K.430), Kwesten (Cowan K.430), Mawes (Cowan K.430), Kaure, Sause.

(2) Tami group: Sko (also called Seko) (Moolenburgh K.507; Cowan K.430, K.506), Sangke (Cowan K.430), Arso (Cowan K.430), Njao (Cowan K.430), Wembi (Cowan K.430), Skofro (Cowan K.430), Ampas (Cowan K.430), Waris (Cowan K.430), Vanimo (Thomas 1941), Kilmeri (Sherman K.887; Capell 1954: 12), Amanab (S).

(3) Arapeshan: Arapesh (Fortune K.845; Laycock 1965*a*), Kombio (S), Mountain Arapesh (S), Torricelli (S), Bambita (S), Wam (S), Yambes (S), Kavu (Klaffi and Vormann K.924), Valman (Klaffi and Vormann K.924; Spölgen and W. Schmidt K.926; W. Schmidt K.925).

(4) Murik group: Murik (J. Schmidt K.902), Angoram (Laycock 1965*a*; Thurnwald 1934 = Tjimundo). Tshamberi (Fortune K.845), Kambot (Laycock 1965*a*).

(5) Monumbo group: Monumbo (Vormann and Scharfenberger K.899; Höltker 1964), Lilau (Capell K.815), Bosngun (Capell K.815), Nubia (Capell

K.815), Makarob (Capell K.815), Igom (Capell K.815), Gamai (Capell K.815), Watam (Pöch K.928; Capell K.815), Tanggum (Capell K.815), Murusapa (Capell K.815), Anaberg (Capell K.815), Atemble (Capell K.815), Gapun (Höltker K.861).

(6) Unclassified: Siaute (Wilkes K.913), Nori (Wilkes K.906), Apris (anon. 1922–3).

In the following word list languages are cited in accordance with the foregoing sixfold division. Cowan (1957*b*) has already presented evidence linking the first and second of these groups.

* 1. above (1) Demta *op* (2) Ampas, Waris *op* 'sky'
 2. belly (1) Waibron Bano-Demenggong, Mekei *anu*, Boven Tor *unu*, Kwesten *un* (2) Kilmeri *el*, Njao *aor* (3) Valman *an*
 3. blood (1) Nimboran *king* 'blood, red', Gresik, Kemtuk *kin* 'blood, red', Kuangsu *king*, Mekei *kin* (2) Njao *kenani*, Wembi *kengga* 'red'
*4. bone (1) Nimboran *dɔng*, Gresik, Kemtuk *don*, Kuangsu *dɔng*, Mekei *dɛn*, Kwesten *rɛn* (5) Monumbo *irun*
 5. bow (1) Sentani *fəla*, Nimboran *pro*, Gresik *pra*, Kemtuk *pre*, Kuangsu *para*, etc. (2) Kilmeri *pareawari*, Arso *fare* (4) Murik *panein*
 6. branch (1) Nimboran, Kemtuk *kali*, Gresik, Kuangsu *kari*, Mekei *kala* (2) Arso *kar* (3) Valman *kolo* (4) Makarob *akari* 'arm', Gapun *akan* 'arm'. ('Arm' and 'branch' are frequently the same words in languages of this area.)
 7. breast (1) Sentani *nime*, Nafri *nimɛ* (3) Valman *nyimi*, Arapesh *nyum*(*ab*) (*ab* is a suffix in the noun class system of Arapesh), (5) Monumbo *nim*, Lilau *nima* (6) Siaute *nime*
 8. to burn (1) Sentani *honə* (2) Kilmeri *suna*
 9. to clean (3) Valman *mentyi* (5) Monumbo *mundigak-*(*tset*) 'cleaning (make)'
 10. crocodile (2) Kilmeri *wari*, (4) Murik *warem* (6) Nori *waroma*
 11. to cut (1) Sentani *akənə* (5) Monumbo *kinat* (6) Apris *akanani*
 12. to drink (2) Sko *memu*, Arso *müy*, Sangke *mea* (5) Murusapa *umam*, Gamai *am*, Tanggum (*niam*)*mi* '(water) drink', Atemble *ami* 'eat, drink', Anaberg (*kpa*)*mi* '(water) drink' (*mi* alone means 'to eat')
 13. dry (1) Sentani (Western dialect) *kərəngkərəng*, Demta *kakere*, Kwesten *karkara* (4) Murik *kakran* 'to dry' (transitive)
 14. to eat₁ (1) Nimboran, Gresik, Kemtuk, Kuangsu, Mekei, etc. *dam* (4) Kambot *sɛmʌ* (5) Makarob *tama*
 15. to eat₂ (1) Kaure *mane* (4) Angoram *aməng*, Murik *mön*

16. eye (1) Gresik *nam*, Kemtuk *numu* (2) Wembi, Waris *nof*, Skofro *nop*
    (3) Arapesh *nabe*(*p*), Yambes *nambep*, Torricelli *nambep*, Kombio
    nambæp, Urat *mbep*, Kavu *nambas* (5) Watam *namak*, Bosngun *lamak*

17. feather (3) Valman *okul* 'hair' (5) Monumbo *okul* ('hair' and 'feather'
    are usually the same word in these languages)

18. fire (1) Nimboran *hake* 'torch', Uria *syauk* (3) Valman *aseka* (5)
    Bosngun *səghak*, Watam *zak*

19. foot (1) Sentani (Central) *oro* (*odo* and *oto* in other dialects), Demta *ari*
    'leg', Kavu *ariu* (5) Nubia *ruɛ*, Gamai *ɔr*, Watam *or* (6) Siaute *iro* 'leg'

20. head (1) Kwesten *bar*, Sause *baqar* (2) Kilmeri *pagu*, Arso *bagar, beger*,
    Wembi *mbangger*, Skofro *mbangger*, Ampas *baghe* (3) Arapesh *berag*,
    Kavu *bagiram*, Mountain Arapesh *bagram* (5) Bosngun *pangən*,
    Watam *pangan*

21. high (1) Kwesten *bwarbwara* 'long', (3) Valman *vur*

22. long (2) Arso *kur* (5) Makarob *ikwar*(*imu*) (Makarob *imu* is suffixed to
    most of the quoted nominal and adjectival forms.)

23. house (1) Kaure *ba*, Sentani *obe* 'young man's house' (2) Sko *pa*, Njao
    fa 'nest' (5) Anaberg *bu*, Atemble *bwi* (6) Nori *bau*

24. leaf (1) Sentani *fe* (2) Arso *fiei*, Njao *feye* (6) Nori *pe*

25. light₁ (adj.) (1) Nafri *fafaxɔ* (2) Wembi *fafka*

26. light₂ (adj.) (1) Boven Tor *forera* (2) Kilmeri *pera* (3) Valman *pal*

* 27. louse (1) Tanahmerah II *ming, min*, Sentani *mī*, Demta *ami* (6) Siaute
    imum, Apris *emana*

* 28. meat (1) Nimboran, Kuangsu *ning*, Demta *nini*, Kwesten *nɪn*, Mawes
    ɛnan (2) Sko *na*, Wembi, Skofro *nenggir*, Ampas *nighi* (3) Murik *nagun*

29. moon (1) Sause *uese* 'star' (2) Kilmeri *wais*, Waris *weis*, Wembi, Skofro,
    Ampas *ues*, Arso *wes*

30. mosquito (2) Arso *meing, ming* (3) Kavu *amungau* 'fly' (6) Nori *emeni*

31. mountain (1) Sentani *mɔxɔ, moko*, Nafri *məxɔ* Tanahmerah II *məkɔ*
    (5) Atemble *mag*, Anaberg *mɛga* 'stone' ('Stone' and 'mountain' are
    frequently the same words.)

32. mouse (2) Sangke *sungkri* 'mouse, rat', Skofro *asanggər* (3) Valman
    singir 'rat'

33. new (1) Sentani *nəmə*, Nafri *nimu*, Tanahmerah II *neme*, Nimboran
    iniim, Waibron Bano-Demenggong *nim* (2) Njao *nomu*, Ampas *nəməi*,
    (3) Arapesh *namu*

34. nose (1) Sause *misik* 'mucus', Boven Tor *mase* (2) Ampas *mosoi*, Waris
    mos (3) Mountain Arapesh *mucuh* (4) Angoram *masik*

35. old (1) Central Sentani *patia* 'old man', Mekei *petüe* 'old man', Nafri
    batua, Tanahmerah II *petua* 'old man' (4) Murik *patengo*

36. palm of hand (1) Sentani *fe* (3) Valman *vi*
37. red (1) Boven Tor *berbere*, Kwesten *birbir* (5) Monumbo *purapura*
38. river (1) Sentani *wi*, Mawes *we*, Uria *uea* (6) Siaute *woi*
39. road (1) Nimboran *tab*, Gresik, Kemtuk, Kuangsu, Mekei *tap* (5) Bosngun *tɔup*, Nubia *tiap*, Makarob *tuap(imu)*, Gamai *taup*
40. short (1) Mekei *tambutə* (2) Njao *timer*, Kilmeri *smori* (3) Valman *tambar*
41. shoulder (1) Central Sentani *dəmə* 'wing', Kemtuk, Kuangsu *tabang* 'wing', Waibron Bano-Demenggong, Mekei *tambang* 'wing' (5) Monumbo *tsombi* (6) Nori *tomu* ('Wing' and 'shoulder' are frequently the same word in these languages.)
42. soft (1) Sentani *malə*, Nafri *mare*, Nimboran *marema* (3) Valman *malei*
43. sour (1) Nafri *mani* (5) Monumbo *manging*
44. to stand (1) Mekei *asü* (3) Arapesh *atu* 'stand up' (5) Makarob *tyu* (6) Siaute *atu*
45. to steal (3) Arapesh *kum* (5) Bosngun *kamba*, Nubia *gamba*, Atemble *kabde* (6) Siaute *kupi* 'to take'
46. stone (1) Nimboran, Kuangsu *domuo*, Gresik *dum*, Kemtuk *dumu*, Waibron Bano-Demenggong *domo*, Mekei *dəmə* (3) Valman *tomul*, Kavu *utam*, Arapesh *utom*, Urat *wutom*
47. sun (1) Boven Tor *guwer*, Kwesten *kwer* (5) Igom *nggaara(ka)* (almost all nouns cited in Igom have a *-ka* suffix), Tanggum *ghaar*, Anaberg *gəra*, Atemble *gar*
48. unripe (1) Central Sentani *nani* 'raw' (2) Sangke *nono* (3) Arapesh *ninabei*
49. vagina (3) Arapesh *uruh* (5) Monumbo *iir*
50. white (1) Nimboran *kekruab*, Kuangsu *kəkrap* (4) Murik *kakrep*.

Since grammatical information regarding almost all of the languages in this group is limited to pronouns, it is not surprising that the evidence linking the subgroups is confined to pronominal forms. The most important is the existence of a third person pronoun of the type *mV* or *mVn*. A typical example is Watam *min* third person absolute pronoun both singular and plural. These and other pronominal forms cited here are largely confined to the Northern group and occur only sporadically elsewhere except that a second person *w* is not uncommon in the Western languages and *g* first person in the Northeastern.

1. First person singular: (1) Nimboran *ngo*, Kuangsu *nga*, Waibron Bano-Demenggong, *kaqu*, Mekei *kat* (2) Arso, Ampas *ka*, Skofro *ki*, Kilmeri *kwo* (5) Makarob *kö*, Igom *ku*, Gamai *aka*, Watam *yak*, Tanggum *ku*, Nubia *nggɔ*, *kö*, Bosngun *nggɔ*, Monumbo *ek*, Lilau *iki*

2. Second person singular (1): (1) Kemtuk *mɔt*, Mekei *umɔt*, Boven Tor *mey* (2) Sko *me*, Sangke *me* (4) Murik, Angoram *mi* (5) Bosngun, Nubia *mɔ*

3. Second person singular (2): (1) Demta, Nafri, Tanahmerah II, Sentani *we* (4) Kambot *wɔn* (5) Gamai *wu, u*, Watam *u*

4. Third person pronoun: (1) Nimboran *m*, element common to all genders and numbers of third person bound subject pronoun, Sentani *mi* third person dual and plural bound object pronoun (2) Sko *me*- his, her, their in kinship terms (3) Arapesh *amun* they (masc.) (4) Murik *mön* 'he, she', *moa* 'they', Chamberi *min* 'he, she' (5) Watam *min* 'he, she, they'. Nubia, Makarob *me* 'they', Bosngun *mwe* 'they', Monumbo *ming* 'they (masc.)', Anaberg, Atemble *mɛ* 'they'.

The next large grouping to be treated, the Southwestern, might also be called Marind-Ok after its two most important subdivisions. Ok is the name proposed by Healey from the common word for 'water' in the Kati and Telefomin languages. In addition to these languages the Awju group to the west of Kati, which incidentally has forms similar to *ok* for 'water', and a Tirio group spoken some distance to the east of Marind, should be included. I have also tentatively included the Kukukuku even farther east of the main body of the Southwestern Group. The relatively few Kukukuku entries are, I believe, due to the sparse documentation available, confined to a few short vocabularies but the evidence as far as it goes is rather striking. There follows an enumeration of sources for languages of this subfamily. In particular the Kati language of the Ok subdivision and the Boazi language of the Marind subdivision are represented by rather numerous vocabularies which are only dialectally different or even the same dialects under separate names.

(1) Tirio group: Tirio (Riley K.550), Tagota (Chalmers K.619), Pisirami (Riley K.636), Arama (Williams 1936), Anima (Williams 1936).

(2) Marind group: (*a*) Marind proper (also called Tugeri), (Drabbe 1955; Geurtjens K.520; Kolk and Vertenten K.524), Gawir dialect (Drabbe 1955, K.434), Boven Mbian dialect (Drabbe 1955, K.434) (*b*) Boazi (Drabbe 1955, K.434), Dea (Capell 1954:52), Konmak = Village Y (Austen K.533; Rentoul K.596), Lake Murray (Murray K.601), Jabga (Nevermann K.444), Babwa = Village Z (Austen K.533), Gabgab (Nevermann K.444), Biak = Village X (Austen K.533) (*c*) Jaqai (Drabbe 1955, K.434), Sohur (Nevermann K.514)

(3) Ok group (*a*) Digul (Geurtjens K.520), Aran (Capell 1954), Niinati (Drabbe K.434), Metomka (Drabbe K.434), Lower Muiu (Austen K.533) Muju (A, Schoorl n. d.) Kandam (Austen K.533), Iongom (Austen K.535), Upper Tedi (Austen K.535), Awin (Austen K.535), Eastern Tedi (Austen K.535),

Western Donaldson (Austen K.535), Upper Fly (Austen K.535), Upper Muju (Keyzer in Austen K.481) Marapka (Austen K.533). (*b*) Telefol (also called Telefomin) (Alan Healey 1962, 1964; Phyllis Healey 1965*a*, 1965*b*; Capell K.814, 1954:52; Kirschbaum K.824), Unkia (Champion 1926–7), 'Plain Country' (Champion 1926–7) (*c*) Mandobo (A), Dumut (Drabbe K.434), Kaeti (Drabbe 1959), Wambon (Drabbe 1959).

(4) Awju group. Awju Sjiagha (Drabbe K.456, 1959), Awju Jenimu (called Oser by Nevermann) (Drabbe K.456, 1959; Nevermann K.500), Awju Pisa (Drabbe K.456, 1957, 1959)

(5) Kukukuku group. Kukukuku (Chisholm K.598), Ashavi (Humphries K. 599), Madinava (Humphries K.600)

In the following word list items are not included which are common only to subgroups 3 and 4 (Ok and Awju) since these are quite numerous.

1. bat (1) Tirio *pitite* (2) Marind *pud* (4) Awju Sjiagha, Awju Jenimu *piriri*
2. to be₁ (2) Jaqai *tVbV* (variable vowels depending in gender and number of the subject) (3) Telefol *tebe*(*min*) 'become, appear'
3. to be₂ (2) Jaqai *ne*, Marind *in* (auxiliary verb) (3) Telefol *n*(*in*) 'be, reside'
4. beard (2) Boazi *meitu*, Konmak *moitu* 'whiskers', L. Murray *motu* 'whiskers' (3) Dumut *matut*, Kaeti, Wambon *matit* (4) Aghu *masü*, Awju Pisa *masi*(*rō*) (*rō* means 'hair')
5. bird (1) Tirio *isupe* (2) Marind *uzub*
6. to bite (1) Tirio *su* 'tooth' (2) Boazi, Konmak *se*, Biak *tsi* (3) Wambon *atigo*, Dumut *atinggo* (4) Awju Sjiagha, Jenimu *ati*, Awju Pisa *su*, Aghu *asi*
7. bitter (2) Boazi *kagha* (3) Niinati *ko-ok*, Metomka *kok*, Dumut *koko*
8. blood (1) Tirio *adada* (2) Marind *do*
9. bone (1) Pisirami *toto* (2) Upper Fly *tu*
10. breast (2) Kandam, Marapka, Niinati, Metomka *muk* (5) Madinava *muk*
11. calf (of leg) (1) Pisirami *caron*, Tagota *sara* (3) Unkia *serip* 'leg'
12. child (2) Jaqai *mak* (4) Aghu *amoko* (5) Madinava *imega-kaivagu*
*13. to come (2) Marind *man* (3) Upper Muju, Digul, Niinati, Metomka *mene*, Mandobo *mend*
14. dark (1) Tirio *dinaia* (2) Marind *dino*
15. to die (2) Boazi *ani* (3) Digul *ain*, Metomka *ane*
16. to drink (1) Tagota *iemo*, Tirio *aiamea* (2) Takai *mi* 'water', Sohur *mui*, *mi* 'water' (3) Kaeti *emi*, Wambon *mi*, Digul *mu* 'liquid' (4) Awju Sjiagha, Jenimu, Pisa *mi*, Aghu *mi*, Oser *mwi*
17. dry (1) Tirio *kakaku*(*gaga*) (For *gaga* as suffix compare *team*(*gaga*) 'sour'.) (3) Upper Muju *koknen*, Niinati *kok*

18. ear (2) L. Murray *kumbit*, Marind *kambet* (3) Digul *kembe*, Metomka *kende*, Mandobo, Kaeti, *kere*(*tap*) (*tap* means 'hole') (5) Ashavi *kata*, Kukukuku *kaitu*, Madinava *kada*

*19. earth (2) Marind *makan* 'earth, underneath', Jaqai *mokon* (3) Telefol *bakan* 'ground' (4) Awju Sjiagha, Jenimu *moka* 'earth, underneath', Awju Pisa *mokã*, Aghu *makã* 'underneath'

20. egg (2) Marind *kana* (4) Awju Sjiagha *gena*, Awju Jenimu *gana*

21. fat (n.) (2) Marind *kaber* (3) Digul *kob*, Niinati, Metomka *kop* (4) Awju Sjiagha and Jenimu *kepe*

*22. fire (2) Boazi *te* (3) Awin *de, di* (5) Madinava *da*

23. flower (2) Boazi *tetek* (3) Kandam *atet*, Niinati *tet*

24. forehead (2) Marind *nani* 'face', Boven Mbian *nenah* 'forehead, face', Gawir *nanih* 'face' (4) Oser *nene*

25. fruit (2) Marind *mokom* (3) Wambon *mok*

26. heavy (2) Marind *kanir* (3) Digul *kun*, Dumut *kakun*

27. head (2) Jaqai *muku* (3) Dumut *muk* 'brains', Kaeti *mök* 'brain' (5) Madinava *maha*

28. husband (3) Telefol *imak* (5) Madinava *imaca*

29. jaw (2) Marind *ete* (4) Aghu *te*

30. lip₁ (1) Pisirami *tapar*, Tagota *taper*, Tirio *diware, taware* 'mouth' (no doubt the same word recorded two different ways) (2) Marind *utup* (3) 'Plain Country' *tobone*

*31. lip₂ (3) Lower Muiu *meia*(*kat*) (*kat* means 'skin') Digul *meya*(*kat*) (5) Ashavi *mea*, Madinava *mega*, Kukukuku *mia* 'mouth' (cf. Indo-Pacific 'mouth')

32. liver (2) Marind *on* (3) Kandam *oni*, Marapka *onyi* (4) Aghu *ũ*

33. long (2) Jaqai *pe* (4) Aghu *pi*

34. lungs (2) Gawir *ahah* (3) Niinati, Metomka *okok*

35. to make (2) Marind *kam* (3) Niinati, Metomka *kamee*

36. man (1) Tirio *tanim* (3) Telefol *tanum*

37. moon (2) Jaqai *kamo* (5) Ashavi *kaamia*, Kukukuku *kama*

38. mouth (2) Boazi *mangkangka*, Konmak *manganga*, Biak *mangat* (3) Metomka, Niinati, *monggot*, Mandobo *mangot*, Dumut *mangkot*, Kandam *mongo*(*tem*) (*tem* means 'hole') (5) Madinava *manga*

39. nail (of finger) (2) Boven Mbian *tek* (3) Digul *tuk*, Niinati *duk* (4) Awju Sjiagha *dogho*

40. new (2) Marind *noh*, Jaqai *nogoke* (4) Awju Sjiagha *noghonggo*

41. nose (2) Biak *megi* (5) Ashavi *imagwo*, Kukukuku *imuk*

42. root (2) Marind *te*, Gawir *itit* (4) Awju Sjiagha *te*, Awju Pisa *tate*, Aghu *tete*

43. rope (2) Jaqai *kikiner* 'sinew' (4) Awju Pisa, Aghu *kikī*

44. to run (2) Boazi *oko* (3) Kaeti, Wambon *ko* (4) Aghu, Awju Pisa *gho*, Awju Sjiagha and Jenimu *xo*
45. saliva (2) Boazi *kasi*, Konmak *kussi*, Marind *kase* (3) Dumut *katet*, Kaeti, Wambon *katet* (4) Aghu *ghasi*, Awju Sjiagha *xate*
46. to say (2) Jaqai *bak* (3) Telefol *bokoo*
47. to see (2) Marind *idih* (3) Kaeti *itigio*, Dumut *eteko*
48. short (2) Jaqai *domakae* (3) Telefol *duumat, duumaak*
49. shoulder (2) Boven Mbian, Gawir *muk* 'upper arm' (3) Kaeti *mak* (4) Awju Pisa *makie*
50. to sleep (1) Tirio *mua* (2) Marind *nu*
*51. to smell (1) Tirio *imsiga* (n.) *intere* (v.) (2) Gawir *imu* (n.) (3) Kaeti *umo*, Niinati *imone*
52. star (1) Tagota *mano* (2) Jaqai *mind*, Marind *mandau* 'moon' (3) Marapka *mindong*, Muju *minod*, (4) Awju Pisa, Aghu *mī*, Awju Sjiagha and Jenimu *mi*
53. to steal (1) Tirio *nono* (3) Metomka *niine* 'seize'
*54. stone (1) Tirio *guma* (2) Konmak *kum*
55. tooth₁ (2) Marind *manggat*, Jaqai *mangger* (4) Awju Pisa and Jenimu, Aghu *maga* (5) Ashavi *mangya*, Kukukuku *menyi*, Madinava *manga*
56. tooth₂ (1) Pisirami, Tagota *kama* (2) Lake Murray, Babwa *kama*
57. tree (1) Tirio *kesowa* (4) Aghu *kesaghe*
58. river (1) Tirio *gawa* (5) Ashavi *hauwa*
59. wing (2) Boven Mbian *pur* (3) Niinati *purukap*, Kaeti *mborō*, Wambon *mburui*, Metomka *mburu*
60. wood (1) Tagota *atiati* (2) Marind *de* 'tree, wood' (3) Kandam, Marapka, Niinati, Metomka, Telefol *at* 'tree, wood' (4) Oser *dadi*

Thanks to the work of Dutch linguists in the Marind, Ok, and Awju languages and research on Telefol by the Healeys, we are much better informed regarding their grammatical structure than is usually the case for Indo-Pacific languages and it will therefore be possible to point to other phenomena than the personal pronouns. The most striking is the following. As will be mentioned in more detail later, grammatical gender is a widely distributed Indo-Pacific characteristic and it is very commonly expressed by vowel alternations. The Marind and Ok language groups share one specific form of this pattern, by which masculine singular is indicated by *e*, feminine singular by *u* and plural for either gender by *i*. It occurs in various aspects of the language but will be illustrated here from the third person pronouns. Examples in the Marind group are Marind *anep* 'he', *anup* 'she', *anip* 'they', Boazi *ndene* 'he', *ndunu* 'she', *ndini* 'they', which are to be compared with

Niinati *ye* 'he', *yu* 'she', *yi* 'them' and similar forms in the Ok languages. The exact correspondence of the two copula verbs in Jaqai, a Marind language, and Telefol seen in the foregoing lexical entries for 'to be' is also striking.

In regarding the ordinary pronominal forms for the Kukukuku languages the only information in print is Capell's note that *na* is the first person pronoun in the dialects of Obi and Kaviropi (Capell 1954: 61). In the Tirio group Riley gives the complete set of independent pronouns for Tirio itself. As will be evident from the following comparisons these pronouns show clear correspondences with other subgroups of the Southwestern stock. The second person plural pronoun in *z* is particularly significant because it is not found elsewhere in Indo-Pacific.

1. First person singular (1) Tirio *nogao*. The one pronoun recorded for Pisirami of this subgroup is *nog* 'thou'. I strongly suspect that it is really the first person singular pronoun, a common reversal under unfavorable conditions of elicitation (2) Marind *nok*, Jaqai *anok*. The remaining subgroups have *n* forms without final *k*. These are very common throughout Indo-Pacific.

2. Second person singular (1) Tirio *ogao* (2) Marind *oh*, Jaqai *ak*, Boazi *gho* (3) Muju *kep*, Telefomin (Capell) *kap*, Dumut, *nggu*(*p*) (*p* is a suffix found in all pronouns in Dumut), Wambon *nggo* 'thy' (4) Awju Sjiagha *go*, Awju Jenimu and Pisa, Aghu *gu*

3. Third person singular (1) Tirio *igi* (2) Marind *ehe* 'this' (masc. sing.) (Marind *h* often corresponds to *k* elsewhere). (3) Kaeti *ege* (4) Awju Sjiagha *ege*, Awju Jenimu *egi*, Awju Pisa *eki*

4. Third person singular (2) Marind *epe* 'that' (masc. sing.) (4) Aghu *efe*

5. Second person plural (1) Tirio *zogao* (2) Boazi *zu*, Marind *z-* objective bound form

We turn to the Southern subgroup which might be called Kiwaiic after the best known language of the family. It includes the Miriam language of the Eastern Torres Straits Islands between New Guinea and Australia. Miriam shows a special relationship to Jibu, Kunini, and Oriomo on the south New Guinea coast and is therefore included in the same subgroup as these latter languages. The westernmost part of subgroup (7) consists of the languages of Frederick Henry Island and the adjacent mainland and is thus split off from the remaining Southern languages by the intrusion of Marind. Some of the languages of subgroup 7 have borrowed extensively from Marind, in particular Jab and Makleu.

(1) Kiwai (Riley K.550; Riley K.636; Wurm K.595), Ipikoi (Chance K.578), Hibaradai (Austen K.574), Hiwi (Austen K.575), Urama (Herbert K.631),

Iwainu (Woodward K.579), Goaribari (anon. K.587; Ray K.547), Kerewa (Woodward K.588), Turama (Herbert 1914–15), Era River (Herbert 1914–15), Mawata (Ray K.636), Domori (Riley K.550), Wabuda (Riley K.550), Sisiame (Riley K.550; Ray K.547), Pirupiru (Riley K.550) Dibiri (Woodward K.567), Karami (Flint K.583), Eme-eme (also called Pepeha) (Johnston K.615), Mahigi (Cridland K.602), Tureture (Riley K.550), Tapapi (Austen K.621) (called Tapari in Klieneberger), Buniki (Ray K.547)

(2) Barika (Johnston K.562), Dugeme (Saunders K.569), Karima (Johnston K.706), Foraba (Lind K.571), Ro (also called Keai or Worugi) (Lind K.616), Sesa (Chance K.618), Tumu (also called Dumu or Kibiri) (Chinnery K.570, Bevan K.438)

(3) Kunini (Riley K.550), Masingara (Ray K.548), Oriomo (Riley K.550), Jibu (Murray K.580), Miriam (Ray K.636), Gijara (Lyons K.572)

(4) Bugi (Chalmers K.563), Dabu (Riley K.550; anon. K.565), Dibolug (Austen K.568), Kibuli (Karius K.585), Ngamai-iki (also called Tupadidi) (Lyons K.614), Mbayaka (also called Jindabira) (Lyons K.606), Agob (Lyons K.556)

(5) Parb (Riley K.550), Sanana (Ray K.548), Dungerwab (anon. K.554), Dorro (Riley K.550), Dapo (Lyons K.542), Nombuio (Lyons K.542), Noraia (Lyons K.542), Potaia (Lyons K.542), Tunjuamu (Herbert K.620), Nenium (also called Wakamara) (Lyons K.613), Karigari (Austen K.584), Moi-e (Rentoul K.608)

(6) Peremka (Riley K.550), Bangu (anon. K.554), Wandatokwe (Lyons K.542), Jiminakana (Rentoul K.581), Nausaku (Austen K.612), Kebanagara (Rentoul K.586), Mani (Nevermann K.444).

(7) Jei (Drabbe K.434; Geurtjens K.437), Toro (Ray K.548), Ngowugar (Nevermann K.444), Kanum (Drabbe K.434; Nevermann, K.444; Boelaars K.427), Moraori (Drabbe K.434; Nevermann K.444; Boelaars K.427), Komelom (also called Mombum), (Geurtjens K.437; Drabbe K.478), Koneraw (Geurtjens K.437), Kimaghama (Drabbe K.472), Riantana (Drabbe K.472) Keladdar (Geurtjens K.437), Teri-Kawalsch (Geurtjens K.437), Ndom (Drabbe K.472), Jab (also called Jelmek) (Drabbe K.465; Geurtjens K.437), Makleu (Drabbe K.645).

A list of probable cognates among these seven subgroups of the Southern subfamily follows. These comparisons are so numerous that the following classes of entries have been eliminated; nearly all those which are found in Indo-Pacific more generally and those involving only subgroups 5 and 6 which are particularly close genetically.

1. ashes (1) Iwainu, Kerewa, Pirupiru, Buniki *tuo*, Kiwai *tuwo*, Domori *tue*
(2) Sesa *tou* 'lime' (3) Miriam *tibi* (7) Jei *tebetebi*, Moraori *timbwe*

2. back (5) Dorro *demi* (6) Peremka *təmen* (7) Komelom *temmeni* 'behind'

3. bad (1) Wabuda *isabe*, Domori *sosauebi* (5) Dorro *sapute* 'bitter, sour', Parb *tabud* 'bitter, sour' (7) Moraori *saf* 'bitter, bad', Kanum *sam* 'bad, bitter', Jei *capaoo* 'bitter'

4. belly₁ (2) Dugeme *kambubu*, Karima *kabubu* (3) Jibu *komu*, Miriam *kem* (4) Bugi *kam*, Agob, Dabu, Kibuli, Kawam *kom*, Dibolug *komo*

5. belly₂ (1) Iwainu, Kerema, Domori, Sisiame, Buniki, Karami *niro* (6) Peremka *nur*, Tokwasa *nar*, Wandatokwe *nernera* (7) Ngowugar *nar*

6. bird (1) Hiwi, Sisiame *ipoipo*, Mahigi *buwa* (2) Sesa *paa*, Ro, Bara, Ibukairi *ba* (3) Gijara *poiai* (4) Bugi, Dibolug, Kibuli, Kawam, Ngamai, Mbayaka *pa* (6) Jiminakana *iviov*

7. black (3) Oriomo *timi* 'black, charcoal', Jibu *timi* (7) Jab *temit* (compare (1) Urama *tamai-idiidi*)

8. blood (1) Tapapi *dede* (2) Barika *eta* (3) Kunini *udi* (4) Bugi *teia*, Mbayaka *tia* (7) Ndom *eth*, Jelmek *ato* 'red', Teri-Kawalsch *duddu*

9. bone (3) Gijara *kuja* (4) Dabu *kut*, Kibuli, Kawam *kutra*, Dibolug *kute*, Mbayaka *kuic*, Agob, Ngamai *kuta*, Bugi *kwetr* (5) Dorro *goart*, Parb, Dunger-wab *kwod*, Tunjuamu *guat* (6) Tokwasa *kuart*, Bangu *kuar*, Kebanagara *guer* (7) Keladdar *kadrowa*, Moraori *nggwar*

10. boy (3) Miriam *kebi*(*le*) (*le* means 'person'), Oriomo *kewar* (6) Peremka *kapar*

11. branch (1) Mawata, Kiwai *atomo*, Sisiame *atema*, Hibaradai, Wabuda *katomo*, (3) Miriam *tam* (6) Jiminakana *idami*, Bangu *tambia* 'hand'

12. to bury (1) Kiwai *ogubiri* (3) Kunini *gapebora*, Jibu *gope-yarenten*, Miriam (*et*) *kobedi* (7) Jab *gubigubi*

13. buttocks (4) Bugi *kum*, Dabu *kuma* (5) Dungerwab *kwomb* (7) Teri-Kawalsch *cumme*, Kimaghama *cöme*(*börö*) (compare *cömedadu* 'anus')

14. cheek (1) Goaribari *tavatava*, Domori *tatamu* 'lower jaw' (4) Bugi *tabe* 'jaw', Dabu *teb* 'jaw' (7) Kanum, Moraori *sama*, Kimaghama, Teri-Kawalsch *cama*

15. child (2) Tumu *pori* (5) Parb *pier* 'boy, son', Dungerwab *pierr* 'son' (7) Jei *por*

16. chin (1) Mawata *baga* 'chin, jaw', Kiwai *baga* (3) Miriam *bag* 'cheek' (6) Bangu *bagi*(*thomba*) 'beard' (*thomba* means 'hair')

17. to come (1) Urama *ouo* 'come here!', Era River *oua* (5) Dorro *awe*, Keraki *auwe* (6) Peremka *awe* (7) Jab *we*

18. crocodile (3) Gijara *kuri* (5) Karigara *karara* (6) Nausaku *kerera* (7) Kanum *keri*

19. to cry (3) Oriomo *iee*, Jibu *ye* (5) Korro, Parb, Dungerwab *ye* (6) Bangu *ye* 'to shout'

20. deep (1) Era River *amuo* 'long' (3) Miriam *muimui* (7) Riantana *moa*, Keladdar *mowi* 'large'

21. egg (1) Hibaradai *kikop*, Hiwi *kikopu*, Wabuda *kikopu* 'egg, fruit' (3) Oriomo *kəp* 'egg, fruit', Jibu *köpö* 'egg, fruit', Kunini *ku-u* 'egg, fruit' (4) Bugi *kapa*, Dabu *kop*, Agob (*pa*)*kop* (*pa* means 'bird'), Kibuli, Kawam *kopa*, Ngamai, Mbayaka *kapo* (5) Dorro *kop*, Nenium *kopo*

22. excrement (1) Mawata, Kiwai, Domori *ne*, Dorro *na*

23. eye (2) Barika, Dugeme *si* (6) Peremka, Wandatokwe, Tokwasa, Mani *si* (7) Kanum *si*, Ngowugar *ji*

24. to fight (4) Dabu *mok*, Kibuli, Kawam *moka* 'war' (5) Tunjuamu *moko*, Keraki *moku*

25. fire (2) Kibiri *ebani*, Tumu *ibani* (5) Dorro *bengi*, Nenium *bonja*, Nombuio *bəsh*, Moi-e *benji* (6) Peremka, Bangu, Tokwasa *meni*, Wandatokwe *mini* (7) Jei *ben, beny*, Kanum *mens*

26. foot (1) Mahigi (*dedi*)*kaba* (2) Ro *kaba*(*kea*) (4) Dibolug *guabo* (5) Potoia *kabkab*, Nombuio *kobkob*, Parb *kabukab*, etc. (6) Bangu *kabokabo*, Tokwasa *kabkab*

27. to give (3) Jibu *ai* (7) Kimaghama *iye*, Riantana *ii*, Makleu *ai*

28. hand (1) Hibaradai (*koto*)*pata*, Hiwi (*la*)*pata*, Turama (*to*)*pata*, Mawata, Sisiame (*tu*)*pata* 'palm' (*tu* means 'hand') (4) Dibolug (*tang*)*pute* (*tang* means 'arm') (5) Tunjuamu *pər* (6) Peremka *pata* 'arm'

29. high (1) Turama *tuturu* 'high, long', Era River *tutu*, Urama, Goaribari *tutu* 'long', Dibiri *tuturuna* 'long' (7) Teri-Kawalsch *tutura* 'long'

30. to hit (5) Parb *toar* 'kill', Dungerwab *turendi* 'strike' (7) Jei *tar*

31. house (2) Kibiri *mi* (4) Bugi *mae*, Dabu, Ngamai, Kibuli, Kawam *ma*, Agob *ma* 'house, village', Mbayaka, Dibolug *mai* (7) Jei *ma*

32. island (5) Dungerwab *koor* (7) Jei *kar* 'earth', Moraori *ke-er* 'earth'

33. jaw (1) Domori *ipu-u* 'upper jaw' (3) Miriam *ibu* 'chin, jaw' (7) Makleu *obo* 'cheek'

34. to know (1) Mawata, Kiwai, Domori etc. *umoro* (3) Oriomo *umre*, Kunini, Miriam *umele* (4) Dabu *umarandangar* (7) Kanum *amara* 'to hear'

35. land (1) Pirupiru *sabo*, Buniki *saboa* 'earth', Hiwi *sobkeri*, Kiwai, Dibiri *sopu* 'earth', etc. (3) Miriam *seb*

36. large (1) Kiwai, Mawata, Sisiame etc. *auwo* (3) Miriam *au* 'large, broad'

*37. leaf (1) Sisiame, Pirupiru *pori* (2) Barika *iboro* (3) Gijara *poringai*, (4) Kawam *per* (7) Komelom *pur*, Mombum *epur*, Koneraw *bur*

38. left (hand) (1) Mawata *pero*, Kiwai, Domori, Sisiame *pere* (3) Miriam *ber* (4) Dabu *barr, berr*

39. light (n.) (1) Sisiame *parapara* (4) Dabu *paraparan* 'white' (7) Mombum *fera-fera* 'bright', Jei, Kanum *pela* 'white'

40. liver (1) Goaribari, Turama, Era River *beru* (7) Kanum, Moraori *mbur*, Jab *bulo* 'heart', Jei *ber(kup)* 'heart'

41. man (1) Kiwai, Domori, Wabuda, Dibiri, etc. *dubu*, Hiwai *ururubi*, Kiwai *arubi*, 'people' (4) Dabu *rabo*, Ngamai *labo* (5) Dorro *darube* (7) Kanum *irebe* 'person', Komelom *idub*, Koneraw *irew*, Ndom *ref* 'person'

42. nail (of finger) (3) Miriam *tanka* (6) Bangu *tanka*

43. navel (1) Goaribari, Hibaradai, Domori, etc. *upuru*, Kiwai *upuru, gupuru* (3) Kunini *opolo*, Jibu *kopöro*, Miriam *kopor*, Gijara *kopul* (7) Jei *ipr, yipel*

44. neck (1) Pirupiru, Buniki *dopa* (2) Bara *doba*, Sesa *dubatugi*, Ro *dabadigi* (3) Miriam *tabo*

45. night (3) Jibu *serin* (4) Kibuli *tirem* (5) Dorro *serumb*, Nenium *cerum*, Karigari, Moi-e *seram*

46. nose (1) Kiwai, Mawata, Domori, Pirupiru, etc. *wodi*, Turama, Era River, Goaribari *wadi* (4) Bugi *wede*, Dibolug *wide*, Mbayaka *wede* (7) Keladdar *wanda*

47. rain (2) Dugeme *mu* 'water', Karima *mowa* 'river' (7) Kimaghama *moa*, Keladdar *mowa*, Teri-Kawalsch *muwo*

48. road₁ (1) Kiwai, Mawata, Urama, etc. *gabo* (3) Kunini *gabe*, Miriam *gab*

49. road₂ (2) Barika *ti*, Foraba, Ro, Bara *tu*, Ibukairi *du* (4) Mahigi *idi*, (7) Keladdar *do*

50. root (1) Sisiame, Pirupiru *sipi* (3) Miriam *sip*

51. saliva (4) Dabu *burme*, Kibuli *bolumi*, Kawam *boluma* (5) Dungerwab, Dapo *berim*, Nombuio, Noraia *berem*, Potaia *beram*, Nenium *borom* (6) Bangu *barim*, Wandatokwe, Tokwasa *berem*

52. sand (4) Bugi *cire* (7) Keladdar *cir*, Koneraw *shire*, Mombum *sir*, Jei *jirojiro*

53. side (1) Mawata, Kiwai *bara* in *bara(hara)* 'rib' = 'side bone', Goaribari *bari* (3) Miriam *ber*

54. to sit (1) Mawata, Kiwai *omioi*, Urama *emie*, etc. (7) Kanum *mi*, Moraori *öme*

55. skin (3) Miriam *gegur* (6) Kebanagara *gogara* (7) Ndom *krikir*

56. snake (2) Barika, Dugeme *kavu*, Kibiri *kapo* (7) Moraori *kaf*

57. soft (1) Dibiri *boroboro* (5) Tunjuamu *abarat* (6) Mani *porfor* (7) Kimaghama *veröverö*

58. star (2) Ibukairi *pure* (4) Dabu, Kibuli, Kawam *piro*, Agob *paro* (7) Kimaghama *böree*, Teri-Kawalsch *mburrewa*

59. stone (3) Miriam *mat* (7) Komelom *mate*, Koneraw *mate*, Kimaghama *mete*, Riantana *metö*, Keladdar *mate*, Ndom *meet*, Jelmek *mata*, Jab *mate*

60. sun (1) Mahigi *ariarima* 'daylight' (3) Jibu *loma*, Miriam *lem* (7) Jelmek *alemu* Jab *alim*, Makleu *olimu*

61. tail (1) Kiwai, Domori, Sisiame *wapo* (3) Miriam *upi* (7) Makleu *wibi*, *ubi*, Jab *übi*, Mombum *pi*

62. thunder (1) Urama *guru* (3) Kunini *gururu*

63. tooth (2) Foraba *sregi*, Ro *sirigi*, Bara *sirige*, Ibukairi *seregea* (3) Miriam *tereg* (5) Parb, Dungerwab *tol* (6) Peremka *tar*, Bangu, Wandatokwe, Tokwasa *ter*, Mani *ser* (7) Kanum *tor*, Moraori *terogh*, Ngowugar *tar*, Keladdar, Teri-Kawalsch *turra*, Koneraw *cire*

64. tree (2) Bara *oru* 'forest' (3) Kunini *uli*, Miriam *lu* (4) Kawam *orong* 'forest', Dibolug *worung* 'forest', Bugi *lu*, Dabu, Kibuli, Kawam *ro* (5) Parb *wel*, Dungerwab *wöle*, Dapo *uel*

65. two (2) Foraba, Bara, Sesa *tamu*, Ro *tambu* (5) Parb, Dungerwab *tumbi*, Tunjuamu *tsumbi*, Agob *rombi*, Moi-e *arombi*, Dorro *rombi*

66. village (1) Hiwi *aba* (2) Foraba *be*, Sesa *pei*, Ibukairi *be*, 'house' (7) Jelmek *ebi* 'house', Teri-Kawalsch *paya* 'house'

67. to vomit (1) Kiwai, Domori, Wabuda *mamaru* (7) Mombum *memori* (noun), Riantana *morömorö*, Ndom *murmur*

68. wind (1) Goaribari *uroma*, Karami *urami* (2) Sesa *weri* (3) Miriam *rubo* (4) Bugi *wəlam*, Kawam *wele*, Ngamai *wudlam* (5) Dungerwab, Dapo *wəlam*, Tunjuamu *wulam* (6) Tokwasa *wawar*

69. wing (1) Kiwai, Pirupiru *tamu* (3) Oriomo *tame* (4) Bugi *dramba*, Dabu *tama* (5) Dungerwab, Dapo *damb* (6) Bangu *tomba* 'shoulder' (7) Moraori *tomof*

70. woman (3) Miriam *mune* 'vagina', Bugi, Ngamai, Mbayaka *mala*, Kibuli, Kawam, Dabu *mara* (7) Kanum *mele*, Jei *menaw*, Makleu *meing*, Mombum *men* 'vagina'

The grammatical evidence in support of the Southern family does not match the abundance of lexical data. The chief reason for this is, I believe, lack of documentation. Only for groups 1 and 7 is there detailed grammatical information. In groups 2 and 6 only the first and second person singular pronouns are reported and elsewhere our sources are almost as scanty. Between groups 1 and 7 there do not appear to be any specific grammatical common features not found elsewhere. My impression is that Kiwai, from which we have most of our data for group 1, is highly innovating grammatically. To

compare a language like Moraori in group 7 with Kiwai in group 1 is like comparing Icelandic with English or Danish.

There does seem to be common to a number of subgroups, however, a second singular and a third person pronoun. The former has an $m \sim b$ base perhaps going back to *mb-*, actually found in several languages. The latter is also labial but is more likely to originate from an original *p* appearing as *f* and *b* also. *f* is never found in the second person or *mb* in the third.

1. Second person singular. (1) Hibaradai, Hiwi *eme*, Tupati *ma* (2) Barika *amai* (3) Kunini *ma(ne)* (compare first person singular *a(ne)*), Oriomo *ma(n)* (all other personal pronouns end in *-n*), Miriam, Gijara *ma* (4) Bugi *bea*, Kawam *bungo* (5) Dorro, Tunjuamu *bom*, Parb *pom*, Dungerwab *pom, pomo*, Keraki *bom, bomo*, Karigari *bir*, Moi-e *bagai* (6) Mani *mba, mpa* (7) Jei *bu*, Kanum *mbo, mpo*, Makleu *obe*

2. Third person pronoun. (4) Bugi *bo* (singular and plural), Dabu *bo* (singular; plural not recorded) (5) Parb *pe* (singular and plural), Dungerwab *pe* (singular; plural not recorded), Tunjuamu *be* (singular; plural not recorded) (7) Kanum *pi* (singular and plural), Moraori (*ngga*)*fi* (singular only), Mombum *aanggib* (singular only), Ndom *ef*, Jab, Makleu *ib*.

Bordering the large Northern group on the east is a substantial group of languages occupying a large part of the Madang district. The nucleus of such a group is already recognized by Ray in 1919 (K.651). It is here called the Northeastern subgroup or alternatively Madang. The unity of this group is quite obvious so that, in the present study, treatment will be limited to a listing of languages and sources. Only one language, Bongu, has received detailed grammatical analysis but Capell has provided brief grammatical sketches of several of other languages (K.815).

Langtub (Dempwolff K.892), Panim (Dempwolff K.892), Mis (Dempwolff K.892), Bongu (Hanke K.854, K.855), Gorendu (Ray K.651), Bogadjim (Hagen K.849; Hanke K.850, K.854; Ray K.651), Sungumana (Hanke K.854 = Sungum of Ray K.651), Wuong (Hagen K.930; Ray K.651), Wenke (Hagen K.929), Uom (Schmitz 1958), Jimjam (Schmitz 1958), Burumana (Hanke K.854; Ray K.651), Koliku (Hanke K.854; Ray K.651), Kaliko (Ray K.651), Male (Hanke K.854; Ray K.651), Damun (Ray K.651), Shongu (Ray K.651), Banara (Capell K.815 = Moando of Tranel K.846), Tombenam (Zoeller K.839), Dagoi (Schebesta K.859), Bonaputa-Mapu (Schebesta K.853), Bunubun (Capell K.815), Ulingan (Capell K.815), Vanembere (Capell K.815), Bunu (Dempwolff K.858; Ray K.651 = Saker of Kaspruś K.822), Rempin (Dempwolff K.910; Ray K.651 = A'e of Kaspruś K.822), Englam (Ray K.651), Em (Kaspruś K.822), Ate

(Kaspruś K.822), Kemba (Dempwolff K.885; Ray K.651), Bawaipa (Dempwolff K.847), Misdao (Dempwolff K.896), Maragum (Ray K.651), Nupanob (Dempwolff K.903; Ray K.651 = Botelkude of Waugh K.867), Matepi (Waugh K.867).

The most extensive of the New Guinea subgroups is the Central subfamily which extends across much of the interior of New Guinea from the Kapauku of Central West Irian to the Kâte group in the Huon peninsula in northeastern New Guinea. As defined here it is basically equivalent to Wurm's Central and North-East New Guinea Phylum (Wurm 1964). Its nucleus is the Central Highland languages first described in detail in Capell (1948–9, K.814). In my 1958 report I included the languages of the Baliem area of West Irian along with the languages of the Central Highlands. Since then I have come to see that the Kapauku (Ekari-Moni) group belongs with the Baliem languages. The inclusion of the Huon Gulf group of Kâte and related languages was largely stimulated by Wurm's suggestion of a connection of the Highland languages in this direction. It has not apparently been hitherto noted that the languages of the Wantoat region as reported by Schmitz and by Davis are affiliated to this Huon group.

Since Wurm's hypothesis of a North-East Phylum seems to have won general acceptance, no supporting evidence will be presented here. Since, further, Wurm (1964 and elsewhere) gives a detailed classification of the Highlands languages, the languages of this group are not listed here but the publications utilized insofar as they are not already to be found in Klieneberger (1957) are listed in bibliography. The Central group then is divided into three subgroups; Kapauku-Baliem (Western), Highlands (Central) and Huon (Eastern) as follows:

(1) Kapauku-Baliem. (*a*) Kapauku (also called Ekari or Ekagi) (Drabbe K.461, K.462; Doble 1960), Moni (Drabbe K.479; Larson and Gordon 1958; Le Roux K.480), Jabi (Le Rouk K.480), Simori (Le Roux K.480), Wolani (Le Roux K.480) (*b*) Dem (Le Roux K.480) (*c*) Uhunduni (Le Roux K.480), Enggipilu (Colijn K.463; van der Water K.480) (*d*) Dani (also called Ndani) (Le Roux K.480; van der Stap 1966; Bromley 1961, 1967) (*e*) Northern Ngalik (Bromley 1961), Oeringoep (Wirz K.480), Sawuri-Hablifuri (Le Roux K.480), Southern Ngalik (Bromley 1961), Peseghem (Broek K.502; Nouhuys K.503; Ranneft K.504)

(2) Highland Languages. (*a*) Gadsup-Auyana-Awa-Tairora (as in Wurm and Laycock 1961) Erima and Tsinyaji (Schmitz 1958) belong here. (*b*) Gende-Siane-Gahuku-Kamano-Fore (as in Wurm and Laycock 1961) (*c*) Hagen-Wahgi-Jimi-Chumbu (as in Wurm and Laycock 1961)

(*d*) Enga-Huli-Pole-Wiru (as in Wurm and Laycock 1961) (*e*) Karam (Pawley 1966; Biggs 1963 = Aförö of Kirschbaum K.841 (*f*) Kutubu (Capell K.814; Williams 1936), Fasu (Loeweke and May 1965, 1966)[3]

(3) Huon Group. As enumerated in McElhanon (1967) with the addition of Matap (Schmitz 1960), Jupna Valley (Schmitz 1960), Kandomin (Schmitz 1960), Wantoat (Schmitz 1960; Davis 1964)

The Eastern group is the last of the major New Guinea stocks to be considered. It falls into ten clearly defined subgroups as follows.

(1) Mailu group. Mailu (also called Magi) (Saville K.733; Lanyon-Orgill K.733; anon. K.735), Domara (anon. K.682), Nemea (Bastard K.766), Dom (also called Domu) (Bastard K.742), Merani (Bastard K.743), Morawa (Bastard K.750), Magori (Bastard K.729), Binahari (Grist K.666), Monomor (anon. K.751), Keveri (anon. K.710), Moikoidi (Hooper K.747), Bauwaki (Grist K.664), Kororo (Blyth K.721), Neme (Bastard K.765), Boli (Hooper K.669), Doriaidi (Bastard K.683), Buari (Ray K.653), Okaudi (Ray K.653), Bori (Ray K.653), Saroa (Ray K.653), Yabura (Ray K.653), Avini (Ray K.653), Lauwa (Ray K.653), O'oku (Ray K.653), Lauuna (Ray K.653), Gebi (Grist K.690), Orai-iu (Bastard K.771)

(2) Binandere group. Binandere (also called Binandele) (King K.667; K.668; anon. 1914–15), Mambare (anon. K.658), Musa River (anon. K.658), Aiga (Beaver and Chinnery K.662; anon. 1914–15), Yoda (Ray K.636), Berepo (Ray K.636), Amara (Ray K.636), Adaua (Ray K.636), Yema-Yarawe (anon. 1914–15), Mawai (anon. 1914–15), Yega (Beaver K.806), Tain-Daware (Beaver K.794), Jegasa-Sarau (Beaver K.700), Jauwa (also called Dobodura) (Beaver K.678), Hunjara (Beaver K.698), Tsia (Pilhofer K.830), Giumu (Strong K.691), Tahari (anon. 1914–15), Aru (Capell 1954: 70 dialect of Tahari), Duvera (Capell 1954: 70 dialect of Tahari), Maiheari (Skelley K.603), Upper Musa (Ray K.653), Bargua (Ray K.653), Totore (Ray K.653)

(3) Dimuga group. Dimuga (also called Nawp) (Cawley K.675; Grist K.676; anon. K.677), Tevi (Strong K.696), Kanamara (Blyth K.695), Gwoiden (Ray K.635), Makiara (Ray K.635)

(4) Elema group. Elema (Ray K.457 = Haira of Baker K.573), Kairi-Kaura (Strong K.582), Uaripi (Ray K.547), Toaripi (also called Motumotu) (Ray

---

[3] To these subgroups should no doubt be added Duna and Mikaru in accordance with Wurm's published statements. However except for two or three short sentences and lexical items in Wurm 1961*a* and Pike 1964 there is no material in print. In general, with the honorable exceptions of Laycock's Ndu and Doble's dictionary of Kapauku, SIL publications give only incidental vocabulary so that in regard to published material at least the earlier work of Capell and other investigators remains basic in this area for lexical data.

K.547; Thomson K.421; anon. K.624), Milareipi (Ray K.547), Orokolo (Ray K.547; Lawes K.772)

(5) Kovio group. Kovio (Ray K.652; Strong K.724), Kuepa (Muscutt K.723), Oru-Lopiko (Egidi K.795), Kunimaipa (Ray K.652; Pence 1964), Sini (Ray K.652), Biaru (Ray K.652), Goiefu (Ray K.652)

(6) Afoa group. Afoa (Ray K.652; Williamson K.687), Tauata (Egidi K.795), Goilala (Armit K.692; anon. K.693), Ambo (Strong 1918–19), Deba (Ray K.652).

(7) Fuyuge group. Fuyuge (also called Mafulu) (Strong K.686; Jackson K.685; Williamson K.687; Egidi K.795), Kambesi (anon. 1917–18 = Tauada of Ray K.652), Sikube (Giulianetti K.786 = Kabana of Chalmers K.540), Korona (Ray K.652), Onunge (Ray K.562), Agita (Ray K.562), Vovoi (Ray K.652), Gomali (Ray K.562)

(8) Mulaha group. Mulaha (also called Iaibu) (Ray K.549, K.562), Manukolu (also called Lakume) (Ray K.549, K.562), Gaira (Ray K.562), Kwale (Ray K.562)

(9) Koita group. Koita (anon. K.719; Ray K.549), Koiari (anon. K.719), Iworo (Ray K.549), Neneba (Macgregor K.767), Gosisi (also called Tobiri) (Macgregor K.767), Kotoi (Brown K.722), Suku (also called Amaseba) (Ray K.549; Giulianetti K.791), Kagi (Brown K.704), Agi (Ray K.549), Hagari (Ray K.549), Uberi (Ray K.549), Maiari (Ray K.549), Nigubaiba (Brown 1918–19), Barai (Brown K.663 = Managalaski of Parlier 1964), Kokila (Ray K.549), Karukaru (Jackson K.708; anon. K.709), Minjori (Blyth K.745), Suambe (Blyth K.789), Uabari (Ray K.652), Wowonga (Ray K.652), Biagi (Ray K.652), Isurara (Ray K.652), Wamai (Ray K.652), Itu (Ray K.652), Iarumi (Ray K.652), Eikiri (Ray K.652), Favele (Ray K.652), Seramina (Brown K.785), Mogoni (Ray K.652)

(10) Namau (also called Maipua) (Holmes K.611; anon. K.604; Ray K.547)

A list of proposed Eastern etymologies follows:

1. ant (2) Mambare *ciri* (9) Neneba *ciri* (borrowing?)
2. arm (7) Kambisa *ia* 'arm, hand', Korona *ya*, Kambesi *ia* 'hand' (9) Minjori, Suambe *iie*
3. ashes (3) Kanamara *ati* (9) Neneba, Hagari *uti*
4. bad (4) Milareipi *ekapu* (5) Kuepa *kaipi*, Kovio *kepip*
5. bird (2) Binandele, Yema-Yarawe, Mawai, Tain-Daware, Tsia *ni*, Guimu *nei*, Tahari *ne* (3) Tevi, Nawp *neni* (7) Neneba *nea* (8) Manukolu *neni*
6. black (1) Bailu *dabaduba* (2) Musa River *duba* (5) Kovio *dubare* 'dark' (7) Mafulu, Kambesi, Sikube *dube*, Korona, Kambesi *dube*, (8) Manukolu *dobo* (9) Neneba *aduve*, Uberi, Maiari *duduba*, Koita *dubu*

7. blood₁ (1) Neme, Doriaidi *dana* (3) Tevi *deniwa*, Dimuga *denip* (7) Mafulu, Korona, Kambesi *tana*, Sikube *tanara*

8. blood₂ (2) Mambare, Binandele, Yega, Tain-Daware *ororo*, Yema-Yarawe *orara* (4) Orokolo *ioru* (8) Manukolu *ro* (10) Namau *aro*

9. bone (1) Bauwaki *ita*, Mailu *ude, uda* (2) Musa River *etu*, Tsia *wetu*, etc. (4) Lepu, etc. *uti* (9) Nigubaiba *ita*, Barai *adu*, Suambe *ate*

10. breast (1) Bauwaki, Doriaidi, Mailu, Dom, etc. *ama* (2) Mambare *ami*, Yega, etc. *emi* (3) Kwatewa *am*, Nawp *aam* (9) Neneba, Gosisi, Suku, Hagari *amu* (10) Namau *ame*

11. cloud (2) Tahari *uni* (5) Kovio *unida*, Kuepa *units* (7) Mafulu *unu(me)* 'cloud, fog'

12. to cut₁ (7) Mafulu *tode*, Kambesi *doda* (9) Koita *dodo*

13. to cut₂ (4) Toaripi *foi* 'cut down' (7) Karukaru *boi*, Suambe *puwo* (9) Neneba *vuvuoi* 'cut off' (10) Namau *opai* 'hack'

14. to dig (3) Kwatewa *tsia* (4) Milareipi *isai*, Lepu *isa* (7) Korona *tsie* (8) Mulaha *isani*

15. ear (5) Kuepa *gatagapu* (7) Kambisa, Kambesi *gaderu* (9) Nigubaiba, Barai *gada* 'ear, to hear'

16. egg (1) Bauwaki, Neme, Boli, Doriaidi *baka*, Nemea, Dom *baha* (3) Tevi *bagua*, Nawp *bagu* (8) Mulaha *abegi*

*17. eye (1) Mailu, Domara *ini*, Nemea *nii* (9) Suku, Agi, Hagari, Sogeri, Maiari, Koita *ni*, Minjori, Suambe *niie*, Koiari *ni* 'eye, face' (cf. Indo-Pacific 'to see')

18. to fear₁ (4) Milareipi *sia* (9) Koiari *si*

19. to fear₂ (1) Mailu, Domara *dobi*, Magi, Keveri *dobi* 'to frighten' (3) Kanamara *tab*

20. fire (1) Orai-iu *isa* (2) Yema-Yarawe *itu* 'tree' (3) Onjob *itu* 'tree' (6) Oru-Lopiko *iti* (9) Karukaru *idi* 'fire, tree' (10) Namau *iri* 'tree'

21. fruit (6) Tauata *eadauda* 'fruit, flower', Afoa *iadaude* 'flower' (7) Sikube *iudede* (8) Manukolu *ibadade*

22. girl (4) Elema, Orokolo, Milareipi, Toaripi *mori* (7) Kambisa, Sikube *amuri* (9) Sogeri *maoro*

23. to give (1) Mailu *mini*, Domara *miniau* (5) Uaripi *miari* (9) Agi *mairo*, Koiari *minu, miru* (different tense forms), Koita *moi*, Barai *ma*, Nigubaiba *maia*

24. to go (1) Mailu *oni* (2) Kororo *ania*, Binandere *ne* (3) Dimuga *aen*, 'went' (10) Namau *enana*

25. head₁ (1) Giumu *iva*, Tahari *iwa* (3) Dimuga *iwa*, Nawp *ewa* (9) Barai *awo*

26. head$_2$ (1) Gebi *ada* (5) Kovio *ade* 'head hair' (6) Afoa *ade*, Ambo *ate* (7) Mafulu *ade, adede*

27. to hear (1) Domara, Dom, Morawa, Monomor *ope* 'ear' Nemea, Binahari *opi* 'ear' (4) Elema *apai* (8) Manukolu *abi* 'ear' (9) Suki, Agi *evi* Nigubaiba *ife*

28. heart (1) Orai-iu *o-oabai* (2) Musa River *uba* (3) Onjob *bua* (7) Mafulu *ua*(*ne*) (9) Barai *ua*

29. leg (1) Nemea *auqu*, Morawa *auqu, au*, Binahari *auqu*, Domara, Magi *au*, Monomor *au* 'leg' (3) Wadewinda *abu*(*umba*) (10) Namau *au* 'thigh'

30. moon (1) Yabura *inua* (2) Yema-Yarawe *inua*, Tsia *inong* (4) Oru-Lopiko *onea*, Kovio *oneau* (5) Afoa *oani*, Tauata, Ambo, Goilala *one*

31. name (2) Musa River *aviri* (3) Kwatewa *ebura* (7) Mafulu *bode* (9) Uberi *evil, ivila*

32. nose (1) Bauwaki, Keveri *iro*, Neme *ilo* (8) Mulaha *ina* (9) Iworo *uni*, Neneba *udi*, Agi *uli*, Karukara, Gosisi, Kotoi, Suki, Hagari, Uberi, Sogeri, Koita *uri*, Minjori, Suambe *ura*

33. rain (1) Binahari *obani*, Nemea *oban* (3) Kwatewa *uben*, Galela *ubin* (9) Uberi, Agi, Koiari, Koita, etc. *veni*

34. red (1) Moikoidi *kaka* 'blood' (2) Yega *kokoi* (9) Nigubaiba *kakai*

35. root (1) Mailu, Domara, Magori *tai* (2) Binandele, Mawai *tai*, Yema-Yarawe *te*

36. saliva (1) Bauwaki, Neme, Doriaidi, Dom *isubu* (7) Mafulu *sabe* (9) Suku *sabai* 'to spit'

37. to see (1) Mailu *eri* (6) Afoa *ali* 'to know' (7) Mafulu *ali, ari* (9) Hagari *elea*

38. to sit (1) Mailu, Domara *auri* (2) Musa River *auiri* 'remain', Tahari *orari* 'remain' (3) Onjob *auro*

39. sky (2) Binandele, Aiga, Yema-Yarawe *utu*, Yega, Tain-Daware, Jegasa-Sarau *utu* 'sky, cloud', Adaua *utu* (9) Neneba, Karukaru *oto*

40. star (4) Uaripi, Milareipi, Toaripi, Lepu *koru*, Elema, Orokolo *kou* (9) Koiari *kolo, koro*, Maiari *koro*

41. to steal (1) Mailu *tere* (2) Mambare *tur* 'to pluck' (4) Toaripi, Lepu *torea*, Elema *korea*

42. to take$_1$ (3) Dimuga *wat* 'to take, have, get' (5) Oru-Lopiko *wade*, Kuepa *wada* 'touch'

43. to take$_2$ (6) Tauata *mi* (9) Kotoi *ma* (10) Namau *miai* 'to take away'

44. to taste (7) Mafulu *tovogi* (9) Suku *teve*

45. wet (1) Orai-iu, Gebi *oru* 'water' (10) Namau *oru*

46. wife (7) Mafulu *amule*, Kambesi *amera*, Sikube *amuri* 'woman' (9) Iworo, Neneba *amuro*

47. word (4) Toaripi *o* (7) Kambesi *wa* 'thing' (9) Koita *uo*

Relatively little grammatical information has been published regarding languages of the Eastern New Guinea subfamily. Within this limited material nothing has been noted as common to the Eastern languages beyond what is generally Indo-Pacific except once again for the pronouns concerning which we have a basis for comparison in a substantial number of languages. Although not uniquely Eastern, it is probably worth pointing out that seven of the ten subgroups have a first person singular *na* and seven have a second person singular *ni* or *nu*. There is an especially close relationship here between subgroups 6 and 7 as shown in (6) Afoa, Tauata *nui* (7) Mafulu, Kambisa *nu*, Mafulu, Sikube *nuni*. Although *na* and *ni* are both very widespread in Indo-Pacific the Eastern group does distinguish itself from neighboring subgroups by the presence and extent of these two pronominal sets. In addition, the following pronoun forms may be cited as common to a number of subgroups of the Eastern group and either unique or at least of very limited distribution elsewhere.

1. Third person singular pronoun. (2) Yema-Yarawe *o* (7) Mafulu *u*, Kambisa *u, hu* (8) Mulaha *o*, Manukolu *oi* (9) Uberi, Koiari (Meroke dialect of Chalmers) *oe* (10) Namau *u*

2. Third person pronoun. (1) Mailu *oma* 'they' (2) Aiga *omo* 'he, she', Jauwa, Hunjara *emo* 'he, she' (3) Dimuga *me* 'he, she', Onjob *mu, ma* 'they' (6) Afoa *ome* 'he, she'

3. First person plural pronoun. (5) Oru-Lopiko *dae* (6) Mafulu *di* (7) Mulaha *-di* 'our'

4. Second person plural pronoun. (1) Mailu *aia* (3) Onjob *ya* (6) Mafulu *yi* (9) Uberi *ia*, Koita *ya*, Koiari *ya-* 'your'

The foregoing enumeration and discussion of the seven major subgroups of New Guinea non-Austronesian languages has omitted the mention of a number of individual languages or groups of languages, some of them substantial. A number of these have been definitely assigned to one or other of the seven subfamilies but too late to be included in the lexical or grammatical comparisons. This additional evidence will be presented in an expanded version of the present study.

As before the discussion will proceed in a roughly west-to-east order. First to be discussed is the language now called Yava by Anceaux (1961) and spoken in the central part of the island of Yapen north of Geelvinck Bay. Data on Yava has been cited under the name Mantembu, one of its dialects, by Cowan in several publications (K.432, 1960). There is also a short word list of Saweroe in Anceaux's unpublished materials (A). Saweroe forms the second main

division within Yava as against all the other dialects including Mantembu (Anceaux 1961). Cowan has presented evidence for the inclusion of Yava within his West Papuan Phylum which consists essentially of the Timor-Alor, Halmahera, and Western New Guinea families in the classification presented here.

All of the material on Yava presented in these works is fragmentary and also, for obvious reasons in Cowan's case, not a random sample since only resemblances to his West Papuan Phylum are presented. The additional material of Anceaux, although sparse, is not selected in this fashion. It tends to support the hypothesis that Yava should be assigned to the Western New Guinea group constituting a separate subgroup within it.

Of the languages of the Bomberai peninsula discussed in Anceaux (1958), three were omitted in the foregoing review. For all of these I have had access to additional material from (A). One of these is Tanahmerah, here called Tanahmerah I to distinguish it from Tanahmerah II of the Northern group, the latter a language closely related to Sentani. I have not been able to classify Tanahmerah I. The other two, Asienara and Iria, are closely related to each other. It has not been previously noted that these two languages are to be connected with the Kamoro group to the east from which they are separated both by the Etna Bay subgroup of the Western New Guinea subfamily and by Irahutu, an AN language.

This brings up the status of the Kamoro group which includes besides the aforementioned Asienara and Iria, Kamoro (Drabbe 1953), Mimika (Dumas K.468), Nagramadu (van der Sande K.482), Angadi (van der Sande K.454), Sempan (Drabbe 1953), Asmat (Drabbe 1953, 1963; Voorhoeve 1965) and Kajakaja (Feuilletau de Bruijn K.466). In my 1958 report Kamoro was included with the Western New Guinea group of the present study under the name Vogelkop-Kamoro. Kamoro is tentatively separated from the larger grouping pending a reassessment of the evidence. Another language omitted from this study is Arare spoken on the Juliana River in West Irian and only known to me from the unpublished word list in (A). This language shows important similarities both with the Kamoro group and with the Western subgroup. It thus tends to strengthen the probability that the Kamoro languages will ultimately be shown to have a special relationship to the Western group.

Situated between the Western and Northern families are a number of languages in the Momberamo River system of north central West Irian. These languages fall into two distinct but possibly related groups as follows: (1) Kauwerawet (Le Roux K.471), Koassa (anon. K.451); (2) Tori (anon. K.451), Tori Aikwakai (Feuilletau de Bruijn K.517), Borumessu (anon. K.451), Sidjuai

(anon. K.451), Pauwi I (anon. K.451), Pauwi II (van der Aa K.501), South River (called Südfluss in anon. K.451). The language of the Goliath Mountain Pygmies although scantily represented in de Kock (1912) shows a small number of striking agreements with the first of the two above groups. There are some indications (e.g. the Tori Aikwakai word for 'water' *kwaru* which resembles a widespread Western New Guinea and indeed Halmahera etymon) that these languages will eventually link up with the Western New Guinea subfamily.

Much farther east in the Sepik River basin of central northern New Guinea, Laycock (1965a) distinguishes a group of closely related languages to which he gives the name Ndu from the common word for 'man' in these languages. Laycock gives detailed linguistic information on the following languages or dialects, Maprik, Wosera, Nyaura, Manambu, Ngala, Kwusaun, Yengoru, and Yelogu. The first two of these are dialects of what is usually called Abelam for which there is additional scattered linguistic information in the ethnological articles of Kaberry (1941, 1941–42). A vocabulary in anon. K.837: 128 labelled 'Middle Section' (i.e. of the Sepik River), Tombenam to Malu also belongs here. Nyaura is a dialect of Iatmul for which some lexical material is given in Bateson (1936).

Laycock (1965a, 1965b) presents data for a certain number of non-Ndu languages of this area. Among these Kwoma, Mayo, and Wongamusin appear to form a group. A very small amount of additional linguistic information regarding Kwoma may be gleaned from the ethnological works of Whiting (1941) and Whiting and Reed (1938). More importantly for present purposes, Capell (1954: 18) gives a full set of Kwoma pronouns including forms differentiated for sex in the second and third person from which it appears that Laycock's second and third singular pronouns are the masculine forms and that he was either reporting a different dialect without this distinction or he neglected to give the feminine forms. At any rate, as noted already by Capell, there is striking agreement with the pronouns of the Ndu. There is also some vocabulary resemblance even within the small amount of lexical material available on the Kwoma group. One may therefore speak at least tentatively of an Ndu-Kwoma family. Further, Ndu-Kwoma is to be considered an additional branch of the Northern New Guinea subgroup. In the general Indo-Pacific word list at the end of this study, forms cited from the Ndu-Kwoma languages are included with the Northern subgroup. Capell's data on Kwoma pronouns adds welcome additional evidence in the form of the third person plural pronoun *mi*. It has been seen that third person plurals in *m* are highly characteristic of the Northern New Guinea subfamily.

On the basis of data in Laycock (1965*a*, 1965*b*) and E. Pike (1964), Iwam and Abau are affiliated with the Ndu-Kwoma group probably as a third branch.

Unpublished excerpts from additional languages of the Sepik district were made available to me in the form of short word lists through the courtesy of the Summer Institute of Linguistics. Some of these are readily classifiable and are listed in their appropriate places in the body of the present paper but a few from Aitape and Amanab subdistricts near the border of West Irian are still unplaced. These are Kwanga, Urim, Yuri, Fas, Busa, Amto, Senagi, and Komberatoro. The latter two are evidently related to each other. Kwanga is the same language as Capell's Womsak (1954: 14) for which he gives three words, two of which figure in the unpublished SIL list and are virtually identical in form. All of these languages may be suspected of affiliation in some fashion with the Northern subgroup.

Of the remaining languages not assigned to any of the major subgroups the most important are the Gogodala (also called Gogodara) languages of south central New Guinea. These include Gogodala (Riley K.550), Gaima (Ray K.547), Girara (Ray K.547), Adiba (Riley K.550), and Waruna (Riley K.550). There are strong indications that the Gogodala languages form an additional subgroup of the Southern New Guinea (Kiwaiic) family. Farther east the following are unclassified: Tate (Strong K.622), Williams River (Chisholm 1914–15), Ondoro (Blyth K.769), Aurama (Brown K.560), Huaruha (Murray K.576), Mumeng I (Vicedom K.900), Mumeng II (Capell K.814: 355–6). Mumeng I and II appear to be distinct but closely related languages reported under the same name. Aurama and Huaruha are closely related.

Finally the dialects of Rossel Island in the Lousiade Archipelago east of New Guinea are probably to be considered a single language. The sources are Macgregor (K.809), Armstrong (K.808), Winter (K.811), and Ray (K.810). There are a few special resemblances to the Central Melanesian subgroup, particularly the characteristic second person plural pronoun *mi*, but the evidence does not seem conclusive for its assignment to this or any other subgroup. It should be remarked that of the languages mentioned here as unassigned to any subgroup, all show a number of lexical items which are either generally Indo-Pacific or occur widely in one or more of the New Guinea subgroups.

Having reviewed the subgroups of Indo-Pacific both outside of and in New Guinea, we now proceed to a consideration of some of the lexical and grammatical features which link these groups together into an overall entity. The discussion is by no means complete, since in regard to grammatical features it is largely limited to features of the pronominal and gender systems and for lexicon almost entirely to those etymologies which occur in at least

three subgroups, one at least of which is outside of New Guinea. Since both in the following grammatical discussion and in the presentation of the lexical evidence there is constant reference to the various subgroups as enumerated earlier in this paper, they are referred to by abbreviations which will be obvious for the most part, but which it seems best to enumerate here: AN (Andaman Islands), TA (Timor-Alor), HA (Halmahera), NB (New Britain), BO (Bougainville), CM (Central Melanesia), TS (Tasmania), WNG (Western New Guinea), NNG (Northern New Guinea), SWNG (South West New Guinea), SNG (Southern New Guinea), CNG (Central New Guinea), NENG (North East New Guinea), ENG (Eastern New Guinea). The languages not assigned to any of these groups are indicated by UNG (Unclassified New Guinea).

The grammatical evidence is presented under eleven headings.

1. First person singular pronouns. There are two very widely distributed types of first person pronouns which will be called here the *n* type and the *t* type. The geographical distribution of these two types among the subgroups will be sketched without citing every individual instance since these are very numerous. Then the relation of the two to each other and their place in certain pronominal patterns involving indicators of other than first person singular will be considered.

The *n* first person pronouns are found outside of New Guinea in the following subgroups: TA, NB, BO, and CM. In TA these are the ordinary forms in all four languages: Oirata *an*, Makasai *ani*, Bunak *ne*(*to*), Abui *ne*(*do*). In New Britain most languages have *ngg* forms but Uasi, which, has been indicated, may be more closely related to Bougainville languages, has *eni*. In the BO group these are the common first person singular forms, e.g. Telei *ne*, Nasioi, and Koromira *n-* 'my'. In CM Nea and Nabalue in Santa Cruz have *ni* but since for these languages *ni* occurs initially in all the pronouns these are perhaps to be analyzed as the general pronominal base plus zero. The other languages of Santa Cruz as well as those of the New Georgia archipelago have *ng* forms except that Savo has a first personal singular object suffix on the verb *-ni*. Since *-i* is common to all these objective forms the base here is presumably *-n*.

On New Guinea WNG has both *n* and *t* forms in different subgroups. In general the Konda-Jahadian group and the Kapaur group have *n*. Examples are Kampong-Baru *ne*(*ri*) (cf. *e*(*ri*) 'thou'), Tarof *ne*(*iga*) (cf. *a*(*iga*) 'thou'), Karas *aan*, Kapaur *On*.

In NNG *ngg* and *k* forms predominate as indicated earlier in the discussion of that group. However in the Ndu-Kwoma subgroup the *n* forms are the common ones in all the Ndu languages, e.g. Maprik *unə* while Kwoma and

Mayo have *an*. There is also the Sko (Tami subgroup) possessive prefix *ne-*. In SWNG *n* forms are the common ones found in all the languages for which we have data. In SNG *n* forms are the normal ones in the western subgroup, e.g. Kimaghama *no*, Riantana *na*, Ndom *ne*. It reappears again the easternmost group where some of the Kiwaiian languages show it, e.g. Hibaradai, Hiwai *na* although most of the subgroup has pronouns like Kiwai *mo*. The Gogodala group which probably affiliates with SNG also has *n* forms. In the vast CNG group *n* is the normal form in virtually every subgroup except for some of the Tairora subgroup of the Central Highlands which have *t*. In contrast I know of no sure case in NENG, which is a stronghold of *t* forms. As indicated earlier *n* forms of the first person pronoun occur in seven of the ten subgroups of ENG. Finally among the unclassified languages there is Huaruha *ano* and *n* forms on Rossel Island. In summary, first person pronoun forms with *n* occur in four of the six extra NG groups and completely predominate in two of these and occur in all of the seven New Guinea subgroup except NENG and are the predominant or even exclusive form in three of these.

The northern subgroup of AN, which embraces all but the southernmost language Önge, has *t* forms. An example is Chariar *tio* 'I' when compared to *ngio* 'thou'. In HA *t* forms are found in all the languages. The normal subject form is *t(o)* in which *o* is a subject indicator. An example is Galela *t(o)* 'I' (cf. *n(o)* 'thou'). The object forms are generally *j(i)* in which *(i)* is an object indicator and *j* is probably a palatalized form of *d*. Similar variations are found in WNG whose *t* and *d* forms have already been connected with the Halmahera pronouns by Cowan.

Roughly half the WNG languages have such *t* and *d* forms while the rest have *n*. Examples are Manikion *tani* (cf. *ban* 'thou'), Madik *ji*, Waipu *tit* (cf. *nin* 'thou'). For the remainder of New Guinea, however, *t* forms are far less frequent than those in *n*. There are however three areas of substantial occurrence; one of these is the Central Highlands subgroup of CNG especially the Tairora languages, e.g. Gadsup *te* but also Benabena -*te* 'my' in a different subgroup. Another area of dominance which has been alluded to earlier is NENG. Palatalized variants occur as in HA and WNG. Examples are Mis *da*, Bongu *aji*, Nupanob *ta, da, ita*. The third *t* area in New Guinea is the Koita subgroup of ENG in which it predominates although some languages have *n* forms. Examples here are Koita, Uberi, etc. *da*. In the same languages the possessive is a prefixed *di-*. Finally Paremka of CNG has *tea* 'I'.

Both of these alternatives are as can be seen of enormously wide distribution, *n* being somewhat more common. Presumably they should both be original. There are, however, three languages in which both *n* and *t* are known to occur and it is, I believe, very strong evidence for the general Indo-Pacific

hypothesis that the division of function is analogous in all of them. It has been already noted that in Makasai on Timor *ani* is the absolute pronoun used in all other uses and *asi* is the possessive with *s* < *t*. In Baham (Patimuni) of the Kapaur subgroup of WNG the absolute pronoun is variously recorded as *antu* or *andu* (cf. Kapaur *ɔn*) while the possessive pronoun is a bound prefix *ta-*. In Benabena of the Central Highlands of New Guinea (Capell's Hofaga) the independent pronoun is *nani* whereas the possessive is a suffix *-te*.

The first person *n* pronoun participates over a very large part of New Guinea in an opposition to second person *k* usually in the form *na*/*ka*, a pattern to be discussed later separately. The vowel following *n* is probably most frequently *a*. In this form it often contrasts with the widely found first person plural *ni*. The number contrast *na*/*ni* has in some instances been extended analogically to the second person. A good example is the 'bene-factive' (dative) set for Moni (western subgroup of CNG), *na* 'for me', *ni* 'for us', *ka* 'for thee', *ki* 'for you'.

2. Where the *na*/*ka* pattern does not dominate, the most common second person singular pronoun is *ngi* or *ni*. I suspect that *ngi* is original and has frequently become *ni* either by direct phonetic change or under the influence of first person singular *n*. In a few cases analogy has probably worked in the opposite direction. This would perhaps explain the *ng* first person forms in NB and CM. An example is Baining (NB) *ngoa, ngu* 'I', *ngi* 'thou'.

The following citations will give some idea of the extent of these second person forms in *n* or *ng* commonly but not by any means always followed by *i*. In some instances the base is simply *n* or *ng*. AN Biada *ngol* (cf. *dol* 'I'), Kede *ngui* (cf. *tui* 'I'). Önge of Little Andaman in the south has *ngi* (but first person singular *mi*), one of the few strong pieces of evidence for its affiliation to Indo-Pacific; HA Galela and other languages *no* 'subject', *ni* 'object', *ngona* 'independent pronoun'. Compare the first persons *to, ji, ngohi* (Loda *ngoji*); NB Baining *ngi*, Taulil *nggi, ngginggi*, Uasi *nini*; CM Savo *no*, Laumbe *inu*, bound form *ngo*, Bilua *ngo*, Baniata *no*. The *o* vowel is distinctive for the Central Solomon subgroup of CM. Santa Cruz has *m*; TS all dialects *ni*(*na*) (cf. *mi*(*na*) 'I'). WNG Amberbaken, Madik, Karon *nan*, Waipu, Moi, etc. *nin*, Aitinjo, etc. *nio*. Patimuni *na-*, Mairasi *ne-* (possessives); NNG Makarob *nɔ*, Tanggum *nu*, Murusapa *na*, Anaberg *nɛ*, Atemble *nə*, Kavu *na*(*k*), Mayo *nə*. In general *m* forms are more frequent in NNG. SWNG, not found, exclusive predominance of pronouns of *na*/*ka* pattern; SNG *m*/*b* forms or *na*/*ka*; CNG, *na*/*ka* pattern is also general in this subgroup but *n* forms are found in the Central Highland, e.g. Gadsup *en*, Binumarien *ane*, Chimbu *ene*, Dom, Sinasina, Tjuave, Sua *ne*, Karam *nant* (cf. *yant* 'I'), Fasu *ne, ni*, Kewa *ne*(*me*). NENG, forms in *n*, particularly *ni* almost exclusively found, e.g. Langtub, Bongu, Saker,

Sungumana, Burumana, etc. *ni*, Banara *no*, Vanembere *ne*, Nupanob *na*, etc.; ENG seven of ten subgroups have *n* forms, e.g. Binandere subgroup Hunjara *ni*(*mo*) (cf. *na*(*mo*) 'I'), Yema-Yarawe, Giumu, Tahari *ni*; Kovio group Kovio, Oru-Lopiko *ni*, Kuepa *ne*; Afoa group Afoa *nu, nui*, Tauata *nui*; Mafulu group Mafulu, Kambisa *nu*; Namau *ni*. Many of the ENG languages show the *na/ni* contrast between first and second person singular. Rossel Island (UNG), e.g. Yela, shows this also *na, nga* 'I', *nyi* 'thou'.

3. First person plural pronoun *ni*. TA Oirata *in*(*te*), Makasai *ini*, Bunak *nei*, Abui *ni*. All these (including presumably Abui for which simply two variants *ni* and *pi* are given) are exclusive; HA object inclusive forms are *na* in almost all the languages; BO Telei, Nasioi *nii*; *ni* 'our' WNG Solowat, Itigo, etc. *ni*(*ti*) exclusive as in TA (cf. *dai*(*ti*) inclusive); NNG Anaberg *ni*, Tanggum *nai*, Sko, Sangke *ne*, Ndu group *nanə, nani, nan*, Kwoma *no*(*ta*) (cf. *ko*(*ta*) 'you'); SWNG Boazi *ni*, Upper Muju *ne*, Telefol *nu*(*ta*). In SWNG we often find singular first person *n* forms with plural suffix used on all pronouns, e.g. Awju Pisa *nu* 'I', *nugu* 'we' SNG Kanum, Koneraw, Ndom, etc. *ni* (western subgroup, elsewhere almost no data except Kiwaian proper which usually has *nimo*); CNG Ndani *ni*(*t*), Ekari *ni*, Sau *ni*(*gi*), Chimbu *no*, Dom *ne*, Hube *nini* (exclusive) etc.; NENG, not found; ENG Dimuga, Onjob *nu*, Jimajima *no*; Tauata *nane*; Mulaha *nai* (inclusive); Koita *no, ni-ni* 'our', Koiari *ni-* 'our'. Rossel Island (UNG) *nu-* 'our'. Note that no Tasmanian pronouns are recorded outside of the first and second person singular.

4. First plural inclusive pronoun. TA and HA agree in having an inclusive pronoun in *⋆p*. TA Oirata *ap-* 'our incl.' Makasai *fi*, Alor *pi*. HA Modole, etc. *po* (subject pronoun).

5. Third person plural pronoun. TA Abui *da-*, *de- de*(*ning*); NB Taulil, Butam, Sulka *ta*; CM Savo *ze*, Bilua *se* (cf. Savo *izi*, Baniata *isia* 'to sleep' with the common *t* and *d* forms elsewhere, e.g. Siwai *at*, Galela *idu*); NNG Sko *tea*, Sangke *te*; Ndu group [*n*]*di*, [*n*]*dʌy*, etc.; CNG Dem *ta*, Hoiyevia, Tarifururo *ti*, Matap *ita*; ENG Yema-Yarawe, Mawai *eto*, Tauata *ote*. In the Gogodala group probably affiliated with SNG we find Waruna, Gogodala *de*, Gaima *da*. Perhaps also the -*t* suffixed to the class prefixes to form the plural in the North Andaman group belongs here.

Although of rather limited distribution this is probably the distinctive Indo-Pacific third person plural pronoun. It is in competition with the *y* singular pronoun often used with a pluralizer or common to both numbers as well as demonstratives often used as third person pronouns without distinction of number.

6. The suffix pronominal pattern (SPP). By SPP is meant here a specific pattern usually suffixed to the verb to indicate the pronominal subject. It is

widely found in the far-flung CNG group. Pawley (1966) has already noted its existence in Karam and gives a comparative table (p. 198) of Karam and a number of languages in the East Central Highlands group. However it can also be traced in the western (Baliem) and eastern (Huon) subgroups of CNG, in SNG, in NNG, and in the Kamoro group (possibly affiliated to WNG). Its characteristic features are first person singular *u, w*, second person singular and first person plural nasal consonant (usually *n*), and third person singular *i, y*. The second and third person plural are usually identical and in CNG at least possibly go back to **aw*. In addition, many of the CNG languages have a dual.

The subject suffixes for the primary verb (i.e. one that always ends a sentence and is never subordinate) in Kanite, a Central Highland language of CNG, may serve as an example.

	SINGULAR	DUAL	PLURAL
1p.	-u	-uq	-un
2p.	-an	-aq	-a
3p.	-i	-aq	-a

This whole pattern doubtless takes its start with the Indo-Pacific *n* pronouns of the second person singular and first person plural already discussed and the common *i, e* or *y* third person pronoun which also figures in the *na/ka/i* prefix pattern which will be discussed later. The first person *-u* is, however, unique and a virtually constant feature.

A good many other instances of this pattern from Central Highlands languages might be cited but this is unnecessary since a number of them are already found in the work of Pawley already cited.

A number of examples are found in languages of the eastern subgroup of CNG, i.e. the Kâte or Huon Gulf group. Naturally enough one or other characteristics of the pattern may be absent in specific instances. The following are the subject suffixes of the perfect tense in Deduae of the Huon Gulf subgroup of CNG.

	SINGULAR	DUAL	PLURAL
1p.	-dua	-diq	-ding
2p.	-dang	-daoq	-dau
3p.	-daq	-daoq	-dau

In this paradigm, *d* is, of course, the tense marker. Note the *u* of the first person and the nasal *ng* in the second person singular and first person plural. In the Huon Gulf languages the *-y* third person singular is apparently not

found although *y* is prominent as the basis of independent third person pronoun. On the other hand we find -*au* in the second and third person plural which may be compared to Fore -*aw*, likewise second and third plural.

Fore is a Central Highlands language. Note also the agreement between Deduae and Kanite in the final glottal stop (*q*) in the dual. Still within CNG what is no doubt the same pattern in attenuated form is encountered in the western (Baliem) subgroup. Van der Stap (1966) in his description of Dani gives the following set as the normal suffixes on the verb to express the person-number category of the subject:

	SINGULAR	PLURAL
1p.	-*y*	-*o*
2p.	-*en*	-*ep*
3p.	-*e*	-*a*

The phonetic value of *y* is not clear. It seems to be [ɪ]. Whether it comes from *u* I am unable to say. Second singular *en* is normal and third singular *e* may well represent *i. The absence of a form with nasal consonant for the plural first person is a real deviation but third person plural *a* is quite normal and the most common form in the Central Highland languages.

In SNG, as has been earlier noted, except for Kiwai proper, our only solid grammatical data are from languages of the westernmost subgroup. Here SSP can be found in Jei (prefixed to indicate object), Kanum (prefixed to indicate object) and Mombum (suffixed to indicate subject). Hence in Mombum its position and syntactic functions are like those of CNG, whereas in the first two languages it differs. The forms in these three languages and their basic agreement with the CNG examples can be seen from the following table.

	JEI	KANUM	MOMBUM
1s.	*wo*	*u*	*u*
2s.	*ne*	*n*	*im*
3s.	*ye* (masc.)	*i*	*i*
	gwe (fem.)		
1p.	*ne*	*n*	*om*
2p.	*ye*	*i*	*om*
3p.	*ye*	*i*	*a*

Finally it occurs in Nimboran in NNG. It functions here in the normal way as a suffixed subject indicator in the verb. However, here we have only a single set which in unmarked form functions in the singular and with the additional presence of *ke* 'dual' or *i* 'plural' added to the verb stem expresses the dual and plural respectively. The unmarked singular forms are first person *u*,

second person *e*, third person: masculine *am*; feminine *um*. In the third person -*m* is most probably the widespread third person *m* of the NNG while, as will be shown later, the vowel alternation for gender represents another widespread pattern. The first person *u*, the coincidence of function and position and the existence of satisfactory explanations for the third person suggesting fairly strongly that this is still another manifestation of the same underlying pattern, SSP, which to my knowledge does not occur outside of New Guinea, serves thus to link three major groups, NNG, CNG, and SNG.

7. By PPP (pronominal prefix pattern) will be indicated a pattern of even broader distribution though still confined to New Guinea. It may appear as a set of independent pronouns but, most characteristically it occurs prefixed to the noun to indicate possession and/or to the verb to express the pronominal object. With regard to the former, there are frequent instances in which it is confined to kin terms, while other nouns express possession analytically by preposing an independent pronoun in accordance with the usual Indo-Pacific order of possessor–possessed or preposing the pronoun followed by a genitive particle. In such instances all non-kin nouns are then expressing pronominal and nominal possession by the same construction.

When prefixed to kin terms there are commonly only three forms used for all numbers. The first and second person forms are *na-* and *ka-* or *ga-* respectively or in a few languages *na-* and *ha-* where *h* can in every case be shown plausibly to originate from *k*. The third person is most frequently *i-*, an obvious point of contact with SSP and the common third person pronoun element *i, y*.

An example is Marind in SWNG where the possessive construction for non-kin terms may be illustrated as follows: *anim end igiz* 'man of name', i.e. man's name; *oh end igiz* 'you of name', i.e. your name. In contrast we find *na-zeb*, 'my, our grandchild', *ha-zeb*, 'thy, your grandchild', *i-zeb* 'his, her, their grandchild'. It also is notable that a general third person prefix undifferentiated for gender occurs in languages like Marind which possesses sex gender in other constructions.

The other chief occurrence is as prefixed, or infixed object in the verb complex. Here number is usually distinguished by a separate plural set while PPP functions for the singular. Often the specific verb which can occur with PPP is a limited and small subset of the verbs. Various other indicators occur in third person, often several different ones in the same language with different verbs. PPP occurs in WNG (subject to some qualification), SWNG, SNG, CNG, and ENG, as well as the unclassified Tanahmerah I.

In WNG, the occurrence of PPP is marginal. In the Kapaur subgroup the first and second person singular pronouns show *n* and *k* respectively, e.g.

Kapaur *on* 'I', *ko* 'thou', and Karas *aan* 'I', *ka(me)* 'thou' cf. *ma(me)* 'he, she, it '. Little is known regarding the grammar of these languages but there are probably no pronominal bound forms. In the Etna Bay group we have a possessive prefix *ka-* 'thy' for Faranjao and the available word list often shows a prefix *ne-* which may mean 'my'. If so we would have the usual possessive prefixes at least for the first and second person.

The three subgroups of SWNG for which we have detailed information, Marind, Ok, and Awju all show this pattern in the languages for which relevant data exist. An example is Marind from which an example of the possessive use with kinship terms has already been quoted. A number of verbs express the pronominal object by prefixing or infixing other variants of this set differentiated for number. The following will illustrate a common pattern of infixing used with certain formative suffixes, e.g. *-ib, kah-ib* 'to bind', *kaha-n-ib* 'to bind me or us', *kaha-h-ib* 'to bind thee', *kaha-z-ib* 'to bind you or them', *kah-ib* 'to bind him, her, or it'. We may suspect that *-ib* is an old auxiliary with prefixed object. An example of prefixing is *na-sak* 'to hit me or us', *ha-sak* 'to hit thee', *u-sak* 'to hit him, her, it', *i-sak* 'to hit you or them'. *u* or *w* is one of the common third person singular variants found in a number of languages.

In SNG the pattern is attested once more in the westernmost subgroup where we have detailed grammatical information. Here it does not occur in bound forms; an example is the subject-independent pronouns of Riantana *na* 'I', [*ng*]*gö* 'thou', *yö* 'he, she, it'.

In CNG, the PPP type is heavily represented in all three main branches, western, central, and eastern. We may illustrate from Dani in the western group. In this language as in Marind and elsewhere the employment of the PPP set as possessives with nouns is confined to a limited set, in this case kin-terms, parts of the body, and personal belongings. The following is an example: *n-eilegen* 'my eyes', *h-eilegen* 'thy eyes', *eilegen* 'his, her eyes', *nin-eilegen* 'our eyes', *hin-eilegen* 'your eyes', *in-eilegen* 'their eyes'. The same set is usually infixed in the verb in different forms for the direct or indirect object e.g. *sal-na-p-in* 'to cover me', *sal-ha-p-in* 'to cover thee', etc., in which *-p-* marks the infix as a direct object and *-in* is the infinitive suffix and *isat-ne-s-in* 'to cook for me', *isat-he-s-in* 'to cook for thee' etc. in which *-s-* marks the dative. As in Marind there is a limited group of verbs which take prefixes: *na-s-in* 'to hit me', *ha-s-in* 'to hit thee', *wa-s-in* 'to hit him', *nina-s-in* 'to hit us', *hina-s-in* 'to hit you', *ina-s-in* 'to hit them'.

Finally in ENG one subgroup, Dimuga, has independent (presumably) pronouns of this type in the first and second person singular, e.g. Dimuga *ne*,

ge; Onjob *na, ga*. Virtually nothing is known of the grammatical structure of these languages.

8. Another pronominal pattern found in languages of the three subgroups SWNG, SNG, and CNG will be illustrated by a verb paradigm from Magobineng of the Huon (eastern) subgroup of CNG, in this case that of a tense labelled Perfect I, with the following suffixes indicating the subject:

	SINGULAR	DUAL	PLURAL
1p.	*-baq*	*-baleq*	*-baneng*
2p.	*-maq*	*-bileq*	*-bineng*
3p.	*-yeq*	*-bileq*	*-bineng*

The significant points which define this pattern are: (1) in the non-singular number or numbers the second and third persons are identical; (2) the first person is distinguished from the non-first persons by a vowel change which is the same for dual and plural if there is a dual.

The vowels involved may be different even in the same language in different tenses. We may symbolize the pattern by the vowel of the first person followed by the vowel of the non-first person, e.g. *a/i* in the above instance. An example of two patterns in the same language is Kate *e/i* and *a/i* in different tenses. The recurrent types are *a/i, e/i, u/i* and *i/e* (i.e. those which occur in more than one language). This pattern is found in a number of languages in the central and eastern branches of CNG. It is also found in Kati of the Ok subgroup of SWNG and in Jei and Kanum of the western subgroup of SNG.

9. Unlike the preceding three items, this one is found both in New Guinea and elsewhere. Abui in TA has a plural *-mana* as in *amui-mana* 'the dead'. Moni in the western subgroup of CNG has plural suffix *-mena* used with relationship terms and Binandere of ENG has two plurals *-mono* and *-mane* of which the former is exclusively used with relationship terms and the latter for the most part. These resemblances might not appear to be particularly significant since they are found in so few languages. However, when one considers that the overwhelming majority of Indo-Pacific languages do not inflect the noun for number at all, or have only a few special, often irregular plurals for a few words such as 'man', 'woman', 'boy', etc., these agreements become noteworthy.

10. A major point of agreement linking a whole series of groups both in and out of New Guinea is the expression of sex gender. The most important pattern once more involves vowel alternations and has received incidental illustration several times in the course of this study. A convenient starting point is the system as it operates in Marind of the SWNG group. As in a number of other instances there are other genders besides masculine and feminine, in this case two others. The chief manifestation of gender is by vowel

change in nominal modifiers, the noun itself usually not having any overt gender mark. The vowel alternation may affect, depending on the specific modifier, a medial or final vowel or both. The four genders are (1) masculine human, (2) feminine human and animals, (3) inanimate, (4) inanimate. Only genders (1) and (2) have a separate plural and this plural is the same for both and always coincides in formal expression with gender (4), the inanimate gender *par excellence.*

The basic and most frequent pattern is: *e* masculine singular, *u* feminine singular, *a* first inanimate gender, *i* second inanimate and plural of masculine and feminine. An example is the adjective 'light' which has the inflections *akek, akuk, akak,* and *akik* in accordance with the above scheme. A few illustrative sentences are: *patur epe papes ka* 'young-man the small is', *kivasom upe papus ka* 'young-woman the small is'. The same pattern appears in a few verb forms which mark the gender of the subject, e.g. *epiziget* 'he sleeps', *upiziget* 'she sleeps', *ipnaiziget* 'they sleep'. In this last *na* is a common marker of third person plural subject.

A further manifestation is found in some nouns in which vowel alternations express sex differences. These are chiefly relationship terms. An example is *anem* 'man', *anum* 'woman', *anim* 'people'.

Although the vowel pattern thus far illustrated is the basic one, certain variations are found. For example *a* instead of *e* for the masculine singular appears in the adjective 'wild' *akhata* masculine, *akhatu* feminine, *akhati* inanimate and human plural and in the pair *zam* 'husband', *zum* 'wife'. Another variant is masculine *i* as in *wananggib* 'son', *wananggub* 'daughter', *wanangga* 'children'. In view of these and other variations in the vowels, and on the assumption that the protolanguage may have shown similar variations, it is not surprising to find a variety of vowel alternations among the various Indo-Pacific languages. One remarkable feature has been found to hold in well over twenty examples with the only clear exceptions being the New Britain and Bougainville languages which consistently reverse it. The masculine singular vowel is always more front than the feminine or, if you will, has a higher second formant. If we write the basic vowels then in the order *i, e, a, o, u,* the masculine vowel will always be to the left of the feminine vowel. A further illustration beyond the Marind examples *e/u, a/u, i/u* cited above is Moraori, a language of the SNG subgroup in which the gender of the verb object is expressed by vowel changes which moreover are different in different tenses. Examples are *termi* 'thou strikest him' (present tense), *torma* 'thou strikest her' (present tense), *kesnemefi* 'you struck him' (before yesterday past), *kosnomafi* 'you struck her' (before yesterday past). Here we find the following vowel changes from masculine to feminine *e/o, i/a, e/a,* all of which conform to the above rule.

The following are the basic facts regarding the distribution of this pattern. In Halmahera we find that while the masculine and femine singular have consonantal indicators, there is agreement with one basic feature of the Marind system. There is a third neuter gender, it coincides with the neutralized plural of the masculine and feminine, and it is expressed by *i* exactly as in Marind. Thus the verb subject pronouns in Galela are masculine singular *wo*, feminine singular *mo*, neuter *i* and plural *i*. On New Britain, Taulil and the closely related Butam likewise show a neuter *i* in contrast with masculine *a*, feminine *e* in the singular. In the plural the widely distributed pronoun *ta* is found.

It seems reasonable to hypothesize that the widespread third person pronoun *i, y-* which functions, as has been seen, in SSP and PPP, but is even more widely distributed, takes its rise from the neuter and plural third person *i*.

In the Nasioi language of Bougainville, Ogan (1966) who reports on the kinship terminology, gives minimal sex contrasts based on vowels which involve the reversed pattern in some instances. Examples are *nuring* 'son', *norang* 'daughter', *naung* 'husband', *naang* 'wife'.

It was noted earlier that the only published evidence regarding the non-Austronesian language of New Ireland is one sentence in Capell (1954) and kinship terminologies in the ethnological report of Chinnery (n. d.). It is therefore all the more remarkable that it should give us further examples of vowel alternation and this time in conformity with the usual New Guinea pattern. From Limalua (p. 17) Chinnery gives *poirang* 'son', *poirung* 'daughter'. The *a/u* alternation is found in Marind, as previously noted.

From New Guinea we find instances of these vowel alternations in NNG, SWNG, SNG, and CNG. In the first of these, NNG, we note that Monumbo with a five-gender system has feminine singular *u*, neuter singular *i*, as in the independent pronouns *uk* and *ik*. The other indicators are consonantal. There is here complete coincidence with the normal *u* feminine and *i* inanimate of Marind. In the Nimboran verb the sex of the third person subject is indicated in the singular by the suffixes *-am* 'masculine', *-um* 'feminine', in which *m* as pointed out in earlier discussion, is the common third person morpheme of NNG. The vowels *a/u* coincide with the New Ireland and Marind patterns quoted earlier.

In SWNG, to which Marind belongs, all the languages of the Ok and Marind subgroups exhibit the features being discussed here. It was indicated in the treatment of SWNG as a group that the agreement between these two subgroups in the particular local variant *e* masculine singular, *u* feminine singular, *i* plural was an important piece of evidence in favor of the unity of SWNG. Examples are so numerous here that they need not be further quoted.

In SNG an example of vowel changes for sex of third person object was noted for Moraori earlier. In CNG, examples are available from the western and central subgroups. The Huon (eastern) subgroups do not have sex gender. Ekari in the western subgroup uses an *i/a* contrast as suffixed verb subjects and *i/o* in the demonstrative. The only example I have found in the central (Highlands) subgroup is in Enga kinship terminology (Wirz K.919) *wanengk* 'son', *wanangk* 'daughter'. There are also instances of vowel contrast for sex in the relationship terms in Asmat of the Kamoro group which is one of those left unclassified in the present study.

Consonantal indicators of sex gender are also common. They are quite various but there are some points of contact among different subgroups. One of these deserves special mention, the occurrence of *m* as a feminine marker. This was already noted for the singular pronoun in the Halmahera languages. In Anceaux's unpublished materials it occurs as verb subject in Aitinjo in WNG, e.g. *tamo* 'he goes', *mamo* 'she goes'. It also occurs in Yava, the language of central Japen (Anceaux 1961) which was tentatively assigned to WNG. The examples here are the Mantembu pronouns *ue* 'he', *uem* 'she', and the verb subject prefix in *more* 'she goes' contrasting with *de* 'he goes'. Since these languages are all assigned by Cowan to his West Papuan Phylum this is important grammatical evidence for this grouping. It does, however, seem to occur in the Telei language of Bougainville, e.g. in the indefinite pronoun singular *niinu* 'masculine singular', *noma* 'feminine singular', *nogo* 'masculine plural', *nomi* 'feminine plural', and with the same plural *g/m* opposition in the independent plural pronouns *igau* 'they, masc.', *emi* 'they, fem.'

11. The formation of a past tense by means of a suffix containing a velar consonant has been widely noted. Where there are several past tenses it tends to be the more remote. The examples recorded come from three subgroups outside of New Guinea and four within New Guinea as follows: AN *-ka* (all languages in the main or northern subgroup); HA *-ka* (where there is evidence, except Galela which has *-oka*); CM Bilua *-ke*; NNG Nimboran *-k-*, Sentani *-ka-* (labelled 'aorist'), Anaberg *-g*; Yelogu *-k-*; Wosera *-k-* (in interrogative sentences); SWNG Aghu *-k-* 'before yesterday past', Jaqai *-k-* or *-b-* (*promiscue* according to Drabbe); SNG Jei *-ag-*, *-eg-* 'remote past', Mombum *-agha-* 'remote past', Ndom *gh*; CNG Dani *-k-* 'remote past', Moni *-ga* 'remote past', *-g-* immediate past; Duna *-gu*, Karam *-k-*, Kewa *-uga-* 'remote past, Ono *-ko-* 'remote past'. This element regularly follows the root. Where indicated as nonfinal here it is followed by obligatory inflectional elements.

Table 1 summarizes the distribution of the preceding items. By 1a is meant the *n* first person pronoun, by 1b the *t* first person pronoun. By 10a is meant

the vowel pattern of gender discrimination, by 10b the occurrence of *m* as a feminine indicator.

TABLE 1

	1a	1b	2	3	4	5	6	7	8	9	10a	10b	11
AN	−	+	+	−	−	+	−	−	−	−	−	−	+
TA	+	+	−	+	+	+	−	−	−	+	−	−	−
HA	−	+	+	−	+	−	−	−	−	−	+	+	+
NB	+	−	+	−	−	+	−	−	−	−	+	−	−
BO	+	−	−	+	−	−	−	−	−	−	+	+	−
CM	+	−	+	−	−	+	−	−	−	−	−	−	+
TS	−	−	+	−	−	−	−	−	−	−	−	−	−
WNG	+	+	+	+	−	−	−	+	−	−	−	+	−
NNG	+	−	+	+	−	+	+	−	−	−	+	−	+
SWNG	+	−	−	+	−	−	+	+	−	+	+	−	+
SNG	+	−	−	+	−	+	+	+	+	−	+	−	+
CNG	+	+	+	+	−	+	+	+	+	+	+	−	+
NENG	−	+	+	−	−	−	−	−	−	−	−	−	−
ENG	+	+	+	+	−	−	−	+	−	+	−	−	−

In evaluating the significance of the table, there are several considerations to be kept in mind. Some of the negative entries are most probably the result of poor documentation. At one extreme we know virtually nothing of Tasmanian grammar whereas, thanks largely to work of the SIL group, we probably know most about the grammar of the CNG languages. Moreover this is an arbitrary selection of grammatical features and, of course, lexical evidence does not figure at all. With all these reservations the following observations may be offered. HA, TA, and WNG corresponding to Cowan's West Papuan Phylum may constitute a 'super-group'. This would receive some support from lexical distributions also. On a similar level there appears to be a nuclear New Guinea group consisting of NNG, CNG, SWNG, and SNG. It is noteworthy that the NENG (Madang) definitely does not belong here and shows just as much resemblance to non-New Guinea as to New Guinea languages. There may also be a similar high-level grouping containing NB, BO, CM and Rossel Island.

It should be emphasized that the above presentation probably by no means exhausts the grammatical evidence. Among other possible lines of investigation that might be pointed out are the following. The existence of separate verb stems for singular and plural (sometimes dual) action is widely reported (e.g. Andamanese, SWNG, CNG, NB). This was not investigated in detail

and might well uncover specific points of contact among the subgroups. Another interesting feature is consonantal ablaut, e.g. the initial consonant alternations of the Halmahera language, consonant alternations of the suffix forms in New Britain languages, and typologically similar alternation in CNG. This was also not investigated here. Whether indeed the question of historical connection among these can be resolved on present evidence is doubtful.

Although not in itself decisive it is perhaps worth remarking that certain typological traits of word order, the direct opposite of those of Austronesian languages, appear with great consistency. The typical Indo-Pacific word order in transitive sentences is subject–object–verb and for the genitive construction the possessor regularly precedes the possessed. In Austronesian languages the object normally follows the verb which often comes first in the sentence and the genitive order is possessed followed by possessor. Among Indo-Pacific languages only the Santa Cruz subgroup of CM has the usual Austronesian order. Every other group has the typical Indo-Pacific order in all or a great majority of the languages.

I believe that the evidence presented here is sufficient to establish the point that the vast majority of non-Austronesian languages outside of Australia, on which judgement is still reserved, have a common origin. For the non-Australian languages, we must of course consider the fact that there are still a fairly large number of languages for which we have little or no data so that nothing can yet be stated concerning them. For Tasmanian the relative paucity of data including the virtual absence of grammatical information necessarily produces a somewhat weaker case than in other instances. Still what evidence we have does point in this direction.

In judging this evidence it is reasonable to ask the reader not to lose sight of the main body of evidence in noting, as he no doubt will, some inaccuracies in citation and some comparisons which perhaps will not stand up in the light of further investigation. In a first attempt, and this is a first attempt, to delineate the basic features of such a vast family, some errors are indeed inevitable.

Regarding the first of these possible inaccuracies in citation, I should point out that, except for an initial period in which I received assistance from George Grace, all of this material was copied from original sources by myself over a period of twelve years during which it occupied a major share of my attention. I have tried to check carefully but in the multiple process of my own copying, secretarial typing, and typographer's typesetting I would be surprised if there were not at least some minor errors.

During this period, particularly as new data were published adding continuous corroboration to old etymologies and grammatical comparisons and giving rise to new ones, more and more points of contact appeared among these languages which, to begin with, appeared to me to be very divergent. My hope is that the present study will help to hasten the long overdue demise of the notion of Papuan as merely a scrapheap of assorted languages bound together by the negative characteristic of being non-Austronesian. May the comparative study of this major linguistic stock, which has been so strangely neglected and whose importance for many Austronesian problems is basic, finally come into its own.

Indo-Pacific Etymologies

1. above	TA Abui *epe* 'upon, at'; BO Konua *piai* 'upwards'; CM Savo *piai* 'go up, upwards'; WNG Mairasi *oβiβi*; NNG Ampas, Waris *op* 'sky'. Demta *op* 'high', CNG Jabi, Ekari *epa* 'sky' Dani *pu* 'up', Agarabi *opu* 'sky', Gende *pea* 'above', Ago, Naga *paiq*, Mape, Wamola *faiq*, Migaba *feiq*, Hube *hoi*, Kate *fai*; NENG Em *fia*; ENG Namau *upai*; UNG Iria, Asienara *ope(ra)*, Asmat *op* Mimika *opo*
2. all₁	AN Biada *doga* 'much'; TA Abui *toka*; NB Taulil *tugus*; BO Siwai *tuki*; WNG Baham *taghüa*; CNG Dibiri *tugarama*; NENG Bongu *jegar*, Sungumana *jagar*; UNG; Asmat *takas*
3. all₂	HA Modelo *odomu*, Tabaru *odumu*; WNG Najarago, Tarof *udumia* 'much', Mairasi *qatmaqia*; NNG Valman *comcom* 'many', Manambu *cʌmʌcʌm* 'much'; SWNG Jaqai *ndom*; SNG Dabu *tomambi* 'all, many', Kimaghama *ndom*; CNG Dem *ondoma*, Ekari *idima*, Moni *edema*, *ontoma*
4. arm (upper)	AN Puchikwar, Juwoi, Kol *ben* 'shoulder blade'; WNG Teminabuan *mbɛng* 'shoulder'; Kapaur *mben* 'shoulder'; NNG Kwesten *fan* 'hand', Murik *pena-garöb* 'shoulder'; SWNG Kati *ben*, Telefol *ben* 'forearm in counting', Metomka *mben* 'arm, hand'; SNG Goaribari, *bena* 'upper arm', Turama, Urama, Iwainu, Kerewa, Era River *bena* 'shoulder'; Eme-eme *beno* 'shoulder'; CNG Jabi *benai*; NENG Saker *ben* 'hand, branch'; UNG

		Asmat, Kajakaja *ban* 'hand', Kauwerawet *pan* 'upper arm', Koassa *ban* 'upper arm'
5.	arrow	AN Kede *tul*, Chariar *taul*, Bale *del*, Juwoi *tol*; WNG Barau *taure*, Mor *taura*, Kampong Baru *tauru*; SNG Karami, Eme-eme *tiri* 'bow', Goari-bari, Kerewa, Turama, Mawata, Kiwai, Wabuda *tere*, CNG Tsaga *telya*, Tijiranki *tela* Fasu *tare*; UNG Kamoro *tearə*, Mimika *tiarə*
6.	ashes	AN Beada *bug*, Bale *buk*; HA Ternate *fika*; WNG Asli-Sidi *bok*, Mogetemin *boh*, Maibrat *buh*, Ajamaru *box*; SWNG Boazi *pokok*, Konmak *pokak*
7.	bark (of tree)	AN Kede *kapo*, Chariar *kaba*; TA Bunak *koma*; Bo Baitsi *kamu*, Siwai *kang*; NNG Maprik, Wosera, Manambu *cəpə*, Yelogu *cəfi*, Nyaura *cəmbə*; SNG Karima *kaibo* 'skin', CNG Kiniageima, Jalimo Anggapuruk (*na*)*kap* '(my) skin', Jupna *gap* 'skin' ENG Namau (*iri*)*kape* '(tree) skin = bark'
8.	to bear (a child)	AN Chariar *otu*, Kede *te* 'child' TA Abui *yadi* BO Telei *utu* 'be born', CM Reef *tou*; WNG Mairasi (*iwɔ*) *atu* '(tree) child = fruit'; NNG Nimboran *tu* 'child', Mekei *dɔ* 'child'; CNG Jabi, Ekari *uta* 'fruit' Pesechem *ot* 'child'; NENG Bongu *at*; ENG Mailu *odi*, Dimuga *otua* 'offspring'
9.	beautiful	HA Pagu *ofiofi*; WNG Mor *iaba* 'good, sweet', Barau *boɛ* 'sweet', Mairasi *iβi* 'sweet', Etna Bay *ivi* 'sweet', Faranjao *iwi* 'sweet', Meringgo *owf*, Mansibaber *eif*; NNG Sentani *foi* 'beautiful, good', Tanahmerah II *poi* 'good', Arapesh *apui*, Ngala *afʌ* 'good'; SWNG Aghu *yafi*; SNG Tunjuamu *poio* 'good'; CNG Moni *poya*, Ndani *op* 'good', Tsaga, Tjiranki *epe* 'good', Augu *eve* 'good', Hube *fea* 'good'; ENG Korona *ifi* 'good', Mafulu *ifa*(*ne*)
10.	belly	CM Nabalue, Noole *bilu* 'guts'; TS Northeast *pla*(*na*), Middle East *peri*(*na*), West *pile*(*re*); NNG Nori *buri* 'heart', SWNG Telefol *bubul* 'heart'; SNG Keladdar *puri*, Jei *buri*, Jab *bulo* 'heart'
11.	blood	BO Nasioi *ereng*; *ereere* 'red', Baitsi *erei*, Koianu *ereng*, Siwai *iri*; CM Baniata *aroa* 'red'; WNG Solowat, Itigo, Bira *aruo*, Madik *el* 'red'; NNG Sko *olo, oli* 'red'; SNG Urama *ora*, Era River *ola*,

Koneraw *irri*, Miriam *arau* 'red'; CNG Moni *era*; NENG A'e *uring* 'red', Bogadjim *leng* 'red', ENG Mambare, Binandere, Yega, Tain-Daware *ororo*, Manukolu *ro*, Namau *aro*

12. bone   TS Southeast *teni*; WNG Madik, Karon *dini*; NNG Mekei *dɛn*, Kwesten *rɛn*, Kamtuk *don*, Nimboran, Kuangsu *dɔng*; NENG Bogadjim *tanu* 'bone, body', UNG East and West Rossell I. *dona*, Yela *donna*, Kwai, Olango, South West Rossell *döna*

13. to break   TA Abui *bal*; NB Uasi *vala*, Taulil *varka*; BO Konua *buri*; SNG Kimaghama *boro*, Ndom *börbör*, Teri-Kawalsch *burriye*, Jei *belna*; UNG Yela *bwari*

14. bush   HA Pagu *bongana*; TS North East *bangala* 'bush, forest'

15. buttocks   AN Biada *dama*, Bale *doamo*, Puchikwar *tomo*, Juwoi, Kol *tome*; TA Oirata *itim*; WNG Etna Bay *sombo*; NNG Monumbo *tsimba*, Yelogu *dyambu*; Sentani *dəmə* 'tail'; Demta *dum* 'tail'; SWNG L. Murray *dumu*; SNG Dorro *temi* 'tail' Keraki *semi* 'tail'; Kimaghama *cöme*, Riantana *cambö* 'tail', Teri-Kawalsch *cumme*; CNG Kate *domba* 'root of tail', ENG Oru-Lopiko *dama(nota)* 'anus' (*nota* probably = 'hole'); UNG Waruna *odoba* 'tail', Mimika *atabu*

16. child   AN Kede *etira* 'boy', Chariar *etire* 'boy', Bogijiab *tire* 'boy, child', Juwoi *tre* 'baby'; TA Oirata *doli* 'small'; NB Taulil *idila* 'small, child'; NNG Kilmeri *turi*; ENG Uaripi, Lepu *aturea*, Mailu *ature*; UNG Olango (Rossel I.) *tier*

17. cloud   TA Bunak *taho*, Abui *tabo*; HA Modole *dipa* 'sky' BO Koianu *tubei-tubei*, Nasioi *tupetupe*; CM Banua, Nea, Noole, Nabalue *duba*; WNG Barau *tuva* 'sky'; NNG Siaute *tapai*, Nori *utepo*, Kamtuk *ndop* 'smoke'; Ngala *tu[m]bu*, Yelogu *tu[m]bu*; SNG Peremka *dapar*, Parb *dabar*, Bangu *davwar*, Dorro *jafarr*, Dabu *dapar*, Kiwai *tobore*, Sisiame, Pirupiru, Buniki *toboro* 'cloud, sky', Dibiri *toboro*, Dungerwab *dabar* 'sky'

18. to come   TA Bunak *man*; Oirata, Makasai *maqu*; Abui *me, mei, mirani*; BO Nasioi *maniai* 'to go'; SWNG Marind *man*, Digul Upper Muju, Niinati, Metomka *mene*, Mandobo *mend*; SNG Moraori *umon*, Teri Kawalsch *muinane*, Ndom *aman*, Jab *man*, Jibu *mane*; CNG Dem *nem*, Simori *meina*, Wolani *mena*, Ekari *mindi*; NENG Sungumana *minjaa*;

ENG Gosisi *mana* 'bring', Namau *mina* 'bring'; UNG Tate *mane*

19. to cry   AN Puchikwar, Kol *war*, Juwoi *yar*, Chariar *olo*; TA Oirata *oore*, Makasai *iar*, Bunak *ɔlɔ*; HA Pagu, Modole *ali*, Galela, Loda, Tabaru *ari*, Sau, Wai *adi*; SNG Kunini *ele*; CNG Sinasina *ale*, Dom *ali*; NENG Bongu *ara* 'scream'; ENG Mailu *ari* 'ceremonial mourning', Iworo *ario*; UNG Tanahmerah I *aire*

20. to dance   NB Taulil *men*; Baining, Butam *mein* NNG Nori *mondai* 'to sing'; SNG Mawata, Kiwai, Wabuda *amaro*, Pirupiru *omaro*, Oriomo *omare*; CNG Tsaga *mali* 'dancing, singing'; NENG Langtub *muni* 'Tanzfest', Bongu *muun-(dawagh)* 'to dance, to sing', Sungumana *mun-(tuwa)* 'to dance'; ENG Oru-Lopiko *aemario* 'song'

21. to die   TA Oirata, Makasai *umu*; NENG Banara *uma*, Bongu, Saker *mo*, A'e *mɔ*, Em, Ate *mau*; ENG Namau *imua*

22. dog   HA Galela, Tabaru *kaso*, Tobelo *kaho*, Modole *qaho*; BO Konua *keesa*; SNG Barika, Dugeme *kase*, Karami *kso*, Ipikoi *gaho*, Hibaradai *gaha*, Eme-eme *gaso*, Mahigi *gahola*; CNG Fasu *kasa*, Deduae, Ono *kasi*, Hube *kazu*

23. ear   NB Taulil *tama(ka)*, Baining *sdem(ki)*, Uasi *lomu* 'to hear'; BO Telei, Koianu, Baitsi *rom*; WNG Kampong-Baru *odaba* 'to hear', Inanwatan, Solowat, Bira *tobo* etc. 'to hear'; NNG Anaberg *rɔmu*; SNG Turama *demeriwai* 'to hear'; NENG Sungumana *damoi*, Burumana, Koliku, Male *damui*, Bongu *dab*, Ate *dabi*, Em *deb*, Rempin *debur*; ENG Giumu, Tahari *doma*

24. earth   AN Bale *moga* 'bottom', Bogijiab *mika* 'underneath', Puchikwar, Juwoi, Kol *mikam* 'underneath'; TA Bunak *mɔɔk*, Oirata *miki* 'mud', Abui *mok*; NB Sulka *miie*; TS Middle East, Southeast *mara*; SWNG Awju Sjiagha *moka* 'earth, underneath', Awju Pisa *mokā*, Marind *makan* 'earth, underneath'; Jaqai *mogon-(ape)* 'underneath', Aghu *makā* 'ground', Telefol *bakan* 'ground'; CNG Gadsup *makai*, Gende *mikai*, Siane *mika*, Benabena *meqi*, Wagamb, Kuno *mar*, Chimbu *magan*, Dem *mok*, Jubi *maki*, Wolani *marai*, Ekari *maki, magi*, Moni *may, mayi*; ENG Minjori *moia*, Suambe *moiia*, Elema *mea* 'land', Haira *mea*, Kairi-Kaura *mio* 'land', Uaripi, Milareipi, Toaripi *mea* 'land'

25. to eat   TA Oirata, Makasai *nawa*, Abui *nei*; BO Nasioi *noi* 'eat, drink', Koianu *nai* 'drink'; HA Galela *ino* 'food'; WNG Konda *no*, Puragi *niga*, Kanpaur *nowa*, Baham *nou*, Patimuni *nawa*, Karas *nanet*; NNG Wembi *nana*, Skofro *nəna*, Ampas *neu*, Sentani *anə* 'eat, drink', Nafri *an*, Valman *nago* 'drink', Arapesh *nig* 'feast'; SWNG Niinati *ane*, Telefol *unin*, Awju Pisa *nī*, Wambon *en*, Aghu *ẽ*; SNG Komelom *noku*, Mombum *noku*, Kanum *anang*, Jibu *nina* 'food', Karima *novu* 'food', Bara *na* 'food', Dabu *none*, Bugi *nana*, Koneraw *nugu* 'food', Goaribari, Turama, Era River *naa* 'food', Tunjuamu *nune* 'food'; CNG Tjuave, Sinasina *nega*, Kuman *neinga*, Dom *neeka*, Tsaga *ne*, Tjiranki *ne*, Uhunduni *no*, Enggipilu *nu*, Dem *nen*, Forei, Gadsup *na*, Tairora *na* ∼ *ne*, Gende *nua*, Jamafi *nowa*, Medlpa *nui*, Kewa *na*, Sau *nee* 'food', Tugi *nano* 'food', Sua *no* 'food', Matap *nang*(*at*) 'I eat, drink', Wantoat *na* ∼ *näp*; ENG Yema-Yerawe *nai*; *nena* 'food', Mawai *nona* 'food', Kwatewa *nat-ni* 'food', Dimuga *nana*, Onjob *na* 'eat, drink' *nana* 'food', Oru-Lopiko *naro* 'eat, drink', Kuepa *neka* 'food', Afoa *nana*, Tauata *nai*, Mafulu *nene*, Kotoi *nananu*, Mulaha *ina* 'drink', Namau *nav* 'eat, drink'; UNG Tate *nove*, Gaima *nou*, Pauwi I *ani*, Kamoro, Sempan *ne*, Asmat *an* ∼ *n*, Mimika *ena*, Yela (Rossel I.) *nna* 'feast'

26. egg   AN Biada *molo*, Bogijiab *mula*, Kede *mulo*; WNG Aitinjo, Mogetemin *make*; SWNG Jaqai *moko*, Awju Pisa *mugo*, Aghu *mügo*; CNG Aua *mokl*, Gawigl, Mogei, Medlpa *mukl*, Wagamb, Kuno *muru*, Nondugl *muro*, Chimbu *muklo*, Dom, Boumai *mile* 'egg, fruit', Sinasina *mile*; ENG Domara *muru*, Mailu *muru-u*[4]

27. to fall   AN Biada *pa*; TA Oirata *ipa*, Bunak *pai* 'arrive', Abui *pe, pei* 'go down'; CM Laumbe *foava, foama*, Banua *voi* 'go down', SNG Teri Kawalasch *paa*; CNG Wantoat *epu* 'come

---

[4] In this and a few other etymologies (and in others confined to New Guinea and not contained in this list) one finds an unusual correspondence of *k* or *g* with *r* or *l*. In the present instance some of the CNG languages with *kl* (i.e. a velar lateral) show the source for these divergent phonetic developments. This sound is probably more widespread than appears from existing sources. Thus the name of the Ekari language of some sources is written Ekagi by Drabbe in his grammar of that language. Doble, a professionally trained linguist, reports this sound to be indeed a velar lateral. What is probably the same sound is described by Feuilletau de Bruijin in somewhat sensational terms as 'Een merkwaardige letter is een achter in de keel gebrouwen *r* die als het ware een overgangsletter vormt tot de *k* en daarmee ook wel verwisseld wordt' (K.517: 144).

down'; NENG Banara *ipapa*, Saker *vo*; ENG Mailu *poo* 'fall, of fruit', Toaripi *foi*

28. female   TA Bunak *fana, pana* 'woman, wife', Oirata *panar* 'female of animals' BO Siwai *panna* 'wife'; WNG Aitinjo *finya* 'women', Mogetemin *fanya* 'woman', Asli-Sidi *finia* 'woman'; CNG Wolani, Moni *pane* 'girl', Karam *pany* 'daughter', Dem *pani* 'female'; NENG Langtub *pano* 'woman, wife'

29. fingernail   AN Bea *pag* 'claw', Bale *poag* 'claw'; NB Sulka *paaga*; WNG Baham *pag*, Teminabuan *pek*; NNG Nafri *faxa*; ENG Amara *foka*; UNG Tate *faha* 'claw'

30. fire   AN Bogijiab, Kede, Chariar *at*; TA Oirata *ada*, Makasai *ata*, Bunak *ɔtɔ*, Abui *ara*; BO Nasioi *nta, tai* 'to burn', Konua *eto* 'to light a fire'; TS Southeast *to*, West *toi*; WNG Madik *yet*; SWNG Boazi *te*, Awin *de, di*, Madinava *da*; SNG Jab *te*, Melmek *ete*, Ipikoi *tai*, Goaribari *taetae* 'burn', Makleu *atha* 'burn'; CNG Jabi, Simori *utu*, Moni *usa*, Tsaga *te* 'burn(intr.)'; NENG Nupanob *ote* 'fever'; ENG Gebi *ita*, Mailu *odaoda* 'hot', Mafulu *ade* 'burn', sogeri *taite* 'hot'; UNG Kamoro, Mimika *uta*

31. fish   BO Telei *topi*, Nasioi *tavi*; WNG Mor *daba*; NNG Atemble *jab*; SWNG Boazi *seve*, Konmak *seva*; SNG Bangu *caua*; NENG Langtub *sieb*; ENG Uaripi, Milateripi *tava* 'mullet', Suku *seva*, Karukaru *sivi*

32. fog   AN Biada, Bale, Puchikwar *pulia*, Juwoi *pulye*, Kol *polia*; TS Middle East *prü'a*, Southeast *bura* 'smoke', WNG Merani *bilo* 'smoke'

33. forest   AN Biada *erema*; TA Oirata *irim, irimi*; NB Baining *inim* 'bush land'; CM Laumbe *aram* 'land'; SWNG Telefol *atem* 'forest', WNG Mailu *arima* 'plot of ground'; UNG Kauwerawet *idim* 'grondgebied'

34. to give   TA Oirata *ina*, Bunak *ani, ini*; NB Baining *een*; WNG Asli-Sidi, Mogetemin *ne*, Purgai *naqe*; NNG Atemble *ani*; CNG Moni *ena*; ENG Afoa *inie*, Tauta *ini*

35. good   HA Galela *tebi* 'beautiful, clean', Tobelo *tebini* 'beautiful, clean', Pagu *tebin* 'clean', Loda *tebini* 'clean'; BO Baitsi *tabei*, Nasioi *tampara*; SWNG Telefol *tambal* 'well'; SNG Miriam *debe*; CNG Karam *te[m]p*; NENG Ate *teiba*; ENG Kambesi *teba* 'sweet'

36. hair   AN Önge *de*; TA Oirata *wata*, Makasai *asa*, Bunak *adu*; HA
                  Tobelo, Modole, Wai *utu*; CM Bilua *tou* 'hair, feather';
                  CNG Moni *tu* 'feather, hair', Dani *esi*, Upper Pyra-
                  mid, etc., (Ndani) (*n*)*iti* 'my hair', Tsaga *iri, id* 'hair,
                  feather', Tjiranki, Samberigi *iri* 'hair, feather', Augu *iri*, Sau
                  iri(*gi*), Kewa *iri*, Fasu *iti* 'hair, feather'; ENG Uaripi *utu*,
                  Kovio *ada*, Binandere, Yema Yarawe, Jegasa Sarau, Yega,
                  Aiga, Tahari *tu*; UNG Mimika *iti* 'feather', Olango *to*
                  'feather', Kwai *doa* 'feather'

37. hand   AN Kede, Charia *tong* 'arm'; TA Oirata, Makasai *tana*,
                  Bunak *dɔn*, Abui *tang*; BO Nasioi *tanka* 'arm, hand'; WNG
                  Kapaur *taan*, Karas *tan*; SWNG Upper Tedi *teni* 'arm',
                  Niinati *tini*, Telefol *teeng*, Upper Fly *teng* 'finger', Kandam
                  tini, Unkia *teng*; SNG Gijara, Dabu, Dibolug, Kibuli,
                  Mbayaka, Agob *tang* 'arm', Ngamai *tan* 'arm'; UNG
                  Goliath *tang*

38. head   TA Bunak *ubul*; BO Siwai *puri*, Telei *pure*, Nasioi, Baitsi
                  borei; CM Savo [*m*]*batu*, Laumbe *vatu*; TS Northeast *pace*;
                  WNG Mnasibaber (*me*)*bir*, Manikion (*me*)*btu*; UNG
                  Kauwerawet (*na*)*par* '(my) head', Koassa (*na*)*bara* '(my)
                  skull', Wamiu *mbada*, Kwai, S. W. Rossell, E. Rossell,
                  W. Rossell *mbara*, Olango *mböda*

39. to hear   TA Bunak *ilek*; BO Siwai *lung*; NNG Siaute *elung* 'listen';
                  CNG Sawuri Hablifuri (*na*)*ruk* '(my) ear', Enggipilu *eloh*
                  'ear' Gafuku *lagh* 'ear'; NENG Bongu *lagh*

40. to hit   TA Oirata *wewese*, Makasai *basi*, Bunak *pas*; HA Loda
                  posana, Modole *poha*; CM Bilua *pazo* 'scourge'

41. horn   AN Biada, Kede *wulu*; TA Oirata *uru*; HA Tobelo *guguli*[5]

42. husband   CM Nea *mʋe*(*nale*), Noole *mʋ*(*nale*), Nabalue *mʋe*(*nalu*)
                  'husband = spouse (male)' cf. *lupnale* 'wife' in Nea etc.;
                  WNG Mairasi *umo*; SWNG Aghu *amo*; SNG Ro *mui* 'man',
                  Mahigi *ami* 'man'; CNG Dom *eme* 'man'; NENG Dagoi
                  mu; ENG Tauata *mu* 'husband, man', Koita *mo* 'male',
                  Yema Yarawe *ema*, Keveri, Bauwaki, Neme *emi* 'man',
                  Kororo, Boli, Doriaidi *eme*, 'man', Tsia *emo* 'man'; UNG
                  Inora *meoa*, Sempan *mao*, Kwai *mma* 'man', Olango
                  mmö 'man'

---

[5] In the Halmahera languages there is a set of initial consonant alternations including one in which
g may alternate with zero.

43. knee   HA Tobelo *buku*, Pagu *bubukul*, Modole *bubuqu*, Tabaru *bubuku*, Loda, *wuwuku*; BO Koianu *poku*; SNG Teri Kawalsch *bugu*; NENG Saker *bakbakam*, Tombenam *ambakan*

44. leaf   NB Uasi *boloxu*; BO Koromira *bara*, Nasioi *para*, Baitsi, Naisioi *pana*; TS Northeast *paroko*, West *paroce*, Middle East *perote*; CNG Sisiame, Pirupiru *pori*, Barika *iboro*, Gijara *poringai*, Kawam *per*, Komelom *pur*, Mombum *epur*, Koneraw *bur*; ENG Manukolu *evarau*, Suku, Koiari, Maiari, Koita *hana*; UNG Mantembu *bara*

45. lip   AN Biada, Bogijiab *pe*, Bale, Puchikwar *pa*; WNG Konda, Jahadian *pa* 'mouth', Kampong Baru *epa* 'mouth'; NNG Nori *epu*, Gamai, Watam *uup* 'mouth'; SWNG Oser *bo*; CNG Jabi, Ekari *ebe* 'mouth' Moni *bay* 'mouth', Mid Grand Valley Dani (*na*)*be* '(my) mouth', Oeringoep *ambe*, Gende *apa*, Naga, Mape, Wamola, Magobineng, Deduae *bibiq*, Kate *bipiq*, Ono *pi*; ENG Amara *pe*, Aiga, Berepo *pe* 'mouth', Binandere, Yema Yarawe, Mawai, Yega, Tain Daware, Jauwa, Hunjara *be* 'mouth', Neme, Morawa, Binahari, Monomor *bebe*, Merani *bebe* 'mouth', Orai-iu *bebeu* 'mouth', Kambisa *uba*, Kairi Kaura *ape*, Uaripi, Haira, Milareipi, Toaripi, Lepu, Orokolo *ape* 'mouth', Mulaha *abe* 'mouth'; UNG Mimika, Asmat *mbe* 'mouth', Gogodala *epo*

46. long   TA Bunak *lɛgul*, Abui *lohu, lougda*, Oirata *lokide* 'high'; HA Tobelo, Galela *luku* 'deep'; NB Uasi *alelaxu*

47. louse   TA Oirata *amin*; NB Uasi *meni*; WNG Kapaur *meng*, Baham *min*, Teminabuan *men*, Amberbakan *im*, Karon, Patimuni *mim*; NNG Apris *emana*, Siaute *imum*, Tanahmerah II *ming, min*, Sentani *mi*, Demta *ami*, Maprik, Wosera *nyəmu*; SWNG Digul *im* 'louse, flea', Marapka *um*, Metomka *im*; SNG Hiwi, Iwainu, Kerewa, Kiwai, Mawata, Domori, Wabuda, Sisiame *nimo*, Jei *nim*, Kimaghama, Riantana *nöme*, Teri Kawalsch *numme*, Ndom *neemön*, Karigari, Moie *ianame*, Dorro *yaname*, Miriam *nem*, Koneraw, Komelom *am*; CNG Moni *amu*, Tsaga, Tjiranki, Samberigi, Tugi *rema*; Siane, Mogei *nema*, Chimbu, Wagamb, Kuno *numan*, Dom, Sinasina, Boumai, Tjuave *niman*, Kate *imeng*, Mape, Wamola, Magobineng, *imang*, Hube *iming*, Bulum *imin*, Ono *emen*, Naga *yɔmeng*, Momale, Migaba, Deduae *yomeng*, Sene *ime*; NENG Bongu, Bogadjim *mana*, Sungumana *manag*, Em, Ate, A'e, Nupanob, Botelkude *mi*, ENG Manukolu

nomone, Mulaha *uman*, Neneba *uma-a*, Agi, Uberi, Koiari, Maiari *umu*, Koita *omo*, Gosisi *mumu*; UNG Mantembu ɛmɛ, Kamoro *mamo*, Mimika *mama*, Kamoro, Sempan *im* 'flea', Gaima *ami*, Yela *yema* 'flea', Olango *iyema*, Southwest Rossell *yiema*

48. male   TA Oirata *nami*, Abui *neng* 'husband'; BO Nasioi *nema(ka)* (pl. *nema*) 'boy'; WNG Baham *namia*, Mor *niamia* 'man'; CNG Ndom *namun* 'husband'; Mombum *nam* 'male, man', Koneraw, Keladdar *nam* 'man'

49. man   AN Biada *liga* 'boy'; HA Galela *roka* 'husband'; Tobelo, Modole *rokata* 'husband' Isam, Tololiku *lokat* 'husband', Ternate *raka* 'husband'; NB Taulil *loka*; BO Siwai *lugang*, Telei *lugang, dugang*

50. meat   TA Oirata *mede* 'to eat'; NB Taulil *mat* 'food', Baining *mes* 'to eat', Butam *matmat* 'eat'; NNG Boven Tor *matan*; SNG Teri-Kawalsch *muje* 'food', Kunini *madzu*, Miriam *med*, Oriomo *mər*, Jibu *möre*, Dabu *mid*; CNG Aua, Wagamb, Gawigl, Mogei, Kuno, Tsaga *minc*, Tjiranki *mincu*, Nangamb *mince*, Kate *mitung* 'piece of meat'; NENG Saker *mizan*, Em *meda*; ENG Mafulu, Korona *mise*, Neneba *misiwa* 'flesh', Gosisi *misive* 'flesh', Suku *misi*, Hagari *ove-misi*, Sogeri *mihika*, Koiari *misika*, Koita *misika-a*, Oru-Lopiko *muditsi*; UNG Asmat *amas* 'sago, food'

51. moon   AN Biada *akar*, Bade *ogar*; NB Baining *xorevet(ki)*; BO Nasioi *kara*; NNG Gamai *karwe*, Gapun *kareb*, Murik *karevan*, Anaberg *gəra* 'sun', Atemble *gar* 'sun', Kwesten *kwer* 'sun'; SWNG Telefol *kaliim*; CNG Siane *ikana*, Gahuku *ikani*, Kamano, Kanite *iqkana*, Asaro *ikeqna*, Upper Asaro *geqna*, Gawigi *kalimp*, Medlpa *katilimb*, Mogei *kalyimp*, Tsaga, Tjiranki *kana*, Pong *karib* 'sun', Keseraua *karip* 'sun', Kandomin *kandam* 'moon, sun'; NENG Mis, Saker, A'e, Nupanob, Kemba, Misdao, Botelkude *kalam*, Em *kalam* 'sun'

52. mouth   NB Baining *m(ki)*; TS North East *mo* 'lip', Middle East *mu* 'lip', South East *moye* 'lip'; SWNG Lower Muiu *meia-(kat)*, Digul *meya(kat)*, Ashavi *mea* 'lip', Madinava *mega* 'lip', Kukukuku *mia*; CNG Kate, Naga, Wamola, Magobineng, etc. *miq*, Tsinyadji *ma*; UNC Mumeng I *mya* Huaruha *mii*

53. mud   AN Biada *lab*; HA Tobelo *lepa*; CM Reef *lapo* 'dirty'

54. name   TA Oirata *neene*, Makasai *nai*, Abui *hanin*, CM Savo *nini*, Bilua *ngi*, Reef *(n)ingʊ*; WNG Kapaur, Teminabuan *ne*,

	Najarago *anaia*, Tarof *anai*, Baham *niye*; NNG Monumbo *inu*; SNG Kimaghama *ne*, Teri Kawalsch *neene*, Oriomo *ne*, Kunini *ngi*; CNG Ekari *ena*; ENG Namau *noi*
55. nose	AN Bea, Bale *mun* 'mucus', Puchikwar, Kol *mina* 'mucus', Juwoi *mine* 'mucus'; TA Abui *min*, Oirata *munikain*; TS Middle East *munu*, North East *minera*; SNG Komelom *mene*, Koneraw *munne*, Nombuio *mi-in*, Potaia *meneni*, Dapo *mənia*, Tunjuamu *menye*; CNG Fore ʌ*moni*, Sau *mine-(gi)*; NENG Bongu *mana*, Bogadjim *mana*; ENG Musa R., Tain Daware, Jauwa *mende*, Adaua, Mawai, Yega, Jegasa Sarau *mendo*, Wadewinda *mudu(amba)* 'nose (my)'; UNG Gaima *mina*, Asienara *minika* 'mucus', Iria *miniqa* 'mucus'
56. old	AN Biada, Bogijiab *tam*, Kede, Chariar *taum*; HA Tobelo *timono* 'to be old'; CM Bilua *tam*; CNG Jei *tamana*, Moraori *tamon*
57. to plait	AN Biada *tepi*; HA Tobelo *topi*
58. to push	AN Bogijiab *tera*; HA Tobelo *tila*
59. rain₁	AN Bogijiab *leke*; NB Uasi *lexa* 'water, river'; TS North East *lega* 'water', Middle East, South East *lia* 'water'; ENG Haira *lahi*, Kairi-Kaura *lahe*, Elema etc. *lai*
60. rain₂	TS West *moka*, North East, West, North *moka* 'water'; WNG Aitinjo *moka* 'wet', Mogetemin *moka* 'wet'; NNG Bosngun, Nubia *mɔk* 'rain, water', Nyaura, Yengoru *mayk*, Maprik, Wosera, Ngala *mac*; SWNG Pisirami, Tagota *mauka* 'water'
61. sea	TA Bunak *mo*; NB Sulka *mau*; TS West *moi* 'water'; WNG Jahadian *mu* 'water'; SWNG Oser *mui, mi* 'water'; UNG Kamoro, Sempan *mi* 'water'
62. to see	TA Oirata *nana, ina* 'eye', Makasai *ena*; WNG Waipu *no*, Jahadian *nu, ni* 'eye'; NNG Anaberg *na*, Nimboran, Kuangsu *nu* 'eye'; CNG Mono *ini*, Forei *qana*, Agarabi *one*, Magobineng *ona*; NENG Banara *na*; ENG Mailu, Domara *ini* 'eye', Nemea *nii*, Suki, Agi, Hagari, Sogeri, Maiari, Koita *ni* 'eye', Minjori, Suambe *niie* 'eye', Koiari *ni* 'eye, face'; UNG Tate *ini* 'eye'
63. sibling (older)	TA Abui *nana*, Bunak *nana* 'older sister'; WNG Baham *ano*; NNG Vanimo *nonei*, Sentani *enea*, Nori *eno* 'sister', Wosera *anyo* 'older brother'; SWNG Wambon *nani* 'older sister', Kaeti *neni* 'older sister', Aghu *eni* 'older sister', Niinati *oni* 'older sister', Awju Jenimu *noni* 'older sister', Dumut *ani*

'older sister'; SNG Kimaghama *nanu*, Kunini, Oriomo *nane* 'older brother', Jei *neny* 'older brother', Moraori *nin* 'older sister', Jab, Jelmek *nana*; CNG Moni *ana* 'older sister'; ENG Tauata *ene* 'older brother', Agi *nana*, Suku *nanave*, Koiari *nani* 'older brother', Tahari *nani* 'woman speaking', *nane* 'man speaking'; UNG Girara *naniwa*

64. to sit₁   HA Tobelo *gogeruku*; TS all dialects *kraka*

65. to sit₂   SNG Makleu *man*, Jab *mön*; CNG Siane *min* 'stay, sit', Gende *mina* 'stay', Kamano *meniqnoe*, Mogei *mana-(munt)*, Kuno *amen-(nyint)*, etc.; NENG Langtub *min* 'stay'; UNG Adiba *mana*, Waruna *mana* 'dwell', Gaima, Gogodala *mana* 'sit, stay', Tanahmerah I *mena(φFeta)*, Inora *mena(vera)*

66. skin   AN Bogijiab *kait*, Bale, Puchikwar, Juwoi, Kol *kaic*; BO Telei *katua*; CM Savo *korakora*; TS North East *kite*; SWNG Kandam, Lower Muju, Marapka, Niinati, Metomka etc. *kat*, Kaeti, Mandobo *kota*, Wambon, Dumut *kotae*; CNG Ekari *kado*; NENG Em *gal*, Bogadjim *gara*, Ate *garav*; WNG Uaripi, Lepu *kauri*, Kovio *koro(si)*, Afoa *goti(pe)*, Tauata *kotsi(pe)*, Goilala *koti(pi)*; UNG Tanahmerah I *katane*, Koassa *korora*, Kauwerawet *kora*

67. to sleep   TA Abui *iti*; HA Modole *idu*, Loda *adu*; BO Siwai, Telei *at*, Nasioi, Koianu *asi*; CM Savo *izi*, Baniata *isia*; TS Middle East, South East *ur*; WNG Mor *ute*; NNG Murusapa *isiatei* 'it lies'; SNG Kunini *ute*, Miriam *ut*, Turama *utua* 'to lie', Era R. *utaa* 'lie', Mawata *utuwa* 'lie down', Kiwai, Domori, Sisiame, Pirupiru *utua* 'lie down', CNG Dem *utawe*, Medlpa *ur* 'sleep (n.)'; NENG Saker *us* 'to lie'; ENG Onjob *utan*, Wadewinda *itun*, Mafulu *ito* 'to lie down'; UNG Kamoro *ete*, Asmat *isi es* (verbal noun + verb)

68. small   AN Biada *ketia*, Bogijiab *ketawa*; NB Baining *kitua*; TS Middle East *kaita*, *kita*

69. to smell   TA Oirata *muee*, Makasai *qamuhu* 'to stink'; HA Tabaru, Loda *ame*; NB Sulka *mmi*, Uasi *mua* 'to breathe'; NNG Maprik, Kwusaun, Yengoru, Yelogu *yamʌ*; SWNG Kaeti *umo*, Gawir *imu* 'smell (n.)'. Tirio *im(siga)* (n.), *im(tere)* (v.)

70. to stand   TS South East *pegera*, West *pegere*; CNG Naga, Mape *pakale*, Wamola *fahale*; NENG Vanembere *figiru*

71. star   TA Bunak *bi*, Makasai *fi(bɛrɛ)*, Oirata *ihi*; CM Reef *(n)ivu*,
                  Banua *vei*, Nea *vui*, Noole *bei*, Nabalue *pupi*; TS North
                  East *poe*; WNG Madik, Karon *bi*, Karas *pop* 'moon';
                  NNG Waris *pai*, Maprik, Wosera *[m]bapmu*, Nyaura
                  [m]bwap, Manambu *[m]bapu*, Kwusaun *[m]bapwʌ*,
                  Yengoru *[m]bwavwʌ*, Yelogu *[m]bwapwʌ*; SWNG Tirio
                  opapa; SNG Karima *baiba* 'moon'; Ro *bafa* 'moon', Karami
                  bube, Ipikoi *opopu*, Urama *piu, biu*, Era R. *piu*, Mahigi *bubel*
                  'moon'; CNG Tsaga, Tjiranki *bui*, Gende *mbei*, Kuman,
                  Sinasina, Dom, Boumai *ba* 'moon', Kate *bɔping*, Mape
                  babang, Womola *bɔpəng*, Magobineng, Sene *bɔbing*; NENG
                  Langtub *boai*, Bongu *baing*, Tombenam *pop*, Bogadjim, *boi*
                  'evening star'; ENG Barai *boio*; UNG Tsinyadji *bai*, Aurama
                  boe 'moon'

72. to steal   TA Abui *taluk*; HA Tobelo *tolik*

73. stone₁   BO Siwai *kubuli*, Nasioi *kapang*; CM Savo *ko[m]borei*; NNG
                  Monumbo *ikupul*; SWNG Tirio *guma*, Konmak *kum*; SNG
                  Jei *kumu*, Dugeme *kamu*, Karima *kamo*, Foraba, Sesa *kabu*,
                  Ro, Ibukairi *kabo*, Bara *kapu*, Karami *akabu*; CNG Aförö
                  ka[m]bö, Chimbu *kombuglo*, Dom *kobule*, Sinasina *kovile*,
                  Boumai *hobure*, Kuno *kimp-* 'mountain'; ENG Boli *gebiro*,
                  Musa R. *gembiro, gambiro*, Mailu, Domara *gomana*, Dom
                  gamu, Merani *goma*, Adaua *gibiri*, Mafulu *kume*, Kambisa
                  kumo 'mountain', Suku *kume* 'mountain'; UNG Koassa *kovi*
                  'mountain, stone', Ondoro *kambe* 'mountain'

74. stone₂   HA Ternate, Tidor *mare*, Pagu, Tololiku *mamaling*, Isam
                  mamalin; BO Siwai, Telei *menu* 'mountain', Nasioi *minani*;
                  WNG Mantion *meni* 'mountain', Manikion *meinyi* 'moun-
                  tain'; NNG Kamtuk *məndüng, mudung* 'mountain', Kuangsu
                  məndung 'mountain', Waibron Bano-Demenggong
                  mendung 'mountain'; SNG Ipikoi *meno*; CNG Fore *ʌmuni*
                  'mountain', Nangamb *maar*, Kuno *muro*, NENG Burumana,
                  Male *mening*, Koliku *meneng*, Panim *menin*, Nupanob
                  man, Botelkude *men*, Bogadjim *mening*, Sungumana *man*
                  'mountain'; ENG Gosisi, Agi, Kagi, Hagari, Uberi, Koiari,
                  Maiari, Koita *muni*, Suku *mune*, Sogeri *muna*, Barai *manu*,
                  Nigubaiba *umari*, Kokila *umare*; UNG Kamoro *omani*, Tate
                  mena 'hill'

75. thing   AN Biada *min*, Bale *ming*; CM Laumbe *mina*, Reef
                  (kele)mengge 'this' = '(this) thing'; NNG Nyaura *məndʌ*

'thing, what', Arapesh *mane* 'what?'; SWNG Kati *man* 'something'; CNG Matap *mina* 'what?'

76. tongue   BO Siwai *mini*, Nasioi *meneng*, Baitsi, Koianu *mene*, Telei *mere*; TS (all dialects) *mena*; WNG Kalabra *men*, Tehit *mar*; NNG Monumbo *menɛp*, Angoram *menöng*, Skofro *mer*; SNG Hibaradai *men*, Miriam *mer* 'language', Kanum *imen*, Jei *amar*; CNG Pesechem *mere*, Ndani *mele*, Oeringoep *amela*; NENG Langtub *muni*, Bongu *muin*, Sungumana *mien*, Bogadjim *mieng*, Em *melal*, Saker *mule*; ENG Manukolu *manane*, Moikoidi, Kororo, Neme, Doriaidi *meana*, Boli *miana*, Dom, Merani *maina*; UNG Kamoro, Mimika *mare*, Sempan *omane*, Asmat *komen*, Kauwerawet *men*

77. tooth   AN Önge *ku, kwe*; TA Bunak *igɔ*; BO Baitsi *kai-i*, Siwai *kewi*; TS Middle East *kayi*; NNG Ngala *akʌ* 'to bite', Manambu *wuk*; CNG Mid Grand Valley Dani, Sawuri-Hablifuri, etc. *(na)ik* '(my) tooth', Ekari *ego*, Naga, Ono, Kate *ki* 'to bite', Sene *eke* 'to bite', Sua *ki*; ENG Maiari *gi*, Sogeri etc. *egi*, Mulaha *kui* 'bites'; UNG Williams R. *ke*

78. underneath   HA Tobelo *timi*; NNG Kwesten *tum* 'earth'; SNG Dibiri *toma* 'earth'; CNG Gende *teme*, Fasu *tomo*, Sene *dome*

79. urine   AN Biada *ulu*, Kede *wile*, Chariar *ili*; TA Oirata *iri* 'urine, excrement' WNG Etna Bay, Mairasi *iri* 'excrement' CNG Mombum *ir*, Koneraw *iyere*

80. to walk   AN Önge *cige* 'leg', Biada, Bogijiab *cag* 'leg', Puchikwar, Juwoi *cok* 'leg' TA Abui *tuku* 'leg, foot'; HA Pagu, Modole, Galela, Tabaru, Loda, Ternate *tagi*; CM Savo *tetegha* 'foot, lower leg'; TS Northeast *tage(na)*, North *taka(ri)*; NNG Arso *taka* 'foot', Sko *tae* 'foot'; SWNG Marind *tagu*, Telefol *tek* 'to go, singular subject'; SNG Kanum, Moraori *tegu* 'foot' (possible borrowing from Marind), Mombum *itögh* 'foot', Bara *togoi* 'leg'; CNG Mid Grand Valley Dani *(ne)sok* '(my) foot', Jalimo Anggapuruk *(nu)juk* '(my) foot' (similar forms in other Dani dialects), Ekari, Wolani *togo*, Mikaru *saga* 'foot', Matap *tag* 'hip'; ENG Mafulu *soge* 'foot', Kambisa *suga* 'foot', Korona *sogo* 'foot', Kabana *suge* 'foot', Sikube *suku* 'foot', Hunjara, Jegasa Sarau *tegi*

'foot'. (For the semantics compare Gresi *masi* 'leg', *masimasi* 'to walk'.)

81. when?   AN Biada *ten*, Bale *tan*; TA Bunak *twen*, Alor *tena*
82. white   NB Uasi *keakea*; NNG Monumbo *kɛakɛak*, Ampas *keke*; SNG Domori, Sisiame *keakea*, Pirupiru, Buniki, Mawata, Kiwai *kea*; CNG Matap *kakaia* 'light (n.)' NENG Ulingan *kia*; ENG Domara *gaia*
83. wing   NB Baining *ighivaret(ki)*; CM Savo *gavara* (Grace, *ghavara*), Baniata *gofona*
84. yellow   HA Tobelo *kurati, gogurati*, Modelo *qoqurati*, Pagu *kulati*; NB Uasi *kuakuala*; BO Baitsi *kakarasa*; NNG Demta *kɛkɛr* 'green', Murik *kuukuur*, Siaute *kala*; CNG Bulum *korotkorot*; ENG Mailu *korakora* 'green', Binandere *korakorara* 'green'

12.1 Appendix: A comparison of Greenberg's and Wurm's classifications of the non-Austronesian, non-Australian languages of Oceania, by Timothy Usher

These tables compare Greenberg's Indo-Pacific family (§12) to the classification of Papuan languages in Wurm (1982). The comparison is divided into two tables. The first gives Greenberg's Indo-Pacific classification. Greenberg's lowest groups are compared to the closest low-level group in Wurm's classification; these are separated by the solid line. A group label of the form ′X/Y′ in the Wurm classification indicates that Greenberg' lowest group is a language (Y) that Greenberg lists separately from other members of Wurm's group (X). The second table gives Greenberg's lower level groups, the Wurm counterparts, and where they fit into Wurm's classification. The position of the Wurm group in Wurm's overall classification is indicated by the affiliations appearing to the right. The rows with the lowest-level groups are numbered so that the two tables can be compared directly (groups are given in the same order in both tables). The tables were prepared by Timothy Usher.

Stephen A. Wurm. 1982. *The Papuan Languages of Oceania*. Tübingen: Gunter Narr.

Abbreviations

CSNG   Central and South New Guinea
TAP   Timor-Alor-Pantar

TABLE 1

	Greenberg: Indo-Pacific			Wurm
1	Andamanese	Andamanese	Andamanese	[not included]
2	West Indo-Pacific	N. Halmaheran	N. Halmaheran	N. Halmaheran
3		Timor	Timor	Timer-Alor-Pantar
4		West New Guinea	WNG-1 [Meninggo etc.]	Kebar (=Mpur)
5			WNG-1 [Meninggo etc.]	E. Bird's Head
6			WNG-1 [Meninggo etc.]	S. Bird's Head
7			WNG-1 [Meninggo etc.]	Mor
8			WNG-1 [Meninggo etc.]	Mairasi-Tanah Merah
9			WNG-2 [Madik etc.]	W. Brid's Head
10			WNG-2 [Madik etc.]	N. Bird's Head
11			WNG-3 [Ajamaru etc.]	C. Bird's Head
12			[not listed]	Borai-Hattam
13			WNG-4 [Kapaur etc.]	West Bomberai
14		unclassified (?WNG)	Yava	Yava
15			[not listed]	E. Geelvink Bay
16		unclassified (?WNG)	Kamoro	Asmat-Kamoro
17		unclassified (?WNG)	Juliana River	Kayagar
18		unclassified	Mamberamo (1) [Koassa, Kauwerawet]	Northern (Kwerba)
19		unclassified	Mamberamo (1) [Goliath Mtn]	Mek
20		unclassified (?WNG)	Mamberamo (2) [Aikwakai, Tori, Tori Aikwakai, Südfluss . . .]	C. Lakes Plain
21		unclassified (?WNG)	Mamberamo (2) [Pauwi]	Warenbori
22		unclassified (?WNG)	Mamberamo (2) [Boromessu]	Taurap
23	Nuclear New Guinea	North New Guinea	NNG-1: Sentani	Sentani
24			NNG-1: Sentani	Nimboran
25			NNG-1: Sentani	Tor
26			NNG-1: Sentani	Mawes
27			NNG-1: Sentani	Kaure
28			NNG-2: Tami	Border
29			NNG-2: Tami	Vanimo
30			NNG-3: Arapeshan	Torricelli/various families
31			NNG-4: Murik	Nor-Pondo
32			NNG-5: Monumbo	Monumbo
33			NNG-5: Monumbo	Taiap (=Gapun)
34			NNG-5: Monumbo	Ramu
35			Ndu-Kwoma	Sepik
36			NNG unclassified: Nori (=Warapu)	Krisa
37			NNG unclassified: Siaute (=Olo)	Wapei/Olo
38			NNG unclassified: Apris	[unidentified]
39		unclassified (?NNG)	Senagi-Komberatoro	Senagi
40		unclassified (?NNG)	Fas	Kwomtari
41		unclassified (?NNG)	Amto	Amto-Musian
42		unclassified (?NNG)	Busa	Busa
43		unclassified (?NNG)	Yuri	Yuri
44			[not listed]	Pauwasi
45			[not listed]	Nagatman
46			[not listed]	Arai
47		Southwest New Guinea	SWNG-1: Tirio	Tirio
48			SWNG-2: Marind (a)	Marind proper
49			SWNG-2: Marind (b)	Boazi
50			SWNG-2: Marind (c)	Yaqay
51			SWNG-3: Ok (a) [except Awin]	Lowland Ok
52			SWNG-3: Ok (b) [except Plain Country]	Mountain Ok
53			SWNG-3: Ok (a/b) [Awin, Plain Country]	Awin

TABLE 1 (*continued*).

	Greenberg: Indo-Pacific			Wurm
54	(Nuclear New Guinea)		SWNG-3: Ok (c)	Dumut
55			SWNG-4: Awju	Awyu
56			SWNG-5: Kukukuku	Angan
57			[not listed]	Oksapmin
58			[not listed]	Bosavi-E. Strickland
59		South New Guinea	SNG-1: Kiwai	Kiwai
60			SNG-1: Kiwai	Inland Gulf
61			SNG-1: Kiwai [Hiwi, Hibaridai, Tapapi]	Pahoturi/Waia
62			SNG-2: Barika	Turama-Kikorian
63			SNG-2: Barika	Teberan-Pawaian
64			SNG-3: Kunini	E. Trans-Fly
65			SNG-4: Bugi	Pahoturi
66			SNG-5: Parb	Nambu
67			SNG-6: Peremka	Tonda
68			SNG-6: Peremka [Jiminakana, Nausaku]	Suki
69			SNG-7: Jei	Yey
70			SNG-7: Jei	Tonda/Kanum [also Ngowugar]
71			SNG-7: Jei	Moraori
72			SNG-7: Jei	Mombum
73			SNG-7: Jei	Kolopom
74			SNG-7: Jei	Bulaka River
75		unclassified (?SNG)	Gogodala	Gogodala
76	Nuclear New Guinea	Central New Guinea	CNG-2: Kapauku-Baliem (a)	Ekagi-Wodani-Moni
77			CNG-2: Kapauku-Baliem (b)	Dem
78			CNG-2: Kapauku-Baliem (c)	Uhunduni
79			CNG-2: Kapauku-Baliem (d)	Southern (Dani)
80			CNG-2: Kapauku-Baliem (e)	Southern (Ngalik-Nduga)
81			CNG-2: Highland (a)	Eastern
82			CNG-2: Highland (b)	East-Central
83			CNG-2: Highland (c)	Central
84			CNG-2: Highland (d)	West-Central
85			CNG-2: Highland (d)	Wiru
86			CNG-2: Highland (e)	Kalam
87			CNG-2: Highland (f)	Kutubuan
88			CNG-3: Huon [incl. Finisterre]	Finisterre-Huon
89		East New Guinea	ENG-1: Mailu	Mailuan
90			ENG-1: Mailu	Yareban
91			ENG-2: Binandere	Binanderean
92			ENG-3: Dimuga	Dagan
93			ENG-4: Elema	Eleman
94			ENG-5: Kovio	Goilalan
95			ENG-6: Afoa	Goilalan
96			ENG-7: Fuyuge	Goilalan
97			ENG-8: Mulaha	Kwalean
98			ENG-9: Koita	Koiarian
99			ENG-10: Namau	Purari
100			[not listed]	Manubaran
101	Northeast New Guinea	Northeast New Guinea	Northeast New Guinea	Madang-Adelbert
102	Pacific	Pacific	New Britain	New Britain
103			Panaras	New Britain
104			Uasi	New Britain
105			C. Solomons	C. Solomons
106			[not listed]	Kazukuru
107			Bougainville	Bougainville
108			Santa Cruz	Reef-Santa Cruz
109	unassigned	unassigned	Rossel Island	Yele
110	Tasmanian	Tasmanian	Tasmanian	[not included]

Table 2

Greenberg: Indo-Pacific	Wurm		
1	Andamanese	[not included]		
2	N. Halmaheran	N. Halmaheran	N. Halmaheran	N. Halmaheran
3	Timor	Timor-Alor-Pantar	Timor-Alor-Pantar	Timor-Alor-Pantar
4	WNG-1 [Meninggo etc.]	Kebar (= Mpur)	Kebar (= Mpur)	Kebar (= Mpur)
5	WNG-1 [Meninggo etc.]	E. Bird's Head	E. Bird's Head	E. Bird's Head
6	WNG-1 [Meninggo etc.]	S. Bird's Head	S. Bird's Head	S. Bird's Head
7	WNG-1 [Meninggo etc.]	Mor	Mor	Mor
8	WNG-1 [Meninggo etc.]	Mairasi-Tanah Merah	Mairasi-Tanah Merah	Mairasi-Tanah Merah
9	WNG-2 [Madik etc.]	W. Bird's Head	W. Bird's Head	W. Bird's Head
10	WNG-2 [Madik etc.]	N. Bird's Head	N. Bird's Head	N. Bird's Head
11	WNG-3 [Ajamaru etc.]	C. Bird's Head	C. Bird's Head	C. Bird's Head
12	[not listed]	Borai-Hattam	Borai-Hattam	Borai-Hattam
13	WNG-4 [Kapaur etc.]	W. Bomberai	W. Bomberai	West Bomberai
14	Yava	Yava	Yava	Yava
15	[not listed]	E. Geelvink Bay	E. Geelvink Bay	E. Geelvink Bay
16	Kamoro	Asmat-Kamoro	Asmat-kamoro	C. & S. New Guinea
17	Juliana River	Kayagar	Kayagar	Kayagar
18	Mamberamo (1) [Koassa, Kauwerawet]	Northern (Kwerba)	Dani-Kwerba	Dani-Kwerba
19	Mamberamo (1) [Goliath Mtn]	Mek	Mek	Mek
20	Mamberamo (2) [Aikwakai, Tori, Tori Aikwakai, Südfluss . . .]	C. Lakes Plain	C. Lakes Plain	C. Lakes Plain
21	Mamberamo (2) [Pauwi]	Warenbori	Warenbori	Warenbori
22	Mamberamo (2) [Boromessu]	Taurap	Taurap	Taurap
23	NNG-1: Sentani	Sentani	Sentani	Sentani
24	NNG-1: Sentani	Nimboran	Nimboran	Nimboran
25	NNG-1: Sentani	Tor	Tor	Tor
26	NNG-1: Sentani	Mawes	Mawes	Mawes
27	NNG-1: Sentani	Kaure	Kaure	Kaure
28	NNG-2: Tami	Border	Border	Border
29	NNG-2: Tami	Vanimo	Vanimo	Vanimo
30	NNG-3: Arapheshan	Torricelli/var. families	Torricelli	Torricelli
31	NNG-4: Murik	Nor-Pondo	Nor-Pondo	Nor-Pondo
32	NNG-5: Monumbo	Monumbo	Torricelli	Torricelli
33	NNG-5: Monumbo	Taiap (= Gapun)	Sepik-Ramu	Sepik-Ramu
34	NNG-5: Monumbo	Ramu	Ramu	Ramu

Table 2 (*continued*).

Wurm continued

1	[not included]			
2	N. Halmaheran	N. Halmaheran	N. Halmaheran	W. Papuan
3	Timor-Alor-Pantar	Timor-Alor-Pantar	S. Bird's Head-TAP	Trans-New Guinea
4	Kebar (= Mpur)	Kebar (= Mpur)	Kebar (= Mpur)	W. Papuan
5	E. Bird's Head	E. Bird's Head	E. Bird's Head	E. Bird's Head
6	S. Bird's Head	S. Bird's Head	S. Bird's Head-TAP	Trans-New Guinea
7	Mor	Mor	Mor	Trans-New Guinea
8	Mairasi-Tanah Merah	Central & Western	Main	Trans-New Guinea
9	W. Bird's Head	W. Bird's Head	Bird's Head	W. Papuan
10	N. Bird's Head	N./C. Bird's Head	Bird's Head	W. Papuan
11	C. Bird's Head	N./C. Bird's Head	Bird's Head	W. Papuan
12	Borai-Hattam	Borai-Hattam	Borai-Hattam	W. Papuan
13	West Bomberai	Central & Western	Main	Trans-New Guinea
14	Yava	Yava	Yava	Geelvink Bay
15	E. Geelvink Bay	E. Geelvink Bay	E. Geelvink Bay	Geelvink Bay
16	C. & S. New Guinea-Kutubuan	Central & Western	Main	Trans-New Guinea
17	Kayagar	Central & Western	Main	Trans-New Guinea
18	Dani-Kwerba	Central & Western	Main	Trans-New Guinea
19	Mek	Mek	Mek	Trans-New Guinea
20	C. Lakes Plain	Tor-Lakes Plain	Northern	Trans-New Guinea
21	Warenbori	Warenbori	Warenbori	isolate
22	Taurap	Taurap	Taurap	isolate
23	Sentani	Central & Western	Main	Trans-New Guinea
24	Nimboran	Nimboran	Nimboran	Trans-New Guinea
25	Tor	Tor-Lakes Plain	Northern	Trans-New Guinea
26	Mawes	Tor-Lakes Plain	Northern	Trans-New Guinea
27	Kaure	Kaure	Kaure	Trans-New Guinea
28	Border	Border	Northern	Trans-New Guinea
29	Vanimo	Vanimo	Vanimo	Sko
30	Torricelli	Torricelli	Torricelli	Torricelli
31	Nor-Pondo	Nor-Pondo	Nor-Pondo	Sepik-Ramu
32	Torricelli	Torricelli	Torricelli	Torricelli
33	Sepik-Ramu	Sepik-Ramu	Sepik-Ramu	Sepik-Ramu
34	Ramu	Ramu	Ramu	Sepik-Ramu

TABLE 2 (*continued*).

Greenberg: Indo-Pacific	Wurm		
35	Ndu-Kwoma	Sepik	Sepik	Sepik
36	NNG unclassified: Nori (= Warapu)	Krisa	Krisa	Krisa
37	NNG unclassified: Siaute (= Olo)	Wapei/Olo	Torricelli	Torricelli
38	NNG unclassified: Apris	[unidentified]	[unidentified]	[unidentified]
39	Senagi-Komberatoro	Senagi	Senagi	Senagi
40	Fas	Kwomtari	Kwomtari	Kwomtari
41	Amto	Amto-Musian	Amto-Musian	Amto-Musian
42	Busa	Busa	Busa	Busa
43	Yuri	Yuri	Yuri	Yuri
44	[not listed]	Pauwasi	Pauwasi	Pauwasi
45	[not listed]	Nagatman	Nagatman	Nagatman
46	[not listed]	Arai	Arai	Arai
47	SWNG-1: Tirio	Tirio	Tirio	Tirio
48	SWNG-2: Marind (a)	Marind proper	Marind proper	Marind proper
49	SWNG-2: Marind (b)	Boazi	Boazi	Boazi
50	SWNG-2: Marind (c)	Yaqay	Yaqay	Yaqay
51	SWNG-3: Ok (a) [except Awin]	Lowland Ok	Ok	C. & S. New Guinea
52	SWNG-3: Ok (b) [except Plain Country]	Mountain Ok	Ok	C. & S. New Guinea
53	SWNG-3: Ok (a/b) [Awin, Plain Country]	Awin	Awin-Pa	C. & S. New Guinea
54	SWNG-3: Ok (c)	Dumut	Awyu-Dumut	C. & S. New Guinea
55	SWNG-4: Awju	Awyu	Awyu-Dumut	C. & S. New Guinea
56	SWING-5: Kukukuku	Angan	Angan	Angan
57	[not listed]	Oksapmin	Oksapmin	Oksapmin
58	[not listed]	Bosavi-E. Strickland	Bosavi-E. Strickland	C. & S. New Guinea
59	SNG-1: Kiwai	Kiwai	Kiwai	Kiwai
60	SNG-1 Kiwai	Inland Gulf	Inland Gulf	Inland Gulf
61	SNG-1: Kiwai [Hiwi, Hibaridai, Tapapi]	Pahoturi/Waia	Pahoturi/Waia	Pahoturi/Waia
62	SNG-2: Barika	Turama-Kikorian	Turama-Kikorian	Turama-Kikorian
63	SNG-2: Barika	Teberan-Pawaian	Teberan-Pawaian	Teberan-Pawaian
64	SNG-3: Kunini	E. Trans-Fly	E. Trans-Fly	E. Trans-Fly
65	SNG-4: Bugi	Pahoturi	Pahoturi	Pahoturi
66	SNG-5: Parb	Nambu	Morehead-Upper Maro	Morehead-Upper Maro
67	SNG-6: Peremka	Tonda	Morehead-Upper Maro	Morehead-Upper Maro
68	SNG-6: Peremka [Jiminakana, Nausaku]	Suki	Gogodala-Suki	Gogodala-Suki
69	SNG-7: Jei	Yey	Morehead-Upper Maro	Morehead-Upper Maro
70	SNG-7: Jei	Tonda/Kanum [also Ngowugar]	Morehead-Upper Maro	Morehead-Upper Maro
71	SNG-7: Jei	Moraori	Morehead-Upper Maro	Morehead-Upper Maro
72	SNG-7: Jei	Mombum	Mombum	C. & S. New Guinea
73	SNG-7: Jei	Kolopom	Kolopom	Kolopom
74	SNG-7: Jei	Bulaka River	Bulaka River	Bulaka River
75	Gogodala	Gogodala	Gogodala-Suki	Gogodala-Suki

Table 2 (*continued*).

Wurm continued

35	Sepik	Sepik	Sepik	Sepik-Ramu
36	Krisa	Krisa	Krisa	Sko
37	Torricelli	Torricelli	Torricelli	Torricelli
38	[unidentified]	[unidentified]	[unidentified]	[unidentified]
39	Senagi	Senagi	Senagi	Trans-New Guinea
40	Kwomtari	Kwomtari	Kwomtari	Kwomtari-Baibai
41	Amto-Musian	Amto-Musian	Amto-Musian	Amto-Musian
42	Busa	Busa	Busa	isolate
43	Yuri	Yuri	Yuri	isolate
44	Pauwasi	Pauwasi	Pauwasi	Trans-New Guinea
45	Nagatman	Nagatman	Nagatman	isolate
46	Arai	Arai	Arai	Arai
47	Tirio	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
48	Marind	Central & Western	Main	Trans-New Guinea
49	Marind	Central & Western	Main	Trans-New Guinea
50	Marind	Central & Western	Main	Trans-New Guinea
51	C. & S. New Guinea-Kutubuan	Central & Western	Main	Trans-New Guinea
52	C. & S. New Guinea-Kutubuan	Central & Western	Main	Trans-New Guinea
53	C. & S. New Guinea-Kutubuan	Central & Western	Main	Trans-New Guinea
54	CSNG-Kutubuan	Central & Western	Main	Trans-New Guinea
55	CSNG-Kutubuan	Central & Western	Main	Trans-New Guinea
56	Angan	Central & Western	Main	Trans-New Guinea
57	Oksapmin	Oksapmin	Oksapmin	Trans-New Guinea
58	CSNG-Kutubuan	Central & Western	Main	Trans-New Guinea
59	Kiwai	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
60	Inland Gulf	Inland Gulf	Inland Gulf	Trans-New Guinea
61	Pahoturi/Waia	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
62	Turama-Kikorian	Turama-Kikorian	Turama-Kikorian	Trans-New Guinea
63	Teberan-Pawaian	Teberan-Pawaian	Teberan-Pawaian	Trans-New Guinea
64	E. Trans-Fly	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
65	Pahoturi	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
66	Morehead-Upper Maro	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
67	Morehead-Upper Maro	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
68	Gogodala-Suki	Central & Western	Main	Trans-New Guinea
69	Morehead-Upper Maro	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
70	Morehead-Upper Maro	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
71	Morehead-Upper Maro	Trans Fly	Trans-Fly-Bulaka River	Trans-New Guinea
72	CSNG-Kutubuan	Central & Western	Main	Trans-New Guinea
73	Kolopom	Kolopom	Kolopom	Trans-New Guinea
74	Bulaka River	Bulaka River	Trans-Fly-Bulaka River	Trans-New Guinea
75	Gogodala-Suki	Central & Western	Main	Trans-New Guinea

Table 2 (*continued*).

	Greenberg: Indo-Pacific	Wurm		
76	CNG-2: Kapauku-Baliem (a)	Ekagi-Wodani-Moni	Wissel Lakes-Kemandogha	Wissel Lakes-Kemandogha
77	CNG-2: Kapauku-Baliem (b)	Dem	Dem	Dem
78	CNG-2: Kapauku-Baliem (c)	Uhunduni	Wissel Lakes-Kemandogha	Wissel Lakes-Kemandogha
79	CNG-2: Kapauku-Baliem (d)	Southern (Dani)	Dani-Kwerba	Dani-Kwerba
80	CNG-2: Kapauku-Baliem (e)	Southern (Ngalik-Nduga)	Dani-Kwerba	Dani-Kwerba
81	CNG-2: Highland (a)	Eastern	Eastern	Eastern
82	CNG-2: Highland (b)	East-Central	East-Central	East-Central
83	CNG-2: Highland (c)	Central	Central	Central
84	CNG-2: Highland (d)	West-Central	West-Central	West-Central
85	CNG-2: Highland (d)	Wiru	Wiru	Wiru
86	CNG-2: Highland (e)	Kalam	Kalam	Kalam
87	CNG-2: Highland (f)	Kutubuan	Kutubuan	Kutubuan
88	CNG-3: Huon [incl. Finisterre]	Finisterre-Huon	Finisterre-Huon	Finisterre-Huon
89	ENG-1: Mailu	Mailuan	Mailuan	Mailuan
90	ENG-1: Mailu	Yareban	Yareban	Yareban
91	ENG-2: Binandere	Binanderean	Binanderean	Binanderean
92	ENG-3: Dimuga	Dagan	Dagan	Dagan
93	ENG-4: Elema	Eleman	Eleman	Eleman
94	ENG-5: Kovio	Goilalan	Goilalan	Goilalan
95	ENG-6: Afoa	Goilalan	Goilalan	Goilalan
96	ENG-7: Fuyuge	Goilalan	Goilalan	Goilalan
97	ENG-8: Mulaha	Kwalean	Kwalean	Kwalean
98	ENG-9: Koita	Koiarian	Koiarian	Koiarian
99	ENG-10: Namau	Purari	Eleman	Eleman
100	[not listed]	Manubaran	Manubaran	Manubaran
101	Northeast New Guinea	Madang-Adelbert	Madang-Adelbert	Madang-Adelbert
102	New Britain	New Britain	New Britain	Yele-Solomons-New Britain
103	Panaras	New Britain	New Britain	Yele-Solomons-New Britain
104	Uasi	New Britain	New Britain	Yele-Solomons-New Britain
105	C. Solomons	C. Solomons	Yele-Solomons	Yele-Solomons-New Britain
106	[not listed]	Kazukuru	Yele-Solomons	Yele-Solomons-New Britain
107	Bougainville	Bougainville	Bougainville	Bougainville
108	Santa Cruz	Reef-Santa Cruz	Reef-Santa Cruz	Reef-Santa Cruz
109	Rossel Island	Yele	Yele-Solomons	Yele-Solomons-New Britain
110	Tasmanian	[not included]		

TABLE 2 (*continued*).

Wurm continued

76	Wissel Lakes-Kemandogha	Central & Western	Main	Trans-New Guinea
77	Dem	Dem	Dem	Trans-New Guinea
78	Wissel Lakes-Kemandogha	Central & Western	Main	Trans-New Guinea
79	Dani-Kwerba	Central & Western	Main	Trans-New Guinea
80	Dani-Kwerba	Central & Western	Main	Trans-New Guinea
81	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
82	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
83	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
84	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
85	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
86	E. New Guinea Highlands	Central & Western	Main	Trans-New Guinea
87	CSNG-Kutubuan	Central & Western	Main	Trans-New Guinea
88	Finisterre-Huon	Central & Western	Main	Trans-New Guinea
89	Mailuan	Eastern	Main	Trans-New Guinea
90	Yareban	Eastern	Main	Trans-New Guinea
91	Binanderean	Eastern	Main	Trans-New Guinea
92	Dagan	Eastern	Main	Trans-New Guinea
93	Eleman	Eleman	Eleman	Trans-New Guinea
94	Goilalan	Eastern	Main	Trans-New Guinea
95	Goilalan	Eastern	Main	Trans-New Guinea
96	Goilalan	Eastern	Main	Trans-New Guinea
97	Kwalean	Eastern	Main	Trans-New Guinea
98	Koiarian	Eastern	Main	Trans-New Guinea
99	Eleman	Eleman	Eleman	Trans-New Guinea
100	Manubaran	Eastern	Main	Trans-New Guinea
101	Madang-Adelbert	Madang-Adelbert	Madang-Adlbert	Trans-New Guinea
102	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
103	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
104	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
105	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
106	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
107	Bougainville	Bougainville	Bougainville	East Papuan
108	Reef-Santa Cruz	Reef-Santa Cruz	Reef-Santa Cruz	East Papuan
109	Yele-Solomons-New Britain	Yele-Solomons-New Britain	Yele-Solomons-New Britain	East Papuan
110	[not included]			

References

Allen, Jerry, and Hard, Conrad. [n. d.] *Languages of the Bougainville District*. Summer Institute of Linguistics (survey carried out in 1963).

Anceaux, J. C. 1958. 'Languages of the Bomberai Peninsula, outline of a linguistic map', *Nieuw-Guinea Studiën* 2: 109–20.

——1961. *The Linguistic Situation on the Islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. 's-Gravenhage: Martinus Nijhoff.

——1965. *The Nimboran Language. Phonology and Morphology*. 's-Gravenhage, Martinus Nijhoff.

Anon. 1914–15. 'Vocabularies of Binandele, Yema Yarawe, Mawai, Tahari and Aiga', *Papua Annual Report* 1914–15.

——1917–18. 'Vocabularies of A, Kambesi, etc.', *Papua Annual Report* 1917–18.

——1922–3. 'Vocabulary of Apris'. *Papua Annual Report* 1922–3.

Baarda, M. J. van. 1891. *Beknopte Spraakkunst van de Galilareesch Taal*. Utrecht.

——1895. *Woordenlijst Galilareesch-Hollandsch*. 's-Gravenhage: Martinus Nijhoff.

—— 1904. 'Het Loda'sch', *Bijdragen tot de Taal-, Land- en Volkenkunde* 56: 317–496.

Bateson, Gregory. 1936. *Naven*. Stanford: Stanford University Press.

Bee, Darlene and Glasgow, Kathleen Barker. 1962. 'Usarufa tone and segmental phonemes', *Oceania Linguistic Monographs* 6: 111–27.

Berthe, L. 1959. 'Sur quelques distiques Buna' (Timor Central)', *Bijdragen tot de Taal-, Land- en Volkenkunde* 115: 336–71.

Biggs, Bruce. 1963. 'A non-phonemic central vowel type in Karam, a "Pygmy" language of the Schrader Mountains, Central New Guinea', *Anthropological Linguistics* 5/4: 13–7.

Bromley, H. Myron. 1961. *The Phonology of Lower Grand Valley Dani*. 's-Gravenhage: Martinus Nijhoff.

Bromley, H. Myron. 1967. 'The linguistic relationships of Grand Valley Dani: A lexicostatistic classification', *Oceania* 38: 286–306.

Brown, A. R. 1914. 'Notes on the languages of the Andaman Islanders', *Anthropos* 9: 36–52.

Brown, L. W. 1918–19. 'Vocabularies of Nigubaiba, Iawarere, Deduri, Seramina, Jari and Pitoni', *Papua Annual Report* 1918–19.

Capell, A. 1954. *A Linguistic Survey of the Southwestern Pacific*. Noumea: South Pacific Commission.

Champion, I. 1926–7. 'Vocabularies of "Plain Country" and Unkia', *Papua Annual Report* 1926–7.

Chinnery, E. W. P. [n. d.] *Studies of the Native Population of the East Coast of New Ireland*. Territory of New Guinea Anthropological Report No. 6 [probably 1925].

Chisholm. 1914–15. 'Vocabulary of Williams River', *Papua Annual Report* 1914–15. [Name given only as Patrol Officer Chisholm.]

Cowan, H. K. J. 1957a. 'Prospects of a "Papuan" comparative linguistics', *Bijdragen tot de Taal-, Land- en Volkenkunde* 113: 70–91.

—— 1957*b*. 'Een tweede grote Papoea-taalgroepering in Nederlands-Nieuw-Guinea', *Nieuw-Guinea Studiën* 1: 107–17.

—— 1957*c*. 'A large Papuan language phylum in West New Guinea', *Oceania* 28: 159–66.

—— 1960. 'Nadere gegevens betreffende de Verbreiding der West-Papoease Taal-groep (Vogelkop, Nieuw-Guinea)', *Bijdragen tot de Taat, Land- en Volkenkunde* 116: 350–64.

—— 1963. 'Le Buna' de Timor: Une langue "Ouest-Papoue"', *Bijdragen tot de Taal, Land- en Volkenkunde* 119: 387–400.

—— 1965*a*. 'The Oirata language', *Lingua* 14: 360–70.

—— 1965*b*. *Grammar of the Sentani Language.* 's-Gravenhage: Martinus Nijhoff.

Davenport, William. 1962. 'Comment on A. Capell, *Oceanic Linguistics Today*', *Current Anthropology* 3: 400–2.

Davis, Donald. 1964. 'Wantoat verb stem classes and affixation', in Kenneth L. Pike (ed.), *Verb Studies in five New Guinea Languages*, Norman: Summer Institute of Linguistics, 131–80.

Deibler, Ellis Jr. 1964. 'The application of matrix theory to Gahuku verbs', *Linguistic Circle of Canberra—Series A*, No. 3: 17–26.

Deibler, Ellis Jr. and Trefry, David. 1963. *Languages of the Chimbu Sub-District.* Port Moresby: Dept. of Information and Extension Services.

de Kock, M. A. 1912. 'Eenige ethnologische en anthropologische gegevens omtrent een dwergstam in het bergland van zuid Nieuw-Guinea', *Tijdschrift der nederlandsch Aardrijkskundig Genootschap* 29: 154–70.

Doble, Marian. 1960. *Kapauku-Malayan-Dutch-English Dictionary.* The Hague: Martinus Nijhoff.

Drabbe, Petrus. 1953. *Talen en dialecten van Zuid-West Nieuw-Guinea.* (Micro-Bibliotheca Anthropos, 11.) Posieus (Freiburg): Anthropos Institut.

—— 1955. *Spraakkunst van het Marind.* Mödling-Wien: Mission House St Gabriel.

—— 1957. *Spraakkunst van het Aghu-dialect van de Awju-taal.* 's-Gravenhage: Martinus Nijhoff.

—— 1959. *Kaeti en Wambon, twee Awju-dialecten.* 's-Gravenhage: Martinus Nijhoff.

—— 1963. *Drie Asmat-dialecten.* 's-Gravenhage: Martinus Nijhoff.

Du Bois, Cora Alice. 1944. *The People of Alor.* Minneapolis: University of Minnesota Press.

Ellen, G. J. 1916*a*. 'Woordenlijst van het Pagoe op Noord-Halmahera', *Bijdragen tot de Taal-, Land- en Volkenkunde* 72: 66–102.

—— 1916*b*. 'Woordenlijst van het Modole op Noord-Halmahera', *Bijdragen tot de Taal-, Land- en Volkenkunde* 72: 103–40.

Fortgens, J. 1928. 'Grammatikale aantekeningen over het Tabaroesch, Tabaroesche volksverhalen en raadsels', *Bijdragen tot de Taal-, Land- en Volkenkunde* 54: 300–545.

—— 1930. 'Ternat'sche teksten', *Bijdragen tot de Taal-, Land- en Volkenkunde* 86: 216–301.

Franklin, Karl J. 1964. 'Kewa verb morphology', in Kenneth L. Pike (ed.), *Verb Studies in Five New Guinea Languages*,. Norman: Summer Institute of Linguistics, 100–30.

Franklin, Joyce. 1965. 'Kewa II: Higher level phonology', *Anthropological Linguistics* 7/5: 84–8.

Franklin, Karl J. and Franklin, Joyce. 1962. 'Kewa I: Phonological asymmetry', *Anthropological Linguistics* 4/7: 29–37.

Frantz, Chester. 1962. 'Grammatical categories as indicated by Gadsup noun affixes', *Oceania Linguistic Monographs* 6: 44–63.

Frantz, Chester and McKaughan, Howard. 1964. 'Gadsup independent verb affixes', in Kenneth L. Pike (ed.), *Verb Studies in Five New Guinea Languages*. Norman: Summer Institute of Linguistics, 84–99.

Futscher, Otto. 1959. *Taulil-Grammatik und naturwissenschaftliche Sammelarbeiten (Neubritannien, Südsee)*. (Micro-Bibliotheca Anthropos, Posieux (Freiburg): Anthropos Institut. 30.)

Gatti, Ricardo. 1906–8. *Studi sul gruppo linguistico Andamanese-Papua-Australiano*. 2 vols. Bologna: L. Beltrami.

Glasgow, David and Loving, Richard. 1964. *Languages of the Maprik Sub-district*. Port Moresby: Dept. of Information and Extension Services.

Grace, George W. 1968. 'Classification of the languages of the Pacific', in Andrew P. Vayda (ed.), *People and Cultures of the Pacific*. Garden City, New York: Natural History Press, 63–79.

Greenberg, Joseph H. 1958. Report on the classification of the non-Austronesian languages of the Pacific. MS.

—— 1960. Indo-Pacific etymologies. MS.

Healey, Alan. 1962. 'Linguistic aspects of Telefomin kinship terminology', *Anthropological Linguistics* 4/7: 14–28.

—— 1964. *Telefol Phonology*. Canberra: Australian National University.

Healey, Phyllis M. 1964. *Telefol Noun Phrases*. Canberra: Australian National University.

—— 1965*a*. *Telefol Verb Phrases*. Canberra: Australian National University.

—— 1965*b*. 'Teléfoól quotative clauses', *Linguistic Circle of Canberra—Series A*, 3: 27–34.

Herbert, C. L. 1914–15. 'Vocabularies of Goaribari, Namau, Urama, Turama and Era River', *Papua Annual Report* 1914–15, No. 24.

Höltker, G. 1964. 'Zur Sprache der Monumbo-Papua in Nordost Neuguinea', *Anthropos* 59: 860–903.

Hooley, Bruce A. 1964. 'The Morobe District—New Guinea', *Oceanic Linguistics* 3: 201–47.

Hueting, A. 1908. 'Iets over de 'Ternataansch-Halmaherasche' taalgroep', *Bijdragen tot de Taal-, Land- en Volkenkunde* 60: 369–411.

—— 1936. 'Iets over de spraakkunst van de Tobeloreesche taal', *Bijdragen tot de Taal-, Land- en Volkenkunde* 94: 259–408.

Josselin de Jong, J. P. B. 1937. *Studies in Indonesian Culture I. Oirata, a Timor Settlement on Kissar.* Amsterdam: Noord-Hollandsche uitgevers-maatschappij.

Kaberry, Phyllis M. 1941. 'The Abelam tribe, Sepik River District, New Guinea', *Oceania* 11: 233–58, 344–67.

—— 1941–42. 'Law and political organization in the Abelam tribe, New Guinea', *Oceania* 12: 79–95, 210–25, 331–63.

Klieneberger, H. R. 1957. *Bibliography of Oceanic Linguistics.* London: Oxford University Press.

Lanyon-Orgill, L. P. A. 1953. 'The Papuan language of the New Georgian Archipelago', *Journal Of Austronesian Studies* 1/1: 122–38.

Larson, Mildred O. and Gordon, F. 1958. 'Preliminary studies in the Moni language', *Bijdragen tot de Taal-, Land- en Volkenkunde* 114: 406–31.

Laufer, C. 1959. 'P. Futschers Aufzeichnungen über die Butam-Sprache (Neu-Britannien)', *Anthropos* 63: 199–212.

Laycock, D. C. 1965*a. The Ndu Language Family (Sepik District, New Guinea).* Canberra: Australian National University.

—— 1965*b.* 'Three Upper Sepik phonologies', *Oceanic Linguistics* 4: 113–16.

Loeweke, Eunice and May, Jean. 1965. 'The phonological hierarchy of Fasu', *Anthropological Linguistics* 7/5: 59–70.

—— 1966. 'Fasu grammar', *Anthropological Linguistics* 8/5: 17–33.

Loukotka, Čestmir. 1957. 'Classification des langues Papoues', *Lingua Posnaniensis* 6: 19–83.

Loving, Richard and Loving, Aretta. 1962. 'A preliminary survey of Awa noun suffixes', *Oceania Linguistic Monographs* 6: 28–43.

Loving, Richard and Bass, Jack. 1964. *The Languages of the Amanab Sub-District.* Port Moresby: Dept. of Information and Extension Services.

Loving, Richard and McKaughan, Howard. 1964*a.* 'Awa verbs Part I: The internal structure of independent verbs', in Kenneth L. Pike (ed.), *Verb Studies in Five New Guinea Languages.* Norman: Summer Institute of Linguistics, 1–30.

—— 1964*b.* 'Awa verbs Part II: The internal structure of dependent verbs', in Kenneth L. Pike (ed.), *Verb Studies in Five New Guinea Languages.* Norman, Summer Institute of Linguistics, 31–44.

Lucht, R., and James, D. 1962. 'The phonemes of Siane', *Te Reo* 5: 12–16.

McCarthy, Joy. 1965. 'Clause chaining in Kanite', *Anthropological Linguistics* 7/5: 59–70.

McElhanon, K. A. 1967. 'Preliminary observations on Huon Peninsula languages', *Oceanic Linguistics* 6: 1–45.

McKaughan, Howard. 1964. 'A study of divergence in four New Guinea languages', *American Anthropologist* 66/4 Part 2: 98–102.

—— 1966. 'Sequences of clauses in Tairora', *Oceanic Linguistics* 5: 1–12.

Nichols, Ruth and Nichols, Ray. 1962. 'Fore phonemes and their interpretation', *Oceania Linguistic Monographs* 6: 128–48.

Nicolspeyer, M. M. [n. d.] 'De sociale structuur van een aloreesche bevolkingsgroep' Rijswijk.

Ogan, Eugene. 1966. 'Nasioi marriage: An essay in model-building', *Southwest Journal of Anthropology* 22: 172–93.

Parlier, Jim. 1964. 'Managalasi verb inflection', *Te Reo* 7: 28–35.

Pawley, Andrew Kenneth. 1966. 'The structure of Karam' Ph.D. dissertation, University of Auckland.

Pence, Alan. 1964. 'Intonation in Kunimaipa (New Guinea)', *Linguistic Circle of Canberra-Series A*, No. 3.

Pike, Eunice V. 1964. 'The phonology of New Guinea Highlands languages', *American Anthropologist* 66: 121–32.

Pike, Kenneth L., and Scott, Graham. 1963. 'Pitch accent and non-accented phrases in Fore (New Guinea)'. *Zeitschrift für Phonetik, Sprachwissenschaft and Kommunikationsforschung* 16: 179–89.

Portman, M. V. 1887. *A Manual of the Andamanese Languages.* London: W. H. Allen and Co.

—— 1898. *Notes on the Languages of the South Andaman Group of Tribes.* Calcutta: Office of the Superintendant of Government Printing.

Roest, J. L. D. van der. 1905. *Woodenlijst der Tobelo-Boeng-Taal.* 's-Gravenhage: Martinus Nijhoff.

Rule, Joan. 1965. 'A comparison of certain phonemes of the languages of the Mendi and Nembi Valley Southern Highlands, Papua', *Anthropological Linguistics* 7/5: 98–105.

Salisbury, R. F. 1956. 'The Siane language', *Anthropos* 51: 447–80.

Salzner, Richard. 1960. *Sprachenatlas des Indopazifischen Raumes.* Wiesbaden: Otto Harrassowitz.

Schmidt, Wilhelm. 1952. *Die tasmanischen Sprachen.* Utrecht-Anvers: Spectrum.

Schmitz, Carl A. 1958. 'Die unzulänglich bekannten Papuasprachen nach L. Birós Aufzeichnungen', *Acta Ethnographica Academiae Scientiarum Hungaricae* 7: 433–8.

Schmitz, Carl A. 1960. *Beiträge zur Ethnographie des Wantoat Tales, Nordost Neuguinea.* Köln, Kölner Ethnologische Mitteilungen No. 1.

Schoorl, J. W. [n. d.] 'Kultuur en kultuurveranderingen in het Moejoe-Gebied', Ph.D. dissertation, University of Leiden.

Scott, Graham K. 1963. 'The dialects of Fore', *Oceania* 33: 280–6.

Stap, P. A. M. van der. 1966. *Outline of Dani Morphology.* The Hague: H. L. Smits.

Strange, Gladys Neeley. 1965. 'Nominal elements in Upper Asaro', *Anthropological Linguistics* 7/5: 71–9.

Strong, W. M. 1917–18. 'Vocabulary of Tevi', *Papua Annual Report* 1917–18.

—— 1918–19. 'Vocabulary of Ambo', *Papua Annual Report* 1918–19.

Thomas, K. H. 1941. 'The natives of Vanimo', *Oceania* 12: 163–86.

Thurnwald, Richard. 1934. 'Adventures of a tribe in New Guinea (the Tjimundo)', in E. E. Evans-Pritchard, Raymond Firth, Bronislaw Malinowski, and Isaac Schapera (eds.), *Essays Presented to C. G. Seligman.* London: K. Paul, Trench, Trubner and Co, 345–60.

Vincent, Alex and Vincent, Lois. 1962. 'Introductory notes on Tairora verb morphology and syntax', *Oceania Linguistic Monographs* 6:4–27.

Voorhoeve, Clemens Lambertus. 1965. *The Flamingo Bay Dialect of the Asmat Language*. 's-Gravenhage: H. L. Smits.

Whiting, J. W. M. 1941. *Becoming a Kwoma*. New Haven: Yale University Press.

Whiting, J. W. M. and Reed, S. W. 1938. 'Kwoma culture', *Oceania* 9: 170–216.

Williams, F. E. 1936. *Papuans of the Trans-fly*. Oxford: Oxford University Press.

Wurm, Stephen A. 1960. 'The changing linguistic picture of New Guinea', *Oceania* 31: 121–36.

—— 1961*a*. 'The languages of the Eastern, Western and Southern Highlands, Territory of Papua and New Guinea', in A. Capell (ed.), *Linguistic Survey of the Southwestern Pacific* (rev. edn.). Noumea: South Pacific Commission.

—— 1961*b*. 'Problems of the verb structure in Central New Guinea', Paper given at the Tenth Pacific Science Congress in Honolulu.

—— 1962. 'The present state of New Guinea (Non-Melanesian or Papuan) and Australian', in Horace G. Lunt (ed.), *Historical and Comparative Linguistics. Proceedings of the Ninth International Congress of Linguists, Cambridge, Massachusetts*. The Hague: Mouton, 575–9.

—— 1964*a*. 'Australian New Guinea Highlands languages and the distribution of their typological features', *American Anthropologist*, 66/4 Part 2: 77–97.

—— 1964*b*. 'Recent developments in linguistic studies of the Australian New Guinea Mainland', *Linguistic Circle of Canberra-Series A*, No. 4.

—— 1965 'Recent comparative and typological studies in Papuan languages in Australian New Guinea', *Lingua* 15: 373–99.

Wurm, Stephen A. and Laycock, D. C. 1961. 'The question of language and dialect in New Guinea', *Oceania* 32: 128–43.

Young, Robert A. 1964. 'The primary verb in Bena-bena', in Kenneth L. Pike (ed.), *Verb Studies in Five New Guinea Languages*. Norman: Summer Institute of Linguistics, 45–83.

Young, Robert A. and Young, Rosemary. 1965. 'The three dimensional classification system of Bena-Bena nouns', Anthropological Linguistics 7/5: 80–3.

Young, Rosemary. 1962. 'The phonemes of Kanite, Kamano, Benabena and Gahuku', *Oceania Linguistic Monographs* 6: 90–110.

Z'Graggen, J. A. 1965. 'Possessor-possessed relationship in the Sakɛr language, NE-New Guinea', *Oceanic Linguistics* 4: 119–26.

Unpublished Sources

(A) Material on languages of West Irian from Notebooks of J. C. Anceaux.

(G) Vocabulary of Savo by George Grace.

(I) Ivens papers. Material on Savo, Bilua, and Laumbe in Library of School of Oriental and African Studies, London.

(S) Vocabulary materials on languages of the Sepik area from Summer Institute of Linguistics.

(V) Material on Uasi, New Britain from Valentine.

This page intentionally left blank

13

Classification of American Indian languages: a reply to Campbell*

In *Language* 64: 3 (1988), Lyle Campbell contributed a review article concerning my book *Language in the Americas* (henceforth C and *LIA*, respectively) that was uniformly negative. I believe that readers of *Language* should be made aware that one year before my book appeared in print, on the basis of a brief exposition of its contents in *Current Anthropology* (Greenberg, Turner, and Zegura 1986), Campbell wrote that my classification of American Indian languages 'should be shouted down' (1986: 488). Under these circumstances, an objective review could hardly be expected. More importantly, the reader of C's review article gets no real notion of the main arguments and contents of the book.

I will first discuss several specific points made by C and then expand on two more general topics, those sections of my book (and the basic results contained in them) that C does not mention and the question of so-called Pan-Americanisms. C claims that my classification is not 'new' (591), referring to my 1956 paper (Greenberg 1960) [§4-Ed.]. But this paper was only four pages long, gave details only for South America and South American outliers in Central America, and presented no linguistic evidence. Even so, if it was not subsequently investigated by others, especially in regard to the new proposals for South America, this was not my fault, but stemmed from a combination of attitudes, essentially a relative lack of interest in the basic problem and a generally negative attitude towards all attempts at broad classification.

C denies that my success in African classification proves anything about the validity of my Amerindian work. This is, of course, true in the sense that it must stand on its own. Still, there should be some presumption that methods successful in one area will also be successful when applied elsewhere. These methods are set forth in great detail in Ch. 1 of *LIA* [§6-Ed.], but C does not

even mention the existence of such a chapter. With regard to African classi-
fication, he says that parts are in dispute, but does not specify which. In fact,
my classification is clearly the basis of present-day African historical linguistics.
There is no alternative classification, and the few proposals cutting across my
four basic groups have received no general support. All disputes have been at
the level of subgroupings, some of which I had said were tentative; and indeed
some of the changes were suggested by me. When one considers that even
today there is no unanimity regarding Balto-Slavic as a subgroup of Indo-
European (IE), similar disputes among the much less studied African lan-
guages should come as no surprise. A Nilo-Saharan newsletter and a *Journal of
Afroasiatic Linguistics* exist, and a comparative dictionary of Afroasiatic, with
no alteration of membership in the family and only one important sub-
grouping change, is proceeding under the editorship of Diakonov (1981). There
is also my Australian classification (*LIA*, 29), with which one may compare
Dixon (1980: 20)—which appeared later and independently.

C asserts that 'the genetic relationship among putative members of
Na-Dene itself has been seriously questioned' (593). He cites Levine 1979, but
fails to mention that an entire chapter of *LIA* was devoted to a refutation of
Levine's critique of Sapir's inclusion of Haida in the Na-Dene family. In the
1987 *Current Anthropology* review of *LIA*, two respondents (Hymes, 662, and
J. D. Sapir, 664) called this refutation 'brilliant'. Campbell and Mithun 1979
endorsed Levine's article emphatically, and Michael Krauss, in the same
volume, said that we should be grateful to Levine for debunking Sapir's thesis
once and for all.

Sapir presented two-way etymologies between Haida and Athabaskan, and
three-way etymologies among Haida, Tlingit, and Athabaskan. Levine cites
seven criteria, most of them in my view irrelevant, for rejecting these
etymologies. Levine repeatedly rejects the same etymology by citing different
criteria. It apparently occurred to no one, in their eagerness to show that
Sapir was wrong, even to note how many etymologies survived all attacks:
seventeen of the three-way etymologies and fourteen of the Haida-
Athabaskan comparisons. In my book I compared three branches of Indo-
European, applying Levine's criteria to the etymologies involving these
three branches in Pokorny's standard comparative dictionary. Only six
three-way comparisons survived when I used some latitude with regard to
Indo-European vowel alternation. If the criteria were strictly applied, none
of the etymologies would survive.

C calls the proposed affiliation of Xinca and Lenca a 'long-shot proposal'
(594), and he cites and attacks the evidence in Lehmann (1920: 767). I noted
Lehmann's proposal, just as I did many others; some I accepted, some I did

not. But what Lehmann and many others have done—comparing just two languages—is exactly what I never do, since it does not lead to an overall classification. To me Xinca and Lenca are related because they are both Chibchan. Any evidence showing that either is Chibchan is relevant. Since Chibchan is a large group, there are innumerable pairs one could compare in isolation, and it would make no more sense than if an Indo-Europeanist first tried to show that Lithuanian was related to Rumanian, then Greek to Albanian, etc. He would be excluding much relevant evidence, and even if he proved all of them he would still not distinguish the IE family. In fact, Xinca and Lenca have no special relationship within Chibchan. Xinca forms occur in thirty-three Chibchan etymologies and twelve Amerind etymologies. The figures for Lenca are similar. In some instances both languages are found; of these, some were cited by Lehmann and some I discovered myself.

C claims that 'most *LIA* examples are lexical' (596). He generally ignores an entire chapter devoted to grammatical evidence. In fact, the line between grammatical and lexical evidence is not always clear because of the process of grammaticalization. Furthermore, I have stressed the importance of agreement in morphophonemic alternations as being almost surely genetic in origin, particularly in cases of arbitrary suppletion. In *LIA* I gave an example of this (46–9), namely, the occurrence in widely separated areas of third person *i-* before consonants alternating with *t-* before vowels. The obscure notions about 'deep grammar' found in the American Indian linguistic literature are of three basic types: (1) agreement in alternations, particularly arbitrary suppletions; (2) agreement in whole or large parts of paradigms (e.g. Sapir's pronominal evidence for Algic); and (3) the marginal survivals in only a few words of grammatical elements that are productive elsewhere. As an example of this latter category one may cite the survival of the Chibchan-Paezan general numeral classifier *kwa-* (originally used for small round objects) on a few numerals in some Chibchan languages (*LIA*, 298–9). Like Sapir, and like the pioneers of Indo-European comparative grammar, I use both grammatical and lexical evidence. The preceding example (and many others that could be cited from my book) are surely examples of 'deep grammar' in any reasonable sense of the word.

C's article contains a serious misunderstanding of my views regarding glottochronology. I agree with C that it is a misuse of glottochronology to employ it as a device to classify languages, and I have never used it that way. In fact I have pointed out additional weaknesses of glottochronology, in *LIA* and other publications, both for classification and subgrouping. The point of Appendix A [§6.1-Ed.], discussed in *LIA* (28–9 [§6-Ed.]), is quite different. It supposes that by other methods we already have distinguished a valid

linguistic stock. Then, based on Brugmann's rule that an etymon is recoverable for the protolanguage if it occurs in two or more separate branches, it is clear that the more ramified a language stock is, and the shorter the period of separation of the branches, the more of the original vocabulary is recoverable. The whole stock can then enter into still deeper comparisons, utilizing quite a large portion of its original vocabulary. The whole argument is directed against the simplistic, but widely held, assumption that after a not very long period the resemblances between two related languages become indistinguishable from chance. This would be true only if there were just two languages in the world.

C suggests that I may have learned about mass comparison from Alfred Kroeber (597). This guess happens not to be true. In my discussions with Kroeber, who taught at Columbia after his retirement and whose colleague I was, we discovered that we had been using the same method. As for my late discovery of mass comparison, in my first classification of African languages I proposed sixteen stocks. This was mainly because linguists had not yet seriously investigated the Eastern Sudan in detail. In this first classification I distinguished stocks about on the level of Algic. I noted many resemblances among them in ways that crisscrossed and led to no clear results. Finally it struck me one day that nothing essential changed methodologically at higher levels. Hence I proceeded to compare all of the African stocks, just as I had previously compared individual languages. I found that most of those in the Eastern Sudan belonged together in a family I named Nilo-Saharan, a grouping now universally accepted.

C criticizes McQuown's 1942 Macro-Mayan hypothesis (598) because one of McQuown's reasons for positing it, the existence of glottalized consonants in Totonac, holds for other languages of the area outside of Macro-Mayan. I accepted Mexican Penutian, of which Macro-Mayan is a major part, for my own reasons. As in the case of Lehmann, C is criticizing another linguist, whose work I certainly appreciated, but whose views I adopted, at least partially, for reasons of my own. It is obvious to anyone who knows my work that I would immediately rule out glottalization as an irrelevant typological feature.

C suggests (597) that I put together my notebooks in accordance with a predetermined classification that was reflected in the notebooks as assembled. But where did I get the 'predetermined classification'? My procedures are clearly described in Greenberg 1960, to which C refers [§4-Ed.]. I started out by comparing about forty words in a large number of South American languages and finding that they fell into several clearly marked groups, such as Andean and Macro-Ge. Only then did I set up notebooks. When I found new

languages they either belonged to one of the previously recognized groupings, whose characteristics became clearer the more languages I used, or formed new groups by themselves.

C accuses me of failing to detect borrowings (599). In many instances I did detect them, in which case I simply omitted the form in question. In other instances I point to them as possible. For example, in the Amerind etymology for 'liver' (240) I noted that some Chibchan-Paezan forms might be borrowed from Spanish *pecho*, and in grammatical section 34 of Ch. 5 I suggested that the pronominal plural -*to* (found only in Fulnio among Macro-Ge languages) might be borrowed from a Macro-Panoan language, a group in which it is well attested. Doubtless I have made some errors in eliminating borrowings, but such inevitable flaws in a pioneering work such as *LIA* can hardly be taken to invalidate the book's basic thesis, the unity of the Amerind family.

In his discussion of the semantics of the etymologies, C finds that 'G's forms are quite permissive in semantic latitude' (600). I believe that I have been extremely sober. As a test, consider the glosses for the Almosan-Keresiouan etymologies 10, 20, etc., up to 100, with the numerals in parentheses indicating the number of instances of a particular gloss in the same etymology: 10: ask (3); 20: belly (4), breast (1); 30: break (1), be hit (1); 40: buy (1), trade (1), buy, sell (1), take (2); 50: come (1), arrive (2) go (1); 60: egg (2), bird's egg (1); testicle (1); 70: finish (4), finish, make, do (1); 80: foggy (1), moist (1), smoke (1), 90: grandfather (6), old man (1); 100: hit, fall into, slap (1), hit (1). I believe that this compares favorably with just about any etymological dictionary. C has taken a few complex etymologies and omitted the connecting semantic links. For instance, C's first example of permissive semantics, 'excrement/night/grass', is actually taken from the Amerind etymology for 'dirty' (*LIA*, 212). When examined in detail this etymology shows great semantic coherence; not only is it a strong etymology in its own right, but it even shows good evidence for the subgroups of Amerind proposed in *LIA*. A better gloss would have been 'black', or perhaps 'dark in color' for the original meaning. It occurs in five subgroups of Amerind. In Almosan its meaning is uniformly 'black', and the distributions throughout North and South America suggests that this was probably the original meaning. Keresiouan shows both 'dark in color' (Iroquoian) and 'green' (Keresan). In Penutian the meaning has shifted completely to 'green' and its close semantic relatives 'grass' and 'blue'. In South America, Macro-Tucanoan shows the meaning 'black' everywhere, except for Canichana 'night' and Shukuru 'Negro'. In Macro-Ge, the original meaning is preserved in Proto-Ge 'black', but in Cayapo and Chiquito the meaning has shifted to 'dirty'. Finally, in the

Equatorial group the meaning is uniformly 'excrement'. Remarkable evidence for the validity of the Penutian grouping is found in the shared semantic shift to 'green' and the reduplicated forms in three main subgroups, Plateau, California, and Mexican (North Sahaptin *tˢəktˢək*, Rumsien *čuktuk*, and Zoque *tˢuhtˢuh*). C has simply omitted all the connecting links. In Pokorny's standard comparative Indo-European etymological dictionary it is easy to find more drastic examples; for instance, under the root *perk-* one finds the meanings 'ask', 'temple', 'prophecy', 'herald'.

Some of my etymologies C rejects as onomatopoeic. Here as elsewhere C does not realize that an etymological dictionary is not meant as a 'proof' of relationship. Some items are of course more cogent in this respect than others. However, all languages have onomatopoeic expressions, including protolanguages. Hence there are in Pokorny, as in every etymological dictionary I have ever seen, some onomatopoeic expressions.

In his discussion of submerged grammatical peculiarities—which are never clearly defined—C points out that even these may be treacherous; he gives two examples (601). One is an apparently arbitrary phonological alternation in which Proto-Mayan resembles Quechua. The unwary reader might think that I had proposed it. It involves, along with other forms, Quechua -*ni* as first person singular, stated to be merely a connecting morph. In fact, in my lengthy listing of *n*- markers for the first person (usually singular) in Amerind languages, I myself eliminated this item, which does not figure in my discussion (*LIA*, 49–50) precisely because internal Quechua data show that it is not a first person marker. C's second example concerns a discontinuous negative in Quechua A and Quiché, a Mayan language. Again I did not suggest this. C notes that the second component in Quiché *ta*, is historically secondary. But one can tell this by bringing Mayan as a whole into the comparison, which is exactly the method I use. I not only did not suggest this (and C does not claim I did), but more importantly the methods I use would eliminate, and in fact did eliminate, both without my mentioning them.

C uses one device, a comparison of Amerind with Finnish, with regard to both my grammatical and lexical etymologies. His point is to show that my results are random by taking Finnish and trying to show that if I had encountered it in South America I would have classified it as an Amerind language. But I would never compare Finnish in isolation. If Finno-Ugric and the larger Uralic group to which it belongs were not already recognized, I would have discovered them. In *LIA* (24) [p. 95-Ed.] there is a table of nine basic words from twenty-five European languages. The fact that Finnish is closest to Estonian, and that their closest relative is Hungarian, and that this group is distinct from IE and Basque appears, literally, from the first word on.

It is a group at this level that should be compared with Amerind, and once more its distinctness is obvious. The large majority of C's forms do not even make it to Hungarian.

C claims that 'the number of the expected accidental matchings will be roughly proportional to the number of languages consulted' (603). It is true that if one adds another language at random there will be an increase in the number of resemblances. However, C fails to ask how many will be found in more than two languages. It is an elementary proposition of probability theory that the probability of multiple accidents is a product of their individual probabilities. Since, of course, we are dealing with fractions, this involves a decrease. The likelihood of finding a resemblance in sound and meaning in three languages is the square of its probability in two languages. In general the probability must be raised to the $n - 1$ power for n languages. Thus, if five languages each showed a total of 8 per cent sound–meaning resemblances pairwise, one would expect on a chance basis approximately 1/25,000 in all five languages. This multiplication of probabilities is discussed in Greenberg (1957: 39) [§2-Ed.], but was earlier noted by Collinder 1949 and no doubt by others as well. More concretely, if I have a group like the Western Romance languages (Italian, French, Spanish, Portuguese), there is an enormous difference between adding Rumanian and adding Basque. If I add Rumanian many three-way resemblances become four-way, etc., and a fair number of new etymologies appear. If I add Basque almost nothing happens.

The remainder of this reply addresses alleged inaccuracies of citation and analysis in *LIA*. C says that nearly all Americanists find shocking distortions in my data. However, Hymes (1987) noted, in regard to Ch. 2 of *LIA*, that ' . . . from the standpoint of Chinookan it is sometimes clearly confirming, sometimes capable of addition, once or twice questionable.' In spite of this and other comments of a like nature, there are of course errors. As I stated in *LIA* (ix), 'the present work is in many respects a pioneering one . . . Although I have exercised great care, it would be miraculous if, in handling such a vast amount of material, there were no errors of fact or interpretation. I will be grateful for any corrections suggested by readers. However, I believe the work should be judged as a whole . . . any user of dictionaries, even of intensively studied language families, will encounter numerous instances in which the same form has been assigned to different etymological entries by different scholars, or even in which the same form has erroneously been included in different etymologies.'

Little has been said in this reply regarding regular sound changes and correspondences. The question was discussed at considerable length in Ch. 1 [§6-Ed.], which C hardly mentions in his review. What happens when one

merely asks whether there is a relationship and then applies the rigid and largely irrelevant criteria used by Levine with regard to Haida and Na-Dene has been shown above. If applied to Indo-European it would quite simply destroy Indo-European comparative linguistics, which is in practice the main branch of historical linguistics. C apparently approves of Levine's methods and results, and similar statements of method by Kaufman and Goddard are even more restrictive. To see how unrealistic they are, the reader may peruse the list of French and English cognates (derived from etymological dictionaries) given in *LIA* (20–21) [p. 91-Ed.]. Using such methods one could disprove the affiliation of Hittite with Indo-European, whereas, in fact, Hittite revolutionized our ideas about the Proto-Indo-European sound system.

From all of C's discussion it would appear that the data assembled by me are so defective and misleading that we must conclude that there are no significant resemblances among the 200-odd independent stocks he posits for the Americas. But this runs against the doctrine of Pan-Americanisms, the first reference to which I find in Campbell and Kaufman (1980: 853), where they talk of 'widespread forms (so-called Pan-Americanisms)'. They use this concept in their criticism of Brown and Witkowski's 1979 article on Mayan-Zoque, a treatment restricted to protovelars. Campbell and Kaufman assert that any attempt to prove that Mixe-Zoque is related to Mayan must exclude fourteen of the proposed etymologies because they are Pan-Americanisms. If Pan-Americanisms are to be considered genetically related forms, this is of course contrary to normal practice. In an etymological dictionary of Germanic, no one excludes forms like English *two* and German *zwei* because this is an Indo-European etymon.

There is no mention in C of the widespread forms *n-* 'first person' and *m-* 'second person'. *LIA* (49–55) contains what I believe is the first detailed enumeration of the distribution of these pronouns, which extend from British Columbia to Chile and occur in every subgroup of Amerind. This distribution cannot be explained either by borrowing or chance. The borrowing of first- and second-person pronouns is very rare. That a highly improbable event should have recurred more than a hundred times exceeds the bounds of credibility. A number of widespread grammatical patterns discussed in Ch. 5, which include irregularities, also cannot be explained plausibly except as the result of genetic inheritance. One wonders why scientists, who should be impartial regarding types of explanation, should avoid genetic explanations in such cases in favor of borrowing over a distance far greater than that covered by Indo-European, and which would require contacts of virtually every language with every other one, given the variety of distributions.

I would like to emphasize the fact that my linguistic classification shows an almost exact match with genetic classification by population biologists and with fossil teeth evidence (Greenberg, Turner, and Zegura 1986). My linguistic classification was arrived at in total independence of this external evidence, and until recently I was unaware of the agreement. Of course it is not probative, but it is still remarkable and should be of interest to readers of *Language.* A recently completed worldwide study of a large number of genes by Cavalli-Sforza and his associates (1988) not only confirms Zegura's results, but shows that from a worldwide perspective the biological differences between North and South American Indians are minimal, if we exclude Na-Dene and Eskimo-Aleut, which I classify separately.

References

Brown, Cecil H., and Witkowski, Stanley R. 1979. 'Aspects of the phonological history of Mayan-Zoquean', *International Journal of American Linguistics* 45: 34–47.

Campbell, Lyle. 1986. 'Comment on Greenberg, Turner, and Zegura', *Current Anthropology* 27: 488.

—— 1988. 'Review article on *Language in the Americas*', *Language* 64: 591–615.

—— and Kaufman, Terrence. 1980. 'On Mesoamerican linguistics', *American Anthropologist* 82: 850–7.

—— and Mithun, Marianne (eds.). 1979. *The Languages of Native America: Historical and Comparative Assessment.* Austin: University of Texas Press.

Cavalli-Sforza, L. L., Piazza, Alberto, Menozzi, Paolo, and Mountain, Joanna. 1988. 'Reconstruction of human evolution: Bringing together genetic, archeological and linguistic data', *Proceedings of the National Academy of Sciences* 85: 6002–6.

Collinder, Björn. 1949. *La parenté linguistique et le calcul des probabilités.* Uppsala Universitets Årsskrift 13.

Diakonov, I. M. (ed.). 1981-. *Sravnitel'no-istoricheskij slovar' afrazijskix jazykov.* Moscow.

Dixon, R. M. W. 1980. *The Languages of Australia.* Cambridge: Cambridge University Press.

Greenberg, Joseph H., Turner II, Christy G., and Zegura, Stephen L. 1986. 'The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence', *Current Anthropology* 27: 477–97.

Hymes, Dell. 1987. 'Review of *Language in the Americas*, by Joseph H. Greenberg', *Current Anthropology* 28: 659–62.

Krauss, Michael. 1979. 'Na-Dene and Eskimo-Aleut', in Campbell and Mithun (eds.) 1979, 803–901.

Lehmann, Walter. 1920. *Zentral-Amerika.* Berlin: Museum für Völkerkundez zu Berlin.

Levine, Robert D. 1979. 'Haida and Na-Dene: A new look at the evidence', *International Journal of American Linguistics* 45: 157–70.

McQuown, Norman. 1942. 'Una posible síntesis lingüística macro-mayence'. *Mayas y olmecas, Segunda Mesa Redonda sobre problems antropológicos de México y Centro América*, Mexico: Sociedad Mexicana de Antropología, 37–8.

Pokorny, Julius. 1959. *Indogermanisches Etymologisches Wörterbuch*. Bern: Francke.

Sapir, J. David. 1987. 'Review of *Language in the Americas*, by Joseph H. Greenberg', *Current Anthropology* 28: 663–4.

14

In defense of Amerind*

14.1 Introduction

During the period 1991–92, five publications concerning my classification of native American languages (1987), hereinafter *LIA*, appeared in this *Journal*. One is a note on Pan-Americanisms by Campbell (1991), three are mainly devoted to a consideration of the data on specific languages or language groups (Poser 1992, Berman 1992, and Kimball 1992), and, finally, there is a review of the volume itself by Rankin (1992). I shall first consider the more general issues concerning the methodology of classification and the overall results when applied to the New World. In the latter part of this article, criticisms of the data presented regarding specific languages and language groups will be considered in detail. In some instances, the criticism of individual items will be shown to be invalid for specific examples from the languages discussed. In these instances, they will be cross-referenced to the relevant part of the initial discussion.

14.2 Methodology

14.2.1 *Multilateral comparison and the comparative method*

It seems to be widely supposed that multilateral comparison and the comparative-historical method are in some way contradictory. This is perhaps most starkly stated by Nichols (1990: 477, n. 1) as follows: 'Greenberg (1987) makes it clear that he believes such grouping cannot be reached by the standard comparative method; a wholly different method, mass comparison, is required.' Where I supposedly said this is not specified. In fact, just the opposite is stated: '...my remarks are not intended as an attack on the validity of comparative linguistics or on the importance of undertaking reconstruction. Rather, the discussion is meant constructively as a way of

taking first steps where the comparative method has not been applied for want of an assured basis in genetic classification' (*LIA*, p. 3) [§6-Ed.].

I said much the same thing in regard to the relationship between the two much earlier (1957) [§2-Ed.], namely, that the methods outlined there did not conflict in any fashion with the traditional comparative method and that they might be viewed as the first step in the method itself, for we cannot begin systematic reconstruction until we know which languages to compare.

Whether classification by multilateral comparison is to be viewed as a necessary preliminary step in order to define valid genetic units without which comparative linguistics cannot proceed or as the first step in the method itself—an alternative which seems more natural—is a matter of definition.

In *LIA* (p. 3), I cited Paul Newman's observation that 'the discovery of sound laws and the reconstruction of linguistic history normally emerge from the careful comparison of languages already presumed to be related.' This citation is referred to approvingly by Calvert Watkins (1990: 292). It should indeed be obvious that, if we mean by the comparative method the discovery of sound laws and the use of these, along with our knowledge of other processes such as analogy, as a method of reconstructing an ancestral language, we cannot begin until we have made a classification. The concept of multilateral comparison is simply an attempt to state more explicitly what this initial step consists of.

The only one of the leading Neogrammarians, to my knowledge, who seriously considered the problem of classification was Berthold Delbrück, Brugmann's collaborator on the famous *Grundriss*. He noted the priority of obvious etymologies to the sound laws based on them in the following terms (1884: 47): 'Since obvious etymologies are the material from which sound laws are drawn and this material can always be increased, therefore new sound laws can be continually discovered and old ones modified.'[1]

Mary Haas (1969: 72) noted in regard to what she called 'inspection', that 'in the twentieth century the method is still being used with effectiveness, whenever new classificatory hypotheses of relatively shallow time depths are needed. Greenberg, who refers to it as "mass comparison of basic vocabulary", has used the method in Africa and South America.' The reference to South America concerns Greenberg (1960) [§4-Ed.], a preliminary statement of the threefold classification of native American languages in which, for the

---

[1] 'Denn da einleuchtende Etymologien das Material sind aus dem Lautgesetze gezogen werden, und dieses Material sich immer vermehren und verändern kann, so können auch immer neue Lautgesetze erkannt und alte umgestaltet werden.' (All German translations are mine.)

first time, extensive stocks in South America, there viewed as subgroups of Amerind, are outlined; these appear in somewhat modified form in *LIA*. Haas's statement should evidently be modified in its reference to time depths. Of course, shallow and deep are relative matters, but most Amerindianists would certainly not regard groups such as Andean or Ge-Pano-Carib as shallow. It should also be noted that specific grammatical markers are equally involved along with basic vocabulary.

The essential identity of my methods to that involved in the discovery of Indo-European has also been noted by Comrie (1989: 92), who states that this is 'the approach that gave us the usual current classification of the Indo-European languages; it is the approach that has given us Greenberg's revolutionary but now established classification of the languages of Africa.'

Along the same lines it may be added that, according to Salmons (1992: 225, n. 8), Terrence Kaufman pointed out during a discussion of long-range comparison at the 1988 Hokan-Penutian Workshop in Tucson, that 'mass comparison' is actually the way that historical linguists begin formulating genetic affinity.

14.2.2 *Intuition and the comparative method*

Related to the foregoing question is Poser's assertion (1992: 202) that I appeal to intuited resemblances rather than regular correspondences as evidence of genetic affiliation. This statement, as is all too usual in the literature, confuses mere genetic affiliation with genetic classification. In this matter I could not be more orthodox. Comparative linguistics proceeds within families, that is, groups of languages having a presumed exclusive common origin like Indo-European, Uralic, or Austronesian. None of these families was discovered by finding regular sound correspondences. Rather, as noted by Newman and Watkins, these followed later as an important, but by no means exclusive, tool in reconstruction. Apparently Poser is opposing this, to him, 'scientific' procedure to intuition, which is presumably unreliable. But sound laws are based on perceived resemblances, which must logically, and in fact chronologically in the history of our knowledge of particular families, precede reconstruction. But what is intuition? I could hardly have produced an overall classification of native American languages with detailed subgroupings from my inner consciousness. What is meant here is rather observation or, as Haas calls it, 'inspection'. In general, linguists do agree on different similarities, which is what is really at stake here. It could be formalized by measuring feature similarities, but this hardly seems necessary. There is obvious intersubjective agreement on these matters, as can be seen from table 7 in *LIA* (p. 24) [p. 95-Ed.]. I believe that any linguist—and indeed any

reasonably intelligent amateur—who examines this table will arrive at precisely the same classification. Moreover, that, for example, Polish, Czech, Russian, Bulgarian, and Serbo-Croation *oko* 'eye' are more like each other than any is to Finnish *silmä*, Estonian *silm*, and Hungarian *sem* 'eye', and that the latter forms are more like each other than any one is to the Slavic forms, is obvious and a matter of observation, not to be attributed to some nebulous notion of intuition. As has already been noted, Delbrück viewed these obvious resemblances as the starting point for sound laws and not vice versa.

14.2.3 *Grammar versus vocabulary: a pseudo-issue*

There has been, to some degree, a division among historical linguists. Some, like Antoine Meillet and, following him, Ives Goddard, claim that no reliance is to be placed on lexical resemblances. On the other hand, in practice most attempts at 'relating' languages in the recent period have taken the form of a set of etymologies illustrating sound correspondences from which a protosound system is reconstructed. In such attempts virtually no attention is paid to grammatical resemblances, except for the occasional inclusion of pronouns as lexical items. I have been incorrectly assigned to the lexical camp by, among others, Baldi (1990: 12), but it is obvious that I use both, as seen, for example, in my African work in which resemblances among the noun-class markers played an important role in the establishment of Niger-Kordofanian. Furthermore, each African family was characterized grammatically, and in *LIA* itself I devoted a whole chapter to grammatical features. In some language families of isolating structure, however, such as Sino-Tibetan, there are virtually no grammatical markers, so that discovery of the family and its subgrouping has in fact rested on vocabulary alone. Yet it would seem that this is one of the most solidly based and universally accepted linguistic stocks in the world.

As I have pointed out before (e.g. Greenberg 1957) [§2-Ed.], what is relevant is simultaneous resemblance in both form and meaning in morphemes, whether they are lexical or grammatical. As in all historical endeavors, we are evaluating separate pieces of evidence. In historical research, evidence is weighed, not counted. The weight of an item depends essentially on the probability of it being merely chance and on its historical stability.

The longer an item, the greater its weight. A morpheme which recurs in several alternants, especially if these are suppletive, has particularly great weight. Thus, the agreement among Germanic languages in the irregular set 'good', 'better', and 'best' is extremely powerful and practically by itself sufficient to prove the relationship of the Germanic languages. However, since irregularities are targets for analogical leveling (e.g. 'good', *'gooder',

*'goodest'), its absence in a language does not in any way prove that it is not Germanic.

A further factor, relevance, is of very great importance in classification. Since the irregular set 'good', 'better', 'best' is only found in Germanic, it is highly diagnostic of membership in this group. Similarly, the suffix -*kwa* in nouns is highly diagnostic among Amerind languages of membership in the Chibchan-Paezan group. In Algic, in the famous pronominal set *n-*, *k-*, *w-*, *m-*, the last, as Sapir saw, was the most cogent and was, as he noted, passed over in discreet silence by Michelson in his polemic. In fact, *n-* 'first person' is diagnostic of Amerind while *m-* 'impersonal' is characteristic of Algic and perhaps, given its occurrence in Salish, of Almosan.

In Rankin's review, vocabularies are disparagingly referred to as 'wordlist linguistics' (1992: 330). Once more it is Haas (1966: 103) who has set the record straight. Referring to Karl Teeter, she states: 'But if he is dismissing all lexical comparison by the adjective "shallow", then he is simply selling the comparative method short.'

14.2.4 *The use of vocabulary evidence: the Chadic evidence*

An epithet is not an argument. To refer disparagingly to vocabularies means that one has failed to consider carefully just what kind of information they give. Given the arbitrariness of the relation between sound and meaning and the essential independence of most vocabulary items, a broad comparison of numerous vocabularies will provide both coherent and historically valid groupings.

An example of this is the treatment of the Chadic subgroup of Afroasiatic in Greenberg (1966: 46). Had I disregarded word lists, I would have eliminated at least 98 per cent of the relevant evidence. Only one language (Hausa) had been studied in any depth. There was a grammar of Angas by a British district officer without linguistic training, a nineteenth-century grammar of Mandara, and a few short grammatical sketches in Lukas (1937), whose material consisted almost entirely of word lists. All of the other sources were word lists, short ones by amateurs as well as those that I compiled in 1953 in the multilingual area of Jos, Nigeria (which included non-Chadic languages). Except for Hausa, even the grammatical material would have been judged 'obsolete' by Geoffrey Kimball.

Up to that time, Marcel Cohen had conjectured that Hausa was Hamito-Semitic, but without drawing firm conclusions; Lukas had distinguished two groups—Chadic and Chado-Hamitic—on typological grounds (sex gender in the latter group) but was equivocal in assigning either or both groups to Hamito-Semitic. Moreover, he had not taken into

account numerous languages only known from vocabularies he had not compiled himself.

In Greenberg (1966: 46), 111 languages are assigned to Chadic, which is divided into nine subgroups, and evidence is given for the membership of Chadic in Hamito-Semitic (renamed Afroasiatic).

Since then there has been an enormous accumulation of new data on Chadic languages. However, a comparison with Diakonoff (1988), the closest we have to a standard work on Afroasiatic, shows the following: allowing for changes in names and the discovery of new languages, his membership list of Chadic is identical to mine. Not a single language now believed to be non-Chadic was included in my original classification, nor was any of the languages I listed found not to be Chadic. Of the nine subgroups, numbers 5, 6, 7, and 8 each consist of a single language or a very small group which I did not venture to assign to a larger grouping. The three original main subgroups (1, 3, and 9) are now viewed as the three major subgroups of Chadic. This occurred by consolidation of the small subgroups I had consciously left unclassified within Chadic. The small group 4 (Masa) is somewhat doubtfully assigned a separate status by Diakonoff. In no case was a language assigned to one of the three main subgroups reclassified to a different one. The intricate subgrouping I proposed for group 1 into subsubgroups and even sub subsubgroups of Chadic remains intact.

These are part of the results referred to by Luc Bouquiaux (cited in *LIA*, p. 2) [§6-Ed.] as completely correct and which he attributed to my intuition. In reality, such results become obvious when one puts together data from a large number of languages, even when documented by material which is both quantitatively sparse and qualitatively far below present standards.

In regard to Chadic it is important to point out that it is by no means an obvious grouping. Diakonoff, probably the leading contemporary scholar in Afroasiatic studies, estimates on the basis of glottochronology (Orël and Stolbova 1995) that the age of Proto-Afroasiatic is at least 11,500 BP, and he considers Chadic to be the earliest branching. This is in the same range as the estimated age of the Folsom culture, considered by most archaeologists the earliest attested culture in the Americas.

An American example is Costanoan. Though it is only known from early and amateur transcriptions, it is not in doubt that Costanoan is a California Penutian language especially closely related to Miwok.

14.2.5 *Language-internal versus language-external explanation*

Poser, especially in his discussion of Yurumanguí, makes the assumption that only language-internal evidence can be used in comparative etymology.

Thus he frequently asserts, regarding morphological analyses relevant to comparative etymologies, that there is not a shred of internal evidence to support it and considers that this closes the case. He construes internal evidence in a peculiarly narrow way. For example, in the list of words and short phrases which is our only source of information for Yurumanguí, a strikingly high proportion of nouns begin with *a-*, often followed by a vowel, such hiatus being relatively uncommon in American languages. Examples include *a-ikan* 'wings', *a-utasa* 'stick', *a-uxui* 'wood', *a-ia* 'water', and *a-usia* 'ears'. Even though no meaning can be attributed to this *a-* (which is very likely a third stage article, that is, a fossilized demonstrative (Greenberg 1978)), its acceptance on diachronic grounds is surely justified. In fact, Poser is willing, as has been every other linguist who worked on Salinan, to abstract from the root an 'articular *t*' prefix which is not a synchronic article. The *a-* of Yurumanguí constitutes, in fact, statistical evidence—a kind that is fundamental in empirical science. A language like Yurumanguí is very much like one that has been deciphered from inscriptions and has a very small corpus. In such cases, external evidence plays a legitimately large role. An interesting case is New Phrygian. All of the inscriptions are variations on the same basic text. 'Whoever adds to this inscription, may he be accursed.' This meaning of the texts is adduced from numerous contemporary Greek examples. As Friedrich (1941) observed, there are two methods we can use. One is the combinatorial, based on text-internal relationships, and the second is the etymological, based on related languages. According to Friedrich, because of the absence of longer texts and the monotony of the semantic content, the combinatorial method is not very useful and so one must rely mainly on the etymological method. Phrygian, as Friedrich observes, is clearly Indo-European. The case therefore might appear superficially different from Yurumanguí, which virtually all American Indianists would consider an unclassified isolate. However, the major evidence for the Indo-European affiliation of Phrygian consists of the fact that 'whoever' is expressed as *ios-ni*, in which *ios-* is obviously like the Indo-European relative pronoun (e.g. Sanskrit *yas*, Greek *hos*) and the demonstrative *semoun* 'this' (sometimes *semou*) is interpreted as the dative of the demonstrative *kʲe-*, most closely resembling Slavic, e.g. Old Church Slavic *semu* 'to this'. No other forms of the relative or the demonstrative are known. Applying Poser's methodology we would have to say that there is not a shred of internal evidence for this analysis, and hence no valid evidence at all, so that we must abstain from classifying Phrygian as Indo-European. In fact, there is unanimous agreement among Indo-Europeanists that it is Indo-European.

14.3 Errors

14.3.1 *The significance of errors*

Leaving aside misunderstandings of my method as somehow in conflict with standard comparative linguistics, the chief thrust of most of the articles considered is that an overall classification can be categorically refuted by showing errors in individual items in a specific language or language group. It is noticeable that Poser does not list among the Salinan items to be discussed the *m* second person, nor does Berman in regard to Kalapuya, simply because these are not found in the etymological chapters they are discussing. Why items found in one chapter, but not another, should be ignored is not explained. It is clear that the consideration of such items transcends a critique based on a single language or a few languages taken separately and in isolation from the remainder.

Doerfer (1966: 111), the leader of the anti-Altaic group and an opponent of all broader comparisons, has expressed his views as follows: '. . . however, it would naturally be naive to assert that in this way the Altaic hypothesis might be refuted. That would be the same as if, for example, one might wish to point to the numerous errors in the Indo-European dictionaries of Fick, Walde-Pokorny etc. and believe that the Indo-European hypothesis would thereby be refuted.'[2]

Kimball seems to realize this when he states (1992: 489) that '. . . *Language in the Americas* is not necessarily wrong in its insights into the deeper relationships of American Indian families to one another . . . .'

The number of errors would have to be vast indeed to refute a classification simply by pointing to its errors. As noted in Goddard (1986: 196), Sapir claimed that well over half of his Algic etymologies would turn out to be correct, a statement that Goddard viewed as 'wildly optimistic'. Yet Goddard accepts Algic as a valid family. As the analyses in §14.7 of this paper will show, errors in my etymologies do not even approach such levels. Moreover, it should be kept in mind that the invalidation of one or two items in a widespread etymology does not invalidate the remainder of the etymology.

It is an almost instinctive reaction of specialists to reject, often with an obvious overtone of strong emotion, any attempt by an outsider to relate 'his' or 'her' group to anything else. Michelson's reaction to Sapir's Algic hypothesis is a representative example. The usual technique is simply to point out errors;

---

[2] '. . . jedoch wäre es natürlich naiv zu behaupten, dass damit die Altaistik widerlegt würde. Das wäre dasselbe als wollte man z.B. die zahlreichen Fehler in den indogermanischen Wörterbüchern von Fick, Walde-Pokorny usw. herausstellen und meinen, damit sei nun die Indogermanistik widerlegt. . . .'

what survives is not indicated. In the present instance, only Poser enumerates the etymologies, in the case of Salinan, in which he has found no error.

An interesting exception, which illustrates some of the relevant points, is that of William Thalbitzer, an eminent expert on Eskimo, in his reaction to Cornelius Uhlenbeck's hypothesis of the relation of Eskimo to Indo-European. After noting that 'an Eskimo linguist . . . would consider some of his etymological items horrific' (1944: 73), he analyzes the ninety-six items presented by Uhlenbeck, rejecting thirty-three (about 34 per cent of those proposed) but then adds some new evidence in favor of the thesis and concludes as follows: 'After all, there will certainly be a great deal left which will serve to support C. C. Uhlenbeck's argument.'

Obviously, not all errors are of equal significance. Kimball recognizes this by dividing his article into two basically separate sections. In the first he includes what he considers to be significant errors in one or more individual items in specific etymologies. In the second he lists by language all of what he considers to be errors. These include minor errors of transcription, but in many instances, 'corrections' which are based on sources that appeared after my book was published, or items like the handouts from Mary Haas's course on Natchez, of whose existence I could not reasonably be expected to know. The total effect, since it is not even accompanied by a list of items to which Kimball takes no exception, is to give a sense of an overwhelming number of errors.

There is a further consideration of major importance. My book does not contain the full evidence for the genetic position of any specific language. The problem of presenting the complete supporting data for an enormous stock is a formidable one and one must stop at some level. An etymological dictionary of Indo-European does not include an entire Proto-Germanic lexicon. Newman and Ma (1966: 219) allude to this problem when they write regarding me: 'He did not provide lexical items and grammatical items which were common to the Chadic languages but not found elsewhere in Afro-Asiatic. His proof of the unity of the Chad family was thereby rendered weaker than it need have been.' Note that in my discussion of Yuki-Gulf (*LIA*, p. 144), I point out that my singling out of etymologies exclusive to this group was exceptional, since I reckoned it to be a single subgroup of Penutian. This consideration is particularly relevant to the discussion of Kalapuya, as explained in §14.7.4.

14.3.2 *The use of early sources*

In regard to Kimball, it will be shown that in the case of Wappo, abstracting from relatively minor errors of transcription, virtually every item is found either in Radin or in the more recent work of Sawyer.

I believe that any discussion of errors should distinguish minor errors of transcription or misattribution of forms to other dialects or closely related languages resulting from a miscount of lines in my notebook from more serious ones. As will be shown in §14.8, many of these apparently more serious ones are in fact not errors at all but forms cited—usually with complete accuracy—from earlier sources. As we saw in the case of Chadic, if it is just a question of classification, the use of such materials on a wide enough scale gives reliable results.

Another source of criticism for some quite acceptable items is semantic nonidentities, which are in fact in most instances quite commonplace semantic equations of the sort that can be found in etymological dictionaries of well-studied linguistic stocks. Thus Berman's objection to relating Kalapuya *mu:kw* 'meat' (Penutian 80) to words meaning 'fish' elsewhere, on the grounds that the Kalapuya form never meant 'fish', shows a lack of understanding of what is involved when words of similar but not identical meanings are assigned to the same etymology. The claim here is rather that a word, which on comparative evidence meant 'fish', has changed its meaning to 'meat' at some point in the past during the period of differentiation of Kalapuya from related languages. Even within a single language like Eskimo, Jacobson has in his dictionary of the language of St Lawrence Island Siberian Yupik (1987) *neqa* glossed as 'food, fish', while for Salliq, a Canadian Inuit dialect, Spalding (1969: 43) gives *niqi* 'meat'. A number of other perfectly acceptable semantic equations rejected by various critics are discussed at appropriate points in §14.7 of this paper.

If, however, as noted even by Doerfer, a classification cannot be overthrown merely by pointing to errors in specific etymologies, how, one might ask, can it be refuted? This can be done only by a better classification, i.e. one which better accounts for the facts. The only alternative to my classification is one which has 150-odd independent stocks in place of Amerind. There is, as is well known, no single standard form of this classification on which Americanists agree. In any version, however, there are a host of facts unaccounted for, of which the first person *n* and second person *m* are simply the most conspicuous. A fuller discussion of these matters is to be found in §14.4 of this paper.

The Chadic example in §14.2.4 shows not only that poor sources can be used, but that early sources, even those which have been superseded by more recent ones, can be used effectively for classificational purposes. Two main types should be distinguished. The first consists of such sources as the vocabularies of various dialects of Yuki found in Barrett (1908), consisting of word lists using an English-based orthography, with perhaps a few brief

sentences but no accompanying grammar. In an entirely different class is Paul Radin's grammar of Wappo. Although it is not up to modern standards of phonetic accuracy, it is rich in vocabulary and in grammatical detail. It was finished in 1917, a time when there were still twenty-odd speakers of Wappo. Moreover, all the examples are taken from a collection of texts (Radin 1924). Surely there must be many valid forms in Radin's work which could not be, or perhaps accidentally were not, elicited in the much more recent work of Sawyer (1965).

As we shall see in detail in §14.7.6, virtually every one of the Wappo forms questioned by Kimball occurs in Radin's grammar. Obviously this is very different from simply making a careless error or citing invented forms. If they are to be rejected simply because they are in Radin's grammar, a reasoned argument for rejecting them *in toto* should have been given. The Kimball article is even more remarkable in that it 'corrects' forms in *LIA*, often in trivial ways, by reference to sources which were published in or after 1987, the year in which my book was published. I should hardly be brought to task for lack of clairvoyance! In fact, my manuscript was submitted to Stanford University Press in 1982, and it was not feasible to make changes in it after that date.

There were, however, some sources predating 1982 which were not used. The reasons are the following. I actually began work on native American languages in about 1951. That a great deal of work was done during the 1950s is evident from Greenberg (1960) [§4-Ed.], a paper given at the Congress of Anthropological and Ethnological Sciences in 1956 and from my joint paper with Swadesh on Jicaque in 1953. My basic purpose was classification. In numerous instances my notebooks were updated when substantial new sources appeared. However, since my work sought to be exhaustive in regard to the coverage of languages, the task was enormous, as evidenced by the fact that on vocabulary alone I compiled twenty-two notebooks on Amerind and one on Na-Dene. To have updated these notebooks endlessly would have been a Sisyphian task. In general, new data were added later when a new source appeared for a language which had only been poorly attested previously. Works which added substantially to our knowledge of languages already reasonably well known were also included. In the great majority of instances, the new data did not change the classification. However, as a comparison of Greenberg (1960) and Greenberg (1987) will show—particularly for South America where earlier data were often poor—there were some important changes in subgrouping. From time to time, in the light of new conclusions, substantial parts of notebooks were recopied in order to bring closely related languages closer together. The work was, however, largely in

abeyance for a considerable period during which I was mainly occupied with typology or classification in other areas.

The actual etymologies were compiled at a much later date during 1980–81. Unfortunately, in the multiple processes of recopying some notebooks, copying etymologies onto cards which were used in writing the manuscript in longhand, which was in turn transcribed by a typist, a substantial number of mostly minor errors accumulated. I was aware of this situation but realized that at least a full year's work would be involved in verifying each form in the original sources. Two factors beyond my control made the situation worse. One was that a new automatic typesetting program was used which was not well suited to linguistic materials. This resulted in some cases in almost random treatment of certain diacritics. When I saw these and other errors, I made numerous corrections on galley proofs. However, by a mischance, these corrections did not reach the press. As soon as the volume appeared I was dismayed to notice them. Almost none of these, however, had any bearing either on the classification or on the validity of the etymologies. In fact, more often than not supposed errors are quite accurate renditions of earlier sources.

14.3.3 *The notebooks*

It should be noted that, contrary to widespread assertions, the notebooks themselves are substantially accurate. Poser, the only critic who actually examined at least one of them and can hardly be accused of prejudice in my favor, stated (1992: 226), 'Judging from the Salinan entries, the notebook is considerably more accurate than *LIA*; it is inaccurate only in nine relatively minor cases.' He does criticize them for using only the transcriptions as they occur in the original sources, saying that this would be excusable in working notebooks. This is just what they are. I had no plans to make them generally available. However, at the suggestion of Bernard Comrie, they were reproduced by the Stanford Library in 1981. At present no other overview of the distribution of lexical items in native American languages exists.

14.4 American classification

14.4.1 *Pan-Americanisms*

For some time now the term 'Pan-Americanism', which was first alluded to (but not under that name) in Campbell and Mithun (1979: 54), has been current in Amerindian linguistic studies. Their statement is as follows: 'Similarly widespread forms must be eliminated; similar forms in many American Indian language groups do little to support a suspected closer

relationship between small subsets of these languages which may come to be compared.'

But how can a 'closer relationship' exist unless there is a wider one? In Campbell's most recent paper (1991: 394, n. 2) their number has unaccountably shrunk 'to a handful of lexical items which seem to recur widely among different American Indian language groups.' However, in a joint paper with Kaufman (1980), where a negative purpose could be served, no less than fourteen of sixty-two cognates involving Mayan and Zoquean proposed by Brown and Witkowski (1979) are eliminated on the grounds that they are Pan-Americanisms. It should be borne in mind that this paper was confined to roots in Mayan and Zoquean that had velars; roughly, one might suppose, about one-third of the roots. In place of fourteen, then, let us substitute forty-two. If just two language groups, now generally believed to be fairly closely related, retain this many Pan-Americanisms, their total number must indeed be very great, so great that Amerind would have been recognized early on as a single stock of the same order as Austronesian or Indo-European.

It is of interest to note that of the fourteen Pan-Americanisms noted by Campbell and Kaufman, seven are easily identified in *LIA*. That there should actually be fourteen with velars alone is highly improbable for the reasons just given. Incidentally, that one-half should be easily identifiable (of which two are Penutian, precisely the subgroup to which I assign Mayan and Zoquean) is evidence enough that in spite of errors inevitable in a pioneering work of vast scope, I have discovered a considerable number of valid Amerind cognates.

In the following list the numbering in Brown and Witkowski (1979) is followed by forms and gloss and then by the numbering in *LIA*. 5 *kam* 'carry, take', Amerind 58 'carry'; 10 *kuš* 'eat', Penutian 16 'bite'; 13 *uk* 'drink', Amerind 87 'drink'; 16 *tek* 'step on, step up', Penutian 231 'stand'; 23 *qul* 'trunk, tree', Amerind 261 'tree'; 31 *qut* 'kneel, crouch', Amerind 156 'knee'; 34 *baq* 'bone', Amerind 90 'dry'.

Campbell and Kaufman have evidently developed different views regarding Pan-Americanisms in recent years. Far from being few, Kaufman (1990: 26) says that 'there is a good deal of similarity among American Indian languages, and that the similarities I have in mind are not due to borrowing.'

In a series of inaccuracies, Campbell (1991: 394, n. 2) claims that, in addition to the present writer, others have also taken the notion of Pan-Americanisms 'as suggestive of genetic inheritance', namely, William Bright and Doris Payne. But Bright, of course, says just the opposite (1984: 15), noting that 'in any case it may be possible to offer a diffusional explanation.' The reference to Doris Payne puzzled me. I examined her entire article, cited

by Campbell, to no avail when it occurred to me that Campbell must have meant to cite *David* Payne, who wrote an article in the same volume on morphological characteristics of lowland South American languages. The latter describes five morphological elements, of which four are to be found in *LIA*. The fifth, a variable vowel, almost completely limited to Arawakan, I noted but excluded from the book because of its narrow distribution. However, Payne says (1990: 76) that 'unlike Greenberg I am not asserting cognacy.'

14.4.2 *The grammatical evidence for Amerind*

It is remarkable that in Campbell's review of *LIA* in *Language*, the reader would not suspect that a large part of chapter 2 and all of chapter 5 are devoted to the grammatical evidence. Liedtke (1991: 37) calls this part of my work brilliant, well founded, and of incomparable value for the demonstration of my thesis. He also states (1991: 37–8) that just through the treatment of the grammatical evidence alone, my publication is certainly a milestone.[3]

In American reviews, except for some attempts to explain *n* 'first person' and *m* 'second person' as due to contact, sound symbolism, or even accident, this part of the book was passed over in silence (except for comments on some minor items by Poser) until Rankin's review.

Although the grammatical evidence in *LIA* extends well beyond pronouns, virtually all of Rankin's criticisms refer to them. Since the two best-known examples of Pan-Americanisms, *n* and *m*, are both pronouns, the significance and bearing of pronominal evidence require separate discussion. Again, it was Haas who put her finger on the problem (1966: 102, n. 8) when, in reference to Boas's dismissal of their historical significance as due to 'obscure psychological causes', she noted that if this were true, then 'even Sapir's prize exhibit for Wiyot-Yurok-Algonkian, the four-way resemblance in personal pronouns, could be ruled out of court!' Indeed, if vocabularies are merely laundry lists, and pronouns, as Rankin says following Meillet, are pretty much the same the world over, we do not have much to work with and it is hard to see how Indo-European or any family could have been discovered in the first place.

It is clear that neither Meillet nor Rankin could have actually examined pronouns all over the world. To begin with, there is more to pronouns than first and second person singular markers. If we include demonstratives, often

---

[3] 'Dieser Teil von Greenbergs Arbeit ist brilliant, fundiert und von unschätzbarem Wert für seine Beweisführung. Aber allein durch die Behandlung der grammatischen Element ist Greenbergs Veröffentlichung sicher zu einem Meilenstein.'

used as third person pronouns, interrogatives, the intersection of personal pronouns with categories such as number, gender, and noun class, and the frequent existence of several pronominal series, bound and/or unbound in different syntactic functions, we have rather rich and, as it turns out, distinctive systems. Thus the Nilo-Saharan pattern *a-* 'first person'/*i-* 'second person'/*e-* 'third person' is different from the widespread Eurasiatic *m-* 'first person'/*t-*, *s-* 'second person' (which extends to Eskimo), and both differ from the South American pattern *i-* (palatalizing) 'first person'/*a-* 'second person'/*i-* (nonpalatalizing) 'third person'. Regarding the much discussed Amerind *n-* 'first person'/*m-* 'second person' pattern, Campbell notes that it is not ubiquitous (i.e. in the Americas); ubiquity is something that neither I nor others have asserted. The widespread occurrence of these pronouns in the Americas was recognized by Alfredo Trombetti, Sapir, Swadesh, Bright (who calls it a Pan-Americanism), and others. How commonplace, obvious, and widely known it is can be illustrated from the almost casual observation by Dixon (1910: 322) in his discussion of Chimariko: 'It will be seen that as in so many American languages, the pronominal stems of the first and second person are based on *n* and *m*.'

Nobody has ever talked of *n* and *m* as North Eurasianisms, Africanisms, or as Oceanicisms. It would clearly be absurd to do so. Their wide distribution in the Americas must simply be accepted as a fact that needs to be explained, not explained away. It should be obvious that, as Sapir noted regarding *n* in a 1918 letter to Speck, 'Getting down to brass tacks, how in the Hell are you going to explain general American *n-* 'I' except genetically.'

What are the alternatives? The inherent absurdity of first and second person pronouns being borrowed from Tierra del Fuego to British Columbia has apparently led to the abandonment of this explanatory hypothesis even by those inclined to attribute quasi-magical powers to language contact. Boas's 'obscure psychological causes', rejected by Haas as inadequate, are presumably a sound-symbolic explanation. More recently, Goddard (1986: 202, n. 5) asserted that 'a gesture equivalent to that used to articulate the sound *n-* is the single most important voluntary muscular effort of a nursing infant.' Presumably this is meant to explain *n-* first person in Amerind and elsewhere. This explanation is sometimes extended to nasals in general, for example, to explain Amerind second-person *m*.

It is interesting to note that European and North Asian *m-* 'first person'/*t-* 'second person', occurring in Indo-European, Uralic, Altaic, Yukaghir, Chukchi-Kamchatkan, and Eskimo-Aleut, have also been explained in sound-symbolic terms. Paasonen (1907), seeking to refute the proposal of a relationship between Indo-European and Uralic, cites Winkler's assertion

(1884: 88) that 'for purely psychological reasons *m* stands for the nearer and *t* for the more distant relation or for separation.'[4] Sound symbolism, to be meaningful, obviously cannot be different in different areas.

A recent attempt to explain personal pronouns symbolically is that of Nichols (1992: 261), where we learn that the 'root consonantism of personal pronouns turns out to have symbolic properties comparable . . . to those of "mama"–"papa" vocabulary.' It consists essentially of a contrast between labial and dental with a favoring of nasal articulation. Thus Northern Eurasian *m/t* and Amerind *n/m* are instances of the same pattern. In other words, the facts presented here are not denied. However, it does not apparently matter which person is designated by a dental, and which by a labial, or whether it is nasal or not! Moreover, the first pattern is simply called Old World and the second New World, disregarding such subtleties as the existence of Eskimo and Aleut in the New World utilizing the Old World pattern and the complete absence of this pattern of labial/dental contrast in Africa. One can only marvel at the intellectual contortions involved in refusing to accept the obvious genetic explanation.

Rankin also objects to the fact that *n* and *m* occur with a variety of vowels and that they are sometimes prefixed and sometimes suffixed. No one, as far as I know, has hitherto voiced these objections. In a completely non-controversial stock, Uralic, in the standard etymological dictionary based on the work of generations of scholars in this field (Rédei 1988: 294), both the first person singular and first person plural pronouns are reconstructed as *mV*, where *V* is an unspecified front vowel and in which some languages (e.g. Vogul) have a vowel preceding the *m* (*äm*, *om*, *um* as dialect variants). The singular and plural are always different, but the vowel alternations between them are unpredictable. In the Italian of Dante's time, the first-person pronoun could occur as a prefix *me-co* 'with me', as a suffix *amar-mi* 'to love me', or as an independent word *a me* 'to me', *mi vede* 'me he/she sees'.

Rankin also misinterprets the *k-* pronoun as referring indifferently to the first and second person (*LIA*, ch. 5, no. 19). This form is second person only in Algic and it is precisely here that it shows its value in that it helps explain why Algonquian forms involving *k-* are found in all instances in which the first person acts on the second person or vice versa (cf. Hockett 1966) as well as in first person inclusive. This situation exactly parallels that of Carib, except that here it is not used for the second person as such but only for first person inclusive (usually dual), first person acting on second, and second person

---

[4] ' . . . aus rein psychologischen Gründen *m* für die nächste, *t* die entferntere Beziehung oder Trennung stehe.'

acting on first. In fact, it is this very *k*- that Liedtke (1989: 285) singles out for special praise when he refers to 'the excellent analysis of the pronominal *k*- element (probably originally first person inclusive in its various reflexes).'[5] Adelaar (1989: 254), in his review of *LIA*, states that, along with *n* first person and *m* second person, *k* 'denoting an inclusive plural or related derived meanings' is among the grammatical similarities in chapters 2 and 5 that deserve close attention.

The diachronic typology of first person inclusives is discussed in Greenberg (1988). A first person inclusive tends to become a general first person plural and, given the fact that a fair number of Amerind languages do not distinguish number in the pronouns, a general first person pronoun and thus a rival to *n*-. This can be seen especially in Andean, Hokan, and Penutian. However, the fundamental distinction between *n* and *k* can be shown quite clearly in Oto-Manguean, where, as discussed in *LIA* (p. 124), a supposed alternation *$n \sim$ *k posited by Rensch (1976) unites the two of them in a single etymology (his no. 123) in which the *k*- forms cited are first person inclusive in accordance with the proposed original meaning, whereas the *n*- forms are either first person singular, first person indifferent to number, or first person exclusive. It is a typological fact that in systems with an inclusive–exclusive distinction in the first person plural, exclusives are often the plurals of first person singulars, but inclusives are virtually never so. The latter are either separate forms or are compounds of the first plus second person.

Rankin also attacks the *t 'third person' (no. 13 in chapter 5 of *LIA*; also discussed in chapter 2 of *LIA*, 46–8). He misinterprets the *t* as a mere 'linking consonant' without historical significance. What we are really dealing with is a widespread suppletive alternation of *i*- third person before consonants and *t*- before vowels, as it actually appears in Choroti (a Mataco langauge of South America). By a well-known process, unmarked forms are reinterpreted as zero forms and then other person forms are prefixed analogically.[6] Once more Adelaar mentions the discussion of this matter in *LIA* as one of the items deserving close attention. He characterizes it as 'the curious correspondences observed in the possessive prefix systems of many New World languages which are not otherwise more closely related.' This is accompanied by a page reference to the section in which the *$i \sim t$* alternation is a major

---

[5] '. . . sei hier nur genannt die exzellente Analyse der pronominalen *k*-elemente (ursprünglich wohl l.Person inklusiv) und ihre verschiedene Reflexe.'

[6] Cf. Bybee (1985), where this property of the unmarked is described. A striking example is Polish in which, based on the unmarked and irregular third person of the verb 'to be' *jest* 'is', a new first person singular *jest-em* was formed together with the rest of the paradigm, excepting the third person plural.

subject of discussion. Fox (1986) notes that the pre-Proto-Mayan system of third person markers *i-/*r is 'remarkably similar to the system reconstructed by Greenberg.' All of these scholars understand that what is involved is a suppletive alternation and not t as a linking consonant.

14.5 External linguistic and nonlinguistic evidence

14.5.1 *Linguistic classifications in other areas of the world*
Strictly speaking, the classification of American languages must be based on the evidence from American languages alone. But clearly if my classification of African languages had not become standard, *LIA* would have received less attention. Moreover, Kaufman (1990: 64, n. 1) and Nichols (1992: 5) have both compared my African and New World classifications. Kaufman wonders how I could have produced such a widely accepted classification in Africa and yet be wrong about the New World. While admitting that there were some original ideas regarding Africa that were correct, and that I brought order to a field in disarray, he then says that I often merely repeated earlier statements or switched labels. In fact, every change in name corresponded to a new idea. One example is Bantu, which in every previous classification had been given separate and independent status. I did not merely shift it to 'West Sudanic' (augmented by the Adamawa-Eastern group, a totally new idea), but rather I placed it within Niger-Kordofanian as especially closely related to a cluster of small languages (Bane) in the Cameroons and then at successively further distance genetically to other subgroupings. All these details are accepted in Wald's article, 'Bantu Languages', in the *International Encyclopedia of Linguistics* (1992).

Nichols (1992: 5) notes that my classification of African languages was 'a paradigm of success'. This, however, she asserts, was not true of my Amerind results, although the method is one and the same. The difference lies with the languages. 'The languages of Africa prove to have been under-analyzed raw data for comparative work, while those of the New World proved otherwise.' The phrase 'underanalyzed raw data' I find unintelligible. In the interest of historical accuracy it should be noted that my African classification was in fact opposed by almost all Africanist linguists for at least ten years. As Flight (1988: 26) noted, 'Throughout the 1950s . . . reaction from the Africanist establishment was unenthusiastic, sometimes overtly hostile.'

14.5.2 *Nonlinguistic evidence*
In his review Rankin described Luca Cavalli-Sforza as jumping on a bandwagon in announcing the basic agreement in the Americas between the correlation of gene frequency patterns in populations and the threefold

linguistic classification. If Rankin had read carefully my reply to Campbell's review (Greenberg 1989) [§13-Ed.], a paper cited in his bibliography, he would have seen in the last paragraph a reference to Greenberg, Turner, and Zegura (1986), which appeared prior to *LIA*. It should be noted that Zegura, not Cavalli-Sforza, is here the physical anthropologist, and that there is an additional third line of evidence, namely, variation in dentition. There are in fact still earlier papers by Turner and Zegura, the first of which, Turner (1971: 147), notes that 'the problem of the peopling of the Americas appears to be a relatively simple one from a dental anthropological viewpoint. The pattern that prehistoric New World variation exhibits is a partitioning into three fairly clear geographical clusters: Arctic Coast (Aleut-Eskimo), Alaska interior Northwest Coast (mostly Na-Dene speaking Indians), and all the rest of North and South America (Indian).' At the time Turner was unaware of Greenberg (1960) but heard about my views from a third party and wrote to me on 3 January 1969 for confirmation. He has more recently stated (personal communication), 'I think the record clearly shows that we identified a tripartite division of New World languages and dental variation independently of one another.'

Since then a fourth line of evidence, namely, mitochondrial DNA, which is independent of nuclear DNA, has been marshaled. Although results in this recent area are to be viewed with caution, Gibbons (1993), in an article in *Science*, finds further confirmatory evidence of the threefold division. There have been a whole series of studies in this area with special reference to the Americas and Siberia by Antonio Torroni and his associates. In the abstract of one of these papers (1992: 153) it is noted that certain site losses 'were observed exclusively in Amerinds . . . thus demonstrating that North, Central and South American Amerinds originated from a common ancestral genetic stock' and that the Na-Dene were a unique sublineage. Given the calculated divergence times, 'the ancestral Nadene migrated from Asia independently and considerably more recently than the progenitors of the Amerinds.' The Eskimo-Aleut did not figure in this particular study, but their distinctness from the Na-Dene and Amerind is generally assumed and, in fact, supported by other mitochondrial DNA data.

It is of interest to note that two other linguists (Lamb 1959 and Swadesh 1960) arrived at linguistic classifications similar to mine. However, my paper was delivered in 1956 and thus has chronological priority regarding language and also precedes Turner's and Zegura's work, so I could not have been influenced by them.

The only kind of argument, as far as I can see, that has been brought against all of this is that Boas, in Campbell's interpretation, rules out

correlations among race, language, and culture. But what Boas really asserted is simply their independence. Whether correlations exist is a question for empirical research. In fact, it is reasonable in certain cases to expect correlations. These include new settlements or migrations into uninhabited areas or ones in which the earlier inhabitants had a very different culture, subsistence base, and area of origin. Such is plausibly the case with the Americas, settled by migrants from different parts of Asia, with initially different gene pools and languages, and with considerable time intervals between them.

14.6 Present status of the classification

In spite of highly vocal opposition in some quarters, there are significant indications either of acceptance or of the fruitfulness of important parts of the classification. In an unsigned article, 'Macro-Gê,' in the *International Encyclopedia of Linguistics* (1992: II: 367–8), my classification of this group—including a whole series of languages which had not hitherto been considered as part of it, such as Chiquita and Fulnio—is simply reproduced from *LIA* without any critical reservations being expressed. A recent SIL publication from Colombia (Huber and Reid 1992: x) states, 'We have chosen to use Greenberg's 1987 classification, although we know some aspects of it are controversial.' It is clear that the basic reason for this choice was the genetic status of Chimila in Colombia, concerning which the authors state (1992: xxi), 'Chimila has been classified in several different language families including Chibchan, Chocoan, Carib and Arawak, but current evidence points to a Chibchan classification (Malone). This is in agreement with the Greenberg classification.' The mention of Malone concerns an unpublished manuscript called 'Chimila: Chibchan, Chocoan, Cariban or Arawakan?' All the languages of Colombia in this publication are classified in accordance with *LIA*. All the material at my disposal on Chimila consisted of earlier sources not up to current standards either in phonetic accuracy or grammatical analysis.

The same may be said in regard to Paya, a language spoken in Honduras. In Greenberg (1960) it was classified as Chibchan on the basis of the material in Lehmann (1920) and this classification was maintained in *LIA*. In what is almost the only new genetic proposal to be made during this period, Lyle Campbell (1979: 942) notes that 'Dennis Holt's recent descriptive work and connected historical studies have conclusively demonstrated that Paya is a Chibchan language.' He cites Swadesh in 1967 and Loukotka in 1968 as having made this assertion but omits any mention of Greenberg (1960) which preceded both of them.

Greenberg and Swadesh (1953) asserted for the first time that Jicaque was a Hokan language. In fact, the hypothesis came from me, as indicated in the article itself, though Swadesh wrote the entire article, including glottochronological data which I myself would not have included and which led Oltrogge (1977) to state that the relationship was established by glottochronology. Jicaque was included in my earliest Hokan note-book because it seemed to fit there. At the time, as Swadesh informed me, Jicaque was considered a great mystery. This notebook is still in my possession and the glottochronological words are checked off in a different handwriting. The Hokan affiliation is now the generally accepted view.

The proposal that there is a special relation between Yukian and the Gulf languages has attracted considerable attention. Liedtke (1991: 37) talks of Greenberg's 'convincing attempt to prove a genetic relationship between Yukian and Penutian. He indicates a particularly close connection between Yuki and Gulf.'[7] Two American linguists have been independently working on the hypothesis of a relationship between Yukian and Gulf, Pamela Munro and Victor Golla. The former (letter of 16 September 1992) states that she 'has found a considerable number of additional similarities between the Yuki and Gulf languages and regards the proposal as worthy of further research.' Golla expresses himself in similar terms when he states (letter of 11 September 1992) that 'the lexical evidence presented in *LIA* is sufficient to make Greenberg's hypothesis of a close connection between Yuki and the Gulf languages (particularly Atakapa and Chitimacha) a good candidate for further investigation.' He also states, 'I think you will agree that the more accurate data on the Yukian side does not detract from—in fact in a couple of cases enhances—the comparison of Gulf forms.' His reference to more accurate data relates to Sawyer and Schlichter (1984) and Schlichter (1985), both of which appeared after my manuscript was submitted in 1982. The only material available to me on Yuki proper (excluding Wappo) were the vocabularies in Powell (1877) and Barrett (1908) and a few citations from Lamb's field notes contained in Shipley (1957).

Incidentally, Shipley cites Wappo forms from Radin (apparently pro-scribed by Kimball) and Costanoan forms from early vocabularies since these were the only materials available. He deserves great credit for having been on the right track in connecting Yukian to Penutian.

---

[7] '...Greenbergs überzeugende Versuch zwischen Yukian und Penutian genetische Verwand-schaft zu beweisen. Er weist auf einen besonders engen Zusammenhang zwischen Yuki und Gulf hin. ...'

An important characteristic of correct hypotheses is their fruitfulness in leading to further investigation and new knowledge. A conspicuous example concerns Berman (1986: 421), who notes regarding Yurok: '*ca:n* "young" is related to *cin* "young man" cited above. I believe that one of these is an old changed form of the other, but I do not know which is which.' Of course, it is not necessary that one should be a changed form of the other within Yurok. They could be the survivals of an old pattern of alternation which is reflected elsewhere in related languages. In fact, these examples fit perfectly with the general pattern of an extremely widespread Amerind etymology which appears in *LIA* as Amerind 125 'girl' in four major subgroups. Subsequent research by Ruhlen (1994), adding much data on Amerind kinship terminology from anthropological sources not utilized in my notebook, shows that this etymology is very widespread, appearing in every subgroup of Amerind. Moreover, the internal vowels show a definite pattern *i* 'masculine', *u* 'feminine', *a* 'common gender'. Further, as I had noted even before perusing the enormous amount of evidence accumulated by Ruhlen, the opposition *i* 'masculine'/*u* 'feminine' appears as a functioning gender system in some Equatorial and Macro-Tucanoan languages (*LIA*, pp. 296–8) and is identical to the Chinook vowel gender distinction. It was further noted by Ruhlen that this same pattern appears in Amerind 62 'child(2)' as *makV* 'child', *mikV* 'son', *mukV* 'daughter'. The feminine form does not appear clearly in my entries, but once more Ruhlen has found convincing evidence for the feminine *u* in such forms as Shiriana (Chibchan) *moko* 'girl' and Pauishana (Carib) *mu'gi* 'daughter'.

Two friendly critics, Hymes and Liedtke, suggested a third person element *u* in addition to the *i* ~ *t* alternation. I was aware of this form but thought its distribution was too sporadic to be included. In view of the intriguing possibility that Proto-Amerind was a sex gender language with the *i/u* opposition as basic, it is possible that these examples of *u* may be survivals of the feminine third person. Note that in Mayan, *u-* is the usual third person possessive but *i-* occurs in the same group as the possessive with kinship terms.

Lest anyone think that this vowel opposition is merely sound-symbolic, it should be noted that Semitic and Egyptian share the exact opposite pattern, namely, *u* 'masculine' and *i* 'feminine', while languages in northern Asia have a feature opposition based on high 'feminine' versus low 'masculine', as in Gilyak *irš* 'mother'/*erš* 'father'.

Perhaps the best summary of my criticism of the approach to *LIA* in the articles which have appeared in *IJAL* up to now is to cite Lamb (1987: 102): 'The volume calls for careful study and for follow-up research; it does not call for criticism based on incomplete understanding of his methods.'

14.7 Individual languages

14.7.1 *Salinan*

Poser does not, at least explicitly, doubt the Hokan affiliation of Salinan. He omits from his comments *moʔ* 'thou' (*LIA*, p. 53) and the articular *t* (*LIA*, p. 47), both of which are treated at length in chapter 2. He does discuss Salinan items in the Hokan and Amerind etymologies of chapter 4 and some of the grammatical items in chapter 5.

In the Hokan section there are thirty-seven Salinan items in the etymologies, of which he says twenty-four call for comment, and in the Amerind etymologies twenty-nine, of which eighteen call for comment. However, the mere fact that an item is selected for comment does not necessarily mean that any real doubt is shed on it. For example, in regard to Hokan 132 'sleep', which contains the San Miguel Salinan entry *p-apa* 'copulate', he notes that 'the segregation of initial /p/ is legitimate.'

The items which he says call for comment fall into two groups. In some instances there are errors in transcription or dialect attribution (practically all of which are correct in my notebooks) which have no effect on the validity of the etymology. Of the remainder almost all contain prefixes, namely, *t-*, *ṭ-*, *p-*, or *k-*; in a few cases questions of segmentation or grammatical analysis are involved. I believe almost all of these are justified for reasons which will be given later.

Of the twenty-four Hokan items which Poser says call for discussion, the following eleven obviously have no mistakes which bear on their validity as etymologies. These, cited by their Hokan headwords, are: 8 'ashes', 46 'eat', 51 'extinguish', 77 'hand', 90 'large', 92 'laugh', 94 'left (side)', 114 'rabbit', 121 'rope', 145 'testicle', and 160 'white'. Of the Amerind items, the following seven fall into the same class: 2 'above(2)', 43 'body(1)', 53 'brother', 87 'water', 175 'make', 217 'say', and 228 'shoulder'.

Regarding segmentation, the most frequent question that arises is in regard to the so-called articular *ṭ* and the connected problem of *t-*, which for reasons set forth below may be considered a variant of *ṭ-*. With regard to *ṭ-*, Alden Mason, our major published source, states: 'It is prefixed to most nouns derived from verbs.... Moreover it or a similar prefix is found with most pronominal possessive prefixes and with certain forms of the verb' (1918: 24). It is in fact a well-behaved stage III article, that is, a formative ultimately derived from a demonstrative, which became a definite article before finally becoming fossilized on nouns, here as a prefix (cf. Greenberg 1978; 1981). Its ultimate origin in a demonstrative was recognized by Sapir and its function as a nominalizer on verbs can be shown on comparative diachronic evidence to

be a secondary one. It is clear that a considerable number of Salinan nouns beginning with *ṭ-* are not analyzable as nominalizations, e.g. *ṭ-am* 'house', so segmented in Mason (1918: 130). It is also clear that in many cases Mason was unsure since internal synchronic evidence is insufficient. In such cases the use of comparative evidence is quite normal. In one case, Hokan 123 'salt', Poser in fact considers the treatment of *ṭ-* as a prefix in *ṭakaiʔ* plausible, even though Mason (1918: 133) did not make this analysis. Gursky (1974: 197) also considers the initial *ṭ-* in his comparative Hokan entry 'louse(3)' to be legitimate, though Mason made no such analysis. If then we allow this kind of analysis, my Hokan entries 61, 62, 87, and 123 and Amerind 87 are legitimate.

The second question that arises is whether the 'articular' *ṭ-* has an alveolar variant *t*. Both Sapir and Gursky evidently thought so. Gursky admits *t-* in *t-ierk* 'liver' in his entry 'liver(2)', saying it could be the nominal prefix, as well in 'louse(2)', where it is given without comment. Here San Miguel has *i'ke* and San Antonio *t-ik'eʔ*, showing clearly that it is a prefix.

It was noted earlier that Mason (and Sapir) considered the prefix *t-* of the possessive (e.g. *t-a-sanat* 'our hide') to be probably the same element. It is a well-attested phenomenon that a stage III article is used in some languages in the possessive construction. This derives from earlier uses as a true definite article with possessives, e.g. Italian *il mio nome* 'the my name'.

There is even variation within the same possessive paradigm of *ṭ-* and *t-* (e.g. *t-um-sanat* 'thy hide', but *ṭ-a-sanat* 'our hide'). Furthermore, we have an entry for 'son-in-law' (Mason 1918: 135) in which the San Antonio dialect has *ṭēleM* and San Miguel *tʰe:lemʔ*. Mason himself was aware of this variation, more examples of which could have been cited, at least in regard to the possessives. He refers to them as follows (1918: 31): '. . . initial *ṭ-* (frequently heard as *t-*)'. In view of all this, there is no good reason for excluding *t-* as a legitimate variant of *ṭ-*. This makes the comparison in Hokan 10 'back' legitimate.

A few etymologies involve the segregation of an initial *p-*, found in some instances in San Miguel but not in San Antonio. Mason conjectured a sound change, namely, loss of *p* in San Antonio. Not all examples are initial. As noted by Sapir, there are also instances in which San Antonio does have a *p* corresponding to *p* in San Miguel. Gursky, in his entry 'bone(2)', accepts it, but with the proviso that it is valid only if *p* is here a prefix. It seems possible that the absence of *p* in San Antonio is due to a phonological change, although it would not be a regular one, while in other instances, as Sapir conjectured, the absence is morphological. At any rate, it does appear that there is a reasonable case for a formative *p* in San Miguel and this would justify my analyses in Hokan 62 'blood' and 132 'sleep' and Amerind 47 'bone' and 255 'throat'. There is also a final *-p* in Amerind 181 'many(2)'.

There is a formative *k* which occurs in Hokan 20 'boil(2)' and 83 'heavy', in which Mason considered the segmentation plausible. In Amerind 182 'many(2)', in addition to -*p* in San Miguel as noted above, there is *k-* in both dialects which Poser evidently accepts as a formative.

This leaves a handful of cases, some of which are doubtless errors, e.g. Hokan 166 'woman(3)', where the cited form is actually Chumash. There is a crossing between Hokan 18 'body' and Amerind 185 'meat'. These two should be consolidated and the Comecrudo entry in the former eliminated. The few remaining items are Hokan 26 'burn'—probably, as Poser notes, a copying error by Mason, Hokan 69 'full(2)', a doubtful analysis on my part, though defensible, and an initial prefix *l-* assumed by Sapir in Amerind 148 'heaven(2)', and which I also used in Amerind 27 'bee', for which no doubt the case is weak. This leaves only Amerind 79 'kill', in which a form 'cause to sleep' is equated with 'kill', and Amerind 137 'hand' containing San Antonio *maʔa:ʔa* and San Miguel *ma:ʔa*, both meaning 'bring, carry'. Here Poser wonders why I did not rather cite *me:n* 'bring, carry'. The probability is that all these are examples of the widespread Amerind root 'hand, give, take' illustrated in *LIA* (p. 158).

14.7.2 *Yurumanguí*

This extinct language of Colombia is known from a single document published in Rivet (1942). As is evident from the title of his article, we are presented, along with the linguistic data, evidence for its Hokan affiliation. I was at first extremely skeptical about this hypothesis given the geographical distance from the nearest Hokan languages, but after entering Yurumanguí in my Chibchan notebook and after comparing it with Andean, the two most likely choices on geographical grounds, and finding that they did not fit in, I compared it with Hokan languages. I found a few—and given the paucity of the material one would hardly find many—instances of clear resemblances to exclusively Hokan roots. Among these the most obvious were Hokan 44 *si-* 'to drink' and 158 *punpun-* 'to bathe'.

Of the twenty-six Yurumanguí forms cited in *LIA* only *ita-(asa)* 'wife', Hokan 161, is incorrect, being, as noted by Poser, an error for *ataisa* 'sister'. In the other cases, the forms are cited correctly though in some instances Poser objects to their grammatical analysis. Several types, however, are clearly justified. One is *a-* prefixed to nouns and discussed earlier in §14.2.6. Another is -*na* suffixed to nouns, for which Poser says there is no internal evidence, but the contrasts (Rivet 1942: 12) between *miti-na* 'calebash used as a bowl' and *miti-ssi-na* 'calebash bowl of small size' as well that between *bai-si-na* 'knife', containing the same diminutive -*si-*, and *bai-kal* 'machete' surely

justify the analysis of -*na* as a nominal suffix. Regarding another nominal suffix, -*sa* and its variants, Poser himself (1992: 217) admits that the suggestion that this suffix forms infinitives is plausible while finding no direct evidence for its use on nouns. It is, however, a widespread typological fact that the infinitive as a nominalization quite commonly involves a formative which is itself nominal, and the items cited here are long enough to suggest morphological complexity.

If, as seems reasonable, one allows *a*-, -*na*, and -*sa* as nominal formatives, the case against Yurumanguí as Hokan virtually disappears.

14.7.3 *Yurok*

Unlike Poser, Berman does not enumerate those etymologies concerning which he has no objections. These are, with their Almosan-Keresiouan and Amerind headwords respectively: AK 4 'all(3)', AK 7 'armpit', AK 8 'arrive', AK 19 'behind', AK 30 'break(1)', AK 31 'break(2)', AK 35 'brother', AK 38 'buttocks', AK 47 'cloud(2)', AK 49 'come(1)', AK 51 'to cook', AK 62 'evening', AK 87 'go', AK 89 'good', AK 90 'grandfather', AK 103 'hot', AK 112 'left(side)', AK 113 'leg', AK 120 'man', AK 135 'name(2)', AK 144 'one', AK 165 'shadow', AK 178 'snow(1)', AK 181 'steal', AK 182 'stick', AK 208 'worm'; Am 8 'arm(2)', Am 50 'breast', Am 53 'brother', Am 69 'cold', Am 76 'dark(1)', Am 87 'drink(1)', Am 116 'flea', Am 149 'hit', Am 158 'know', Am 171 'long', Am 220 'see', Am 232 'sit(1)', Am 252 'swallow'. This is a total of thirty-nine, twenty-six of which are Almosan-Keresiouan and thirteen Amerind. However, the Almosan-Keresiouan entry 'armpit' is a duplicate of Amerind 'arm(2)'.

There remain eleven entries which Berman questions. Two of these, *rakw*- 'to be able' (AK 1) and *kän*- 'open' (AK 45), which Berman is unable to identify, are really Wiyot forms. Wiyot *r* in 'to be able' derives from **n* as usual and corresponds to that consonant elsewhere in this etymology.

In Amerind 24 (*LIA*, p. 89), I have Yurok -*pʔos* 'beard'. According to Berman the stem is *mepʔos*, which is certainly true synchronically. However, there is a well-attested process by which an unmarked form is reinterpreted as a zero form and then affixes of marked categories are added to it. We have already seen an example in §14.4.2 of the same process in regard to Amerind *t*-, a third-person marker reinterpreted as zero to which markers of other persons were then prefixed. This process is abundantly illustrated in Watkins (1962). Haas (1958*a*: 165) recognizes this process in the Algic etymology for 'forehead' when she compares Yurok *moLKʷoh* 'head' to forms in Algonquian without the *m*-, noting that 'Yurok has incorporated the indefinite pers. pref. *m*-'.

This leaves a relatively few items regarding which Berman is presumably right, but for which, in some instances at least, I believe I should not be faulted since I was drawing reasonable conclusions based on the data available to me. This did not include Berman's supplement to Robins's dictionary which appeared in 1982. In the case of Amerind 219 'say(3)', involving the irregular Yurok verb *hegol* 'to say', I was well aware of the intensive and verb pluralizer *-eg-*, but in his dictionary Robins normally gives the unaugmented form only. For example, in the grammatical section of his book (1958: 81), he mentions the pair *koʔmoy-* and *kegoʔmoy* 'to hear', but in the lexicon he only cites *koʔmoy-*. Hence when, for the highly irregular verb 'to say', he gives the stem as *hegol-* (as he does also in the grammatical section (1958: 73)), it was reasonable for me to interpret this as the stem. So too with the verb *hego:(s)-* 'to shout', which I took to be etymologically connected. The same reasoning holds for 'carry' (AK 42), Yurok *negem-*, cited in exactly that form in Robins (1958: 228) with no indication that it contained the infix *-eg-*. In regard to 'hold' (AK 101), *ʔekonem*, I once more reproduced what I found in Robins (1958: 269). where there is no hint of morphological complexity. Berman asserts that the first part contains *ʔekoh* 'hold on, grasp', but this verb is not found in Robins, as indicated in Berman's comment that he (Robins) did not know of it. There is no reference by Berman to anything published regarding his analysis of *ʔekonem* as *ʔekoh* + *n-(eg)-em*, so how could I possibly have known about it? Regarding the variant *pleli* for 'broad' (Am 52), this form came from Kroeber. Berman analyzes it as containing a morpheme *-ʔel-* 'a suffix referring to houses' and refers to Robins (1958: 95) where indeed as part of a classifier system it is given as 'big (in relation to houses)'. However, further down the page there is simply *pleʔloy-* 'to be big'. None of the adjectives except 'big' referring to houses in the table of classifiers at the top of the page contains *-ʔel-*, nor does such an element occur in Robins's lexical section, which includes formatives. Once more, I had no basis for making this analysis and Berman does not refer to its being in print anywhere. I do not assert that he is wrong since he is the expert on Yurok, only that I could hardly be blamed for not making this rather speculative analysis without any basis in the published material. There remain a few other items on which Berman is certainly right and which involve errors on my part. But of course all this has no real bearing on the classification of Yurok.

14.7.4 *Kalapuya*

The problem of finding published sources for the Oregon subgroup of Penutian was an especially difficult one. I was unaware of Jacobs's unpublished field notes at Washington. I did examine his Kalapuya texts, but there

was neither a glossary nor a grammatical section. I could perhaps have utilized these texts but only with enormous effort. What I did obtain was a microfilm of a manuscript from the Boas Collection in the Library of the American Philosophical Society (Angulo and Freeland, n. d.) that contained a series of word lists semantically arranged. Both Angulo and Freeland were certainly highly reputable linguists of the period, but they had not segmented the articles or possessive prefixes. A check of the original microfilm shows that I faithfully reproduced their material. For reasons given in §14.3.2 the item *mu:ku* 'meat' (Penutian 80), objected to on semantic grounds, is acceptable. In addition, Penutian entries 12 'bad', 133 'know(2)', 178 'person', 213 'shoot', 239 'swallow', as well as Amerind 40 'blood', 97 'earth', and 250 'sun(3)' are not objected to. More importantly, since, as explained in §14.3.1, etymologies confined to a level such as that of the Oregon branch of the Penutian sub-group of Northern Amerind were not included in the book, so the evidence in Frachtenberg (1917) and especially the two-way glottochronological comparison with Takelma by Swadesh (1965) were omitted except when involved in etymologies for Penutian as a whole. The material in Frachtenberg and Swadesh seemed to me sufficient to show the affiliation of Kalapuya to the Oregon subgroup.

14.7.5 *Natchez, Muskogean, Gulf, and Yukian*

The number of languages and the extent of the material discussed from these groups in Kimball (1992) are such that for the sake of brevity I only discuss two in detail, Natchez and Wappo. With regard to Muskogean in particular, I wish merely to point out that Kimball has frequently 'corrected' my material on the basis of either unpublished work, or material that appeared subsequent to or in the same year that *LIA* appeared, e.g. Kimball (1987; 1988; 1991), Munro and Willmond (1988), or, in several cases, after my manuscript had been submitted in 1982, a fact of which Kimball, of course, was not aware. I was at fault in sometimes using the English index to Byington's (1915) dictionary of Choctaw, of whose problems I was, in fact, aware, so that I often checked with the Choctaw–English sections but I did not do so in every case. I wish to thank Kimball for these corrections. It does not appear that he questions the Natchez-Muskogean genetic connection.

Regarding Natchez, I used the published sources that were available. These consisted basically of Swanton (1907; 1924) and Haas (1956; 1958*b*). Kimball 'corrects' my forms by including material from class handouts in a course given by Haas and from her as yet unpublished dictionary. A special problem concerns that of verb forms, that are usually followed by one of a number of auxiliaries. As is evident from Haas's treatment, which auxiliary is used is

irrelevant for comparative purposes. Moreover, in Haas (1958*b*), Natchez verb forms are cited without their auxiliaries. Swanton normally gives verb forms as stems, not followed by a hyphen and omitting the auxiliaries. To take one example, my citation under Penutian 68, Natchez *t^{s}i*: 'to fall', is 'corrected' (Kimball 1992: 491) to *ci:-haku?iš* also meaning 'to fall'. This form is not found in Haas (1956), so presumably Kimball got it from a class handout or Haas's unpublished dictionary.

My source was Swanton *dzī:* 'fall, lie' (1924: 72). On the basis of Haas's description I tacitly changed *dz* to *t^s* as with other presumably nonphonemic voiced stops in Swanton. Length is faithfully reproduced; Kimball's form is, of course, only notationally different. To take another example, my citation under Amerind 97 *?inoo* is absolutely identical with that in Haas (1956: 67). In what sense this is an error is difficult to see.

The following entries are not questioned by Kimball: Penutian 7, 17, 19, 202, 203, 218 and Amerind 140, 165, 210. In the following list I take all of the Natchez forms corrected by Kimball in the list on pages 491–2 of his article. In each case I give the source in Swanton or Haas. Unless marked (H) for Haas, the form will be understood to derive from Swanton. In these cases, since the paginations in Swanton 1907 (513–78) and 1924 (46–75) do not overlap I merely give the page number. Similarly, since the paginations do not overlap, citations from Haas (1956) will have 61–72 as numerical limits and those from Haas (1958*b*), 231–64. There is one item from Gallatin (1848) marked as G. The interested reader can readily check these forms with the list in Kimball's article and with *LIA*. I should add that there is clearly a *ha-* prefix to verb stems, so its omission in *LIA* is not therefore an error. The only major mistake, as far as I can see, is my form *ala* 'come' (Penutian 8) which I now see derives from a misreading of a somewhat complicated sentence in Swanton (1924: 58) in which *ala* is meant as Muskogean, not as Natchez. With this exception, and also of course those items to which Kimball takes no exception, I list for each Natchez entry in *LIA* the form found in the original source and the place it is found in accordance with the conventions just described.

Penutian 9. *o:no:xk* 'briar' (48); 14. *pa:k-* 'beat' (H239); 39. *ba:la* 'shut, close' (65); 50. *do:p^h* 'cut' (66); 53. *wat(a)* 'die' (68); 68. *dzi:* 'fall, lie' (72); 69. *do:t* 'descend' (74); 79. *hak* 'afire' (69); 85. *únu* 'berry' (518); 86. *pi* 'swell' (H252); 89. *kus, gus* 'give' (517); 109. *k^{w}eye:* (H247); 110. *ta* 'strike' (518), *da* 'hit, kill' (74); 130. *ox* 'finish killing' (74); 146. *pulu* 'to lighten, lightning' (66); 160. *ihi* 'mouth' (74); 170. *toowa* 'night' (G); *du:wa:sidi* 'spend the night' (58); 171. *hash* 'old' (517); 213. *weL* 'speak, talk' (75); 240. *pes, pe^{h}l* 'sweep' (518); 245. *wits* 'tell' (75); 261. *pa:* 'to plant' (48); 262. *kittip-* 'turn' (H237); 267. *we:* 'go about' (pl.) (66); 273. *pacak-* 'wet' (H253).

Amerind 2. *abo:* 'head' (65); 11. *puk, pukaʔh* 'to smoke' (521); 27. *mo:m* 'bee' (48); 45. *luk* 'to boil' (70); 47. *tso:x* 'to dry' (73); 49. *kets* 'to break' (517); 52. *ʔepet-kup* 'wide' (H67, 244), *pet* 'to spread' (H66, 244); 55. *ʔinu* 'name' (H248); 76. *timu:ya* 'yesterday' (77); 87. *kun* 'water' (68); 97. *ʔinoo* 'earth' (H67); 198. *mayuk* 'dark, night' (67); 213. *be* 'go' (pl.) (66); 221. *eL* 'to see' (74); 226. *lem* 'to shine' (74); 234. *hebe:š* 'bark' (69), *ebesh* 'bark' (520); 235. *nu:* 'sleep' (69); 238. *mis-* 'stink' (H244); 255. *nAsh, noch* 'throat' (73); 259. *dzu:* 'tree'; 265. *yo:ba* 'to rain' (73); 270. *mai* 'to love' (67).

The only other language I consider in detail is Wappo. My basic sources were Radin (1929) and Sawyer (1965). Every Wappo form cited in *LIA* is to be found in these sources. In addition, earlier data exist in Powell (1877) and Barrett (1908). Kimball's procedure was usually to identify every Wappo gloss in *LIA*, then to look it up under its semantic heading in Sawyer's English–Wappo vocabulary, and then to assume it was wrong if it differed in any way. In some instances, however, Kimball corrected an item in *LIA* by finding an item in Sawyer with a similar or identical form but with a totally different meaning.

The following are a few examples of the pitfalls of this procedure which I illustrate before giving a systematic list of the sources of my Wappo forms.

Under Amerind 96, I cite *ɔma* 'earth' found in Radin (1929: 181) and elsewhere. Note that Barrett also recorded Wappo *o:ma* 'earth, world' (1908: 73), all suggesting that this is a genuine form. In addition, in a section at the end of Sawyer's book called 'Wappo–English Affixes' (pp. 125–8), of which Kimball was evidently unaware, we find the entry *ʔoma* 'around, all around, unspecified location in the general environment' as well as *ʔomi* 'around, all around the ground'. In addition to all this, there are several other entries in Sawyer's vocabulary that show what is clearly the same lexical item, namely, under 'dig', we find the entry *ʔoma eli* 'to dig' (as against *ʔeli* 'to dig out') in which *ʔoma* plausibly means 'ground' as well as under 'take', *ʔoma hopehse* 'take care of a place'. Kimball corrects this well-attested form for 'earth' to *ʔehna* 'earthworm'!

Another example is Amerind 126 *mi* 'go', corrected by Kimball to *miʔ* 'you (sg.)'. In Radin we find *mi* 'go' on page 115 and elsewhere. Another verb, Amerind 128 *-ya-* 'to go', is described in Radin (1929: 46) as '[a]n old verb meaning 'to go' only found in texts now. . . .' There are several other mentions of this verb in Radin. In Sawyer (1965: 127) we find *-ya.miʔ* 'going to (see *miʔ*) verb trans. indefinite' and under *-miʔ* (1965: 27) 'transitive indefinite'.

In a number of instances Kimball has 'corrected' perfectly accurate forms cited by me by using entries in Sawyer that involve the present indefinite of the verb which Sawyer used as the citation form. As we have seen, Kimball

was unaware of the affix listing in Sawyer, to which one should add the verb affixes found in his vocabulary under the heading, 'vb'. An example is Amerind 98 'eat', under which I give Wappo *paʔ*, *paʔe* 'eat' (Radin 1929: 174), found also as Sawyer *paʔmi*, as well as *ʔopahmi*, the forms cited by Kimball. These latter contain the Wappo present indefinite suffix *-mi*.

The following Wappo citations are not questioned by Kimball: Penutian 24, 32, 38, 58, 61, 69, 102, 110, 145, 152, 173, 176, 193, 226, 254, 272, 275; Amerind 43, 53, 137, 140, 150, 165, 204, 261.

In the following list my source in Radin's grammar is given simply by page number; those from Sawyer are marked with S. A page number here is superfluous because the vocabulary is alphabetized.

Penutian: 4. 'all(3)' *mi-muli* 'world' (181), *mul'* 'all' (S): 9. 'arrow' *luka^h* 'bow' (186): 21. 'body' *wil* 'body' (191); 23. 'boil(3)' *kɔ^h* 'boil' (167); 39. 'close (v.)' *pɔn* 'close' (170); 41. 'cold' *tˢatˢa*, reduplicated from *tˢa* 'cold' (179); 42. 'come' *tˀɔɪ* 'arrive' (177); 43. 'cook(1)' *yoko* 'cook' (172); 46. 'cover' *sa^h* 'cover' (175); 50 'cut' *tsipu* 'cut off' (179); 53. 'die (1)' *ɔtɛwɪ*, *ɔtɛ* 'be dead' (162); 66. 'eye' *hucˀi* (S) 'eye'; 68. 'fall(2)' *tsɛ* 'drop' (179); 78. 'fire(2)' *hɛl* 'fire' (183); 81. 'flea' *tšˀoʼtɛˀ* 'flea' (190); 84. 'fly(1) (v.)' *yɔk·ɔ* 'fly' (181); 101. 'hard(2)' *tˀɔʔɛ* 'hard' (194); 105. 'hear' *pikakhiʔ* 'listening' (S); 118. 'hot(3)' *šɔi* 'be burnt, scorch' (164); 120. 'house(2)' *hil·iʔ* 'build a house' (S); 125. 'keep' *piʰnɛ* 'keep' (175); 128. 'kill(3)' *lipu* 'kill many' (168); 132 'know(1)' *hat* 'know' (165): 135. 'large(2)' *taʔɛya* 'be heavy' (163); 142. 'lie down' *yɔkɛ* 'lie' (181); 148. 'live(2)' *nɔmi* 'live' (173); 151. 'look(2)' *pɛ+hɛ+l* 'look around' (166), *pɛ* 'look' (174); 161. 'much(1)' *lɛʼa* 'much' (193); 170. 'night(2)' *ʔuʼtšuwa* 'night' (182); 171. 'old' *hiʼn-hasi* 'old man sun' (183), *has* 'old' (192), cf. *hin* 'sun' (S), *-has* 'Mr., old' in affix list (S); 177. 'person(1)' *k·aʼniʔ* 'human being' (185), *kˀanih* 'person; sometimes of people who are dead to avoid use of specific names', cf. also *kˀanih ʔeniya* 'a mean person' (S); 191. 'road' *mits·* 'road' (186); 197. 'run(1)' *puli* 'run away' (175); 202. 'say(2)' *hai, ha* 'say' (165); 203. 'say(3)' *wilɛ* 'tell' (172), *walɛ* 'call' (180); 232. 'stand up(1)' *yoka/yokal* 'get up' (172); 236. 'stone(1)' *weʼi* 'flint' (191), cf. Barrett *we:*; 237. 'stone(2)' *ɔtsɔl·a* 'rock' (181); 238. 'swallow(1)' *lɛkɛ* 'swallow' (169), *ma-lɛʼkˀɛ* 'swallow it' (34); 243. 'take' *pɪta* 'take' (174); 268. 'wash(1)' *tsɔ* 'wash' (174).

Amerind: 49. 'break (v.)' *ho-kˀɛtsɛ*, *kˀɛtsɛ* 'cut' (166); 65. 'clean(v.)' *takˀɛ* 'to clean' (177); 68. 'cloud(2)' *pɔʼhi* 'smoke' (188); 72. 'come' *na* 'come up' (171); 76. 'dark' *sum* 'evening' (188); 84. 'dig(3)' *ɛli* 'dig a hold' [*sic*] (161); 85. 'dirty' *šikaʼtɪs* 'blue' (137); 87 'drink(1)' *uki* 'drink' (164); 89. 'drink(3)' *mey* (S); 96. 'earth(3)' *ɔʼma* 'earth' (181); 97. 'earth(4)' *nuʼiˀ* 'sand' (187); 98. 'eat(1)' *paˀ* (174); 108. 'feather(1)' (misnumbered 109 by Kimball) *putšaya* 'abundant hair' (188); 114. 'fire(3)' *ma, maha* 'burn low, smoulder' (168), *ma* 'burn (itr.)'

(170); 120. 'foot(1)' *pɔi* (19), *p·ɔ* 'kick' (170); 126. 'go(1)' *mi* (115); 128. 'go(3)' -*ya*- 'go' (100); 170. 'long(1)' *ɛla* 'get deep' (161); 182. 'many(3)' *mul* 'all' (120); 220. 'see(1)' *naw* (S); 239. 'smell(2)' *ku^h* 'stink' (168); 254. 'thin' *kut^ʔi:ya* 'small (sg.)' (S); 262. 'two' *p'a'la* 'human twins' (188); 269. 'wing' *k'apɛ* (185).

14.8 Conclusion

Even in this lengthy rebuttal it has not been possible to address all the criticisms that have been leveled at my classification of New World languages. Thus omission of, for example, discussion of some of the material from specific languages does not indicate that these criticisms are necessarily valid. For example, I did not consider the Yuki data. As noted by Golla in §14.6 above, more recent data on these languages actually improves the case for Yuki-Gulf. Similarly, in spite of some errors and the evidence of more recent data, F. D. Winston (1966: 168), reviewing my comparison of the Niger-Congo subgroups, noted that 'despite the errors, the language groups arrived at were identical with Greenberg's. The fact emerges that Greenberg's method can tolerate a high degree of inaccuracy.' What I have tried to do is to explicate the relationship between multilateral comparison and the comparative method of reconstruction. I have argued that classification is the indispensable first step in the comparative method itself and reconstruction and the discovery of sound laws subsequent enterprises whose success depends on the success of the initial step—the discovery of valid families on any level.

Furthermore, I have once again stressed that it is the comparison of the historically most stable and basic morphemes, whether grammatical or lexical, over a wide area that leads to a classification of the languages into families at various levels. It was precisely in this way that the Indo-European family was discovered before even the term 'sound law' existed in linguistics. It was by these methods that I reduced the bewildering diversity of African languages to four families. My application of this method to Native American languages was no different than that in Africa and there is no reason to think that these methods should not have equal success in the Americas. The number of important grammatical items, including shared irregularities in the case of the third person *i-/t-* and numerous lexical items that permeate Amerind language and are not found in the Old World, virtually guarantee the validity of the Amerind family. That the threefold division I have posited has been independently identified by patterns of dentition and of gene distribution only provides further support.

I have taken some pains to address the issue of 'errors in the data', especially since many scholars—both Americanists and others—have assumed

uncritically that the expert criticisms of my work with regard to errors were valid, without checking either my notebooks or the relevant literature. We have seen that the great majority of so-called errors were not in fact errors at all, and that the remainder could hardly be expected to affect the classification at all.

Another major point of this paper, as set forth in broad outline in §14.5.2, is that nonlinguistic evidence in an area of new settlement may be expected to coincide, at least roughly, with the linguistic classification. The extent of the agreement in the present case is such that mere accident is ruled out on any reasonable statistical basis and is indeed obvious without any such mathematical underpinnings.

References

Adelaar, Willem F. H. 1989. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Lingua* 78: 249–55.

Angulo, Jaime de, and Freeland, Lucy L. n. d. 'Kalapuya semiology, word lists semantically arranged'. MS. no. 293 in the Boas Collection of the American Philosophical Society, Philadelphia.

Baldi, Philip. 1990. 'Introduction: the comparative method', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin and New York: Mouton de Gruyter, 1–13.

Barrett, S. A. 1908. 'The ethno-geography of Pomo and neighboring Indians', *University of California Publications in American Archaeology and Ethnology* 6/1: 7–332.

Berman, Howard. 1986. 'A note on the Yurok diminutive', *International Journal of American Linguistics* 52: 419–21.

—— 1992. 'A comment on the Yurok and Kalapuya data in Greenberg's *Language in the Americas*', *International Journal of American Linguistics* 58: 230–48.

Bright, William. 1984. *American Indian Linguistics and Literature*. Berlin: Mouton.

Brown, Cecil, and Witkowski, Stanley. 1979. 'Aspects of the phonological history of Mayan-Zoquean', *International Journal of American Linguistics* 45: 35–47.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. (Typological Studies in Language, 9.) Amsterdam: Benjamins.

Byington, Cyrus. 1915. *A Dictionary of the Choctaw Language*, ed. John R. Swanton and Henry S. Halbert. Bureau of American Ethnology Bulletin no. 46. Washington, DC: Government Printing Office.

Campbell, Lyle. 1979. 'Middle American languages', in Lyle Campbell and Marianne Mithun (eds.), *The Languages of Native America: Historical and Comparative Assessment*. Austin: University of Texas Press, 902–43.

—— 1991. 'On so-called Pan-Americanisms', *International Journal of American Linguistics* 57: 394–9.

Campbell, Lyle, and Kaufman, Terrence. 1980. 'On Mesoamerican linguistics', *American Anthropologist* 82: 850–7.

Campbell, Lyle, and Marianne Mithun. 1979. 'Introduction', in Lyle Campbell and Marianne Mithun (eds.), *The Languages of Native America: Historical and Comparative Assessment.* Austin: University of Texas Press, 3–69.

Comrie, Bernard. 1989. 'Genetic classification, contact and variation', in Thomas G. Walsh (ed.), *Synchronic and Diachronic Approaches to Linguistic Variation and Change.* Georgetown University Round Table on Language and Linguistics 1988. Washington, DC: Georgetown University Press, 81–93.

Delbrück, Berthold. 1884. *Einleitung in das Sprachstudium* (2nd edn.). Leipzig: Breitkopf and Härtel.

Diakonoff, I. M. 1988. *Afrasian Languages.* Moscow: Nauka.

Dixon, Roland B. 1910. 'The Chimariko Indians and their Language'. *University of California Publications in American Archaeology and Ethnology* 5, no. 5. Berkeley: University of California Press.

Doerfer, Gerhard. 1966. 'Zur Verwandschaft der altaischen Sprachen', *Indogermanische Forschungen* 71: 81–123.

Flight, Colin. 1988. 'The Bantu expansion and the SOAS network'. *History in Africa* 15: 261–301.

Fox, James. 1986. 'Comment on Greenberg, Turner, and Zegura, "Settlement of the Americas"'. *Current Anthropology* 27: 488–9.

Frachtenberg, Leo S. 1917. 'Comparative studies in Takelma, Kalapuya and Chinookan lexicography, a preliminary paper', *International Journal of American Linguistics* 1: 175–87.

Friedrich, Johannes. 1941. 'Phrygia (Sprache)', in Georg Wissowa (ed.), *Pauly's Realencyclopädie der Altertumswissenschaft* (2nd edn.), vol. 20, no. 1. Stuttgart: Metzler, 867–82.

Gallatin, Albert. 1848. 'Hale's Indians of North West America and vocabularies of North America, with an introduction', *Transactions of the American Ethnological Society*, 2. New York: Bartlett and Welford, xxiii–clxxx, 1–130.

Gibbons, Ann. 1993. 'Geneticists trace the DNA trail of the first Americans', *Science* 259: 312–13.

Goddard, Ives. 1986. 'Sapir's comparative method', in V. Cowan, M. K. Foster, and K. Koerner (eds.), *New Perspectives in Language, Culture and Personality.* Amsterdam: John Benjamins, 191–210.

Greenberg, Joseph H. 1966. *The Languages of Africa.* The Hague: Mouton.

——1978. 'How does a language acquire gender markers?' in Joseph H. Greenberg *et al.* (eds.), *Universals of Human Language*, vol. 3. Stanford: Stanford University Press, 47–82.

——1981. 'Nilo-Saharan movable *k-* as a stage III article (with a Penutian typological parallel)', *Journal of African Languages and Linguistics* 3: 5–12.

——1987. *Language in the Americas.* Stanford: Stanford University Press.

——1988. 'The first person dual as an ambiguous category', *Studies in Language* 12: 1–18.

Greenberg, Joseph H. and Swadesh, M. 1953. 'Jicaque as a Hokan language', *International Journal of American Linguistics* 19: 216–22.

Greenberg, Joseph H., Turner II, Christy G., and Zegura, Stephen L. 1986. 'The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence', *Current Anthropology* 27: 477–97.

Gursky, Karl-Heinz. 1974. 'Der Hoka-Sprachstamm: Eine Bestandsaufnahme des lexikalischen Beweismaterials', *Orbis* 23: 170–215.

Haas, Mary. 1956. 'The Natchez and Muskogean languages', *Language* 32: 61–72.

——1958*a*. 'Algonkian-Ritwan: the end of a controversy', *International Journal of American Linguistics* 24: 159–73.

——1958*b*. 'A new linguistic relationship in North America: Algonkian and the Gulf languages', *Southwestern Journal of Anthropology* 14: 231–64.

——1966. 'Wiyot-Yurok-Algonkian and the problems of comparative Algonkian', *International Journal of American Linguistics* 32: 101–7.

——1969. *The Prehistory of Languages*. The Hague: Mouton.

Hockett, Charles F. 1966. 'What Algonquian is really like', *International Journal of American Linguistics* 34: 58–93.

Huber, Randall Q. and Reid, Robert. 1992. *Vocabularios comparativos, palabras selectas de linguas indígenas de Colombia*. Santafé de Bogotá: Instituto Lingüístico de Verano.

Jacobson, Steven A. 1987. *A dictionary of the St Lawrence Island/Siberian Yupik Eskimo*. Fairbanks: Alaska Native Language Center, University of Alaska.

Kaufman, Terrence. 1990. 'Language history in South America: what we know and how to know more', in Doris L. Payne (ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press, 213–41.

Kimball, Geoffrey. 1987. 'A grammatical sketch of Apalachee', *International Journal of American Linguistics* 53: 30–8.

——1988. 'An Apalachee vocabulary', *International Journal of American Linguistics* 54: 387–98.

——1991. *Koasati Grammar*. Lincoln: University of Nebraska Press.

——1992. 'A critique of Muskogean, "Gulf," and Yukian materials in *Language in the Americas*', *International Journal of American Linguistics* 58: 447–501.

Lamb, Sydney M. 1959. 'Some proposals for linguistic taxonomy', *Anthropological Linguistics* 1/2: 33–49.

——1987. 'Comment on *Language in the Americas*', *Current Anthropology* 28: 1012.

Lehmann, Walter. 1920. *Zentralamerika*. 2 vols. Berlin: Reimer.

Liedtke, Stefan. 1989. 'Review of *Language in the Americas*', *Anthropos* 84: 283–5.

——1991. *Indianersprachen*. Hamburg: Buske.

Lukas, Johannes. 1937. *Zentralsudanische Studien*. Hamburg: Frierichsen, de Gruyter.

Malone, Terrell A. n.d. 'Chimila: Chibchan, Chocoan, Carib or Arawakan?' MS.

Mason, J. Alden. 1918. 'The language of the Salinan Indians', *University of California Publications in American Archaeology and Ethnology* 14: 1–154.

Munro, Pamela, and Willmond, Catherine. 1988. 'Chickasaw: an etymological dictionary'. MS.

Newman, Paul, and Ma, Roxana. 1966. 'Comparative Chadic: phonology and lexicon', *Journal of African Languages* 5: 218–51.

Nichols, Johanna. 1990. 'Linguistic diversity and the first settlement of the New World', *Language* 66: 475–521.

——1992. *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.

Oltrogge, David F. 1977. 'Proto Jicaque-Subtiaba-Tequistlateco: a comparative reconstruction', in David Oltrogge and Calvin Rensch (eds.), *Two Studies in Middle American Comparative Linguistics.* Austin: University of Texas Press, 1–52.

Orël, N. N., and Stolbova, O. 1995. *Hamito-Semitic Etymological Dictionary.* Leiden: Brill.

Paasonen, Heikki. 1907. 'Zur Frage nach der Urverwandschaft der finnisch-ugrischen und indo-europäischen Sprachen', *Finnisch-Ugrische Forschungen* 7: 13–31.

Payne, David. 1990. 'Morphological characteristics of Lowland South American languages', in Doris L. Payne (ed.), *Amazonian Linguistics: Studies in Lowland South American Languages.* Austin: University of Texas Press, 213–41.

Poser, William J. 1992. 'The Salinan and Yurumanguí data in *Language in the Americas*', *International Journal of American Linguistics* 58: 202–29.

Powell, J. W. 1877. *Tribes of California.* (Contributions to North American Ethnology, vol. 3.) Washington, DC: Government Printing Office.

Radin, Paul. 1924. 'Wappo texts (first series)', *University of California Publications in American Archaeology and Ethnology* 19: 1–147.

——1929. 'A grammar of the Wappo language', *University of California Publications in American Archaeology and Ethnology* 27. Berkeley: University of California Press.

Rankin, Robert L. 1992. 'Review of *Language in the Americas* by Joseph H. Greenberg', *International Journal of American Linguistics* 58: 324–51.

Rédei, Károly. 1988. *Uralisches Etymologisches Wörterbuch*, vol. 1. Wiesbaden: Harrassowitz.

Rensch, Calvin Ross. 1976. *Comparative Otomanguean Phonology.* Bloomington: Indiana University Press.

Rivet, Paul. 1942. 'Un dialecte Hoka colombien: le yurumanguí', *Journal de la Société des Américanistes*, n.s., 34: 1–59.

Robins, R. H. 1958. *The Yurok Language.* (University of California Publications in Linguistics, 15.) Berkeley and Los Angeles: University of California Press.

Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy.* Stanford: Stanford University Press.

Salmons, Joe. 1992. 'A look at the data for a global etymology: *tik "finger" ', in Garry W. Davis and Gregory K. Iverson (eds.), *Explanation in Historical Linguistics.* Amsterdam: Benjamins, 207–28.

Sawyer, Jesse O. 1965. *English–Wappo Vocabulary*. (University of California Publications in Linguistics, 43.) Berkeley and Los Angeles: University of California Press.

Sawyer, Jesse O., and Schlichter, Marie-Alice. 1984. *Yuki Vocabulary*. University of California Publications in Linguistics, 101. Berkeley and Los Angeles: University of California Press.

Schlichter, Marie-Alice. 1985. 'The Yukian language family', Ph.D. dissertation, University of California, Berkeley.

Shipley, William. 1957. 'Some Yukian-Penutian lexical resemblances', *International Journal of American Linguistics* 23: 269–74.

Spalding, A. E. 1969. *Salliq, an Eskimo Grammar*. Ottawa: Education Branch, Department of Indian Affairs and Northern Development.

Swadesh, Morris. 1960. 'On interhemisphere linguistic connections', in Stanley M. Diamond (ed.), *Culture in History: Essays in Honor of Paul Radin*. New York: Columbia University Press, 894–924.

——1965. 'Kalapuya and Takelma', *International Journal of American Linguistics* 31: 237–40.

Swanton, John R. 1907. 'The ethnological position of the Natchez Indians', *American Anthropologist* 9: 513–78.

——1924. 'The Muskhogean connection of the Natchez Language', *International Journal of American Linguistics* 3: 46–75.

Thalbitzer, William C. 1944. 'Uhlenbeck's Eskimo-Indoeuropean hypothesis: a critical revision', *Cercle Linguistique de Copenhague* 1: 66–96.

Torroni, Antonio, *et al.* 1992. 'Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations', *Genetics* 130: 153–62.

Turner, Christy G., II. 1971. 'Three-rooted mandibular first permanent molars and the question of American Indian origins', *American Journal of Physical Anthropology* 34: 229–41.

Wald, Benji. 1992. 'Bantu Languages', in William F. Bright (ed.), *International Encyclopedia of Linguistics*, vol. 1. New York: Oxford University Press, 157–60.

Watkins, Calvert. 1962. *Indo-European Origins of the Celtic Verb*. Dublin: Dublin Institute of Advanced Studies.

Watkins, Calvert. 1990. 'Etymologies, equations and comparanda: types and values and criteria for judgment', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin: Mouton de Gruyter, 298–304.

Winkler, Heinrich. 1884. *Uralaltaische Völker und Sprachen*. Berlin: F. Dümmler.

Winston, F. D. 1966. 'Greenberg's classification of African languages', *African Language Studies* 7: 169–79.

This page intentionally left blank

15

Does Altaic exist?*

Since the writings of Clauson, and more recently Doerfer, it appears that most specialists in the Altaic languages no longer believe that the three groups of traditional Altaic, namely Turkic, Mongolian, and Tungusic, are related; their resemblances are to be attributed to borrowing, or in some cases to accident or sound symbolism.

The term 'traditional Altaic' is here used purposely, that is, without reference to Korean, Japanese, or for that matter Uralic.[1] This is not because I believe that the Altaic languages are genetically isolated. In fact, in my view (Greenberg 1987: 332), they belong to a much larger grouping, Eurasiatic, along with other languages besides those just mentioned above. Moreover, considerations deriving from these wider connections will figure in some instances in an essential way in the following discussion.

There are two separate questions involved here. Are the Altaic languages related to each other? If they are, do they constitute a valid genetic grouping, that is, a set of languages which have a single exclusive common ancestor, Proto-Altaic, which gave rise to three groups of languages and no others?

I believe that the answer to the first question, that of mere relationship, is overwhelmingly positive. That to the second is more difficult, but on the balance I rather strongly endorse a positive answer here also.

Recently in several publications, Miller (1991a, 1991b) has defended the traditional view. His arguments are largely phonological, especially the existence of two reconstructed pairs of liquid phonemes l_1, l_2, r_1, and r_2, which within Altaic are only distinguished in non-Chuvash Turkic. Miller believes that l_1 and l_2 have separate reflexes in Japanese. There are also instances in which Turkic merges a number of phonemes in j, namely d, j, n, and n^y.

[1] It seems clear to me that languages like Korean, Japanese, and Uralic stand apart from traditional Altaic. Thus, Poppe (1960: 8), who includes Korean, shows it as a separate branch from the rest of Altaic, and it figures comparatively infrequently in his etymologies.

In such instances in order to account for the usual anti-Altaicist scenario in terms of borrowing from Turkic into Mongolian (with some reverse borrowing) and then from Mongolian into Tungusic, the borrowing has to be pushed back to a time so early that it becomes indistinguishable from Proto-Altaic, that is, when Turkish still distinguished *d*, *j*, *n*, and *n*^{y}, and all the Altaic languages outside of non-Chuvash Turkic displayed a difference between l_1 and l_2 as well as r_1 and r_2. At such a time the languages would all have had a sound system which is identical with that reconstructed by Ramstedt, Poppe, and others for Proto-Altaic.

Miller also alludes to the cogency of the grammatical data regarding verb derivation in Ramstedt (1912) and Poppe (1973). I agree with him on all of this, but I believe that he has omitted the most powerful evidence of all, that based on personal, demonstrative, and interrogative pronouns.

This material is, of course, familiar, but the anti-Altaicists have, as will be shown, carefully avoided presenting it in a coherent way, and where they have, have sought to explain it away in an unconvincing fashion as the result of factors other than common genetic inheritance.

I will begin with the first and second person pronouns. In the first person singular in non-Chuvash Turkic, some languages, e.g. Osmanli Turkish, have nominative singular *ben* and a stem *ben-* which, except for an internal variation in the dative (*bana*), is found in all the oblique cases. Most Turkic languages, however, have *men* rather than *ben*, and all have *-m* as the first person singular marker in verb forms. The fundamental form then is *me-n*, in which *-n* (often called pronominal *n* by Altaicists) has as its original function a mark of the oblique, ultimately of genitive origin. In non-Chuvash Turkic, this *-n* has spread analogically to the nominative. In Chuvash, however, which represents a separate branch of Turkic, this did not occur. The nominative here is *e-pe** in which *e* is a deictic element, and the oblique stem is *man-*.

This irregular alternation between nominative and oblique recurs in Mongolian in which the nominative is *bi* and the genitive *min-u* and Tungusic, e.g. Evenki, with nominative *bi* and genitive *min-i*. The forms *men* and *min* are much more widespread than Altaic, including Uralic (e.g. Finnish *minä* 'I') and Indo-European. Indo-European appears here as an important link in this chain. On the basis of Baltic, Slavic, and Indo-Iranian, Szemerényi (1970: 197) reconstructs **mene* for the genitive. In Baltic and Slavic, the form in *-n* has been extended to all the oblique cases as in Altaic.

The Indo-European evidence is important because it provides a confirming instance for the oblique case function of the form in *-n*. This is

presumably the same -*n* which occurs in the oblique cases of *r/n* stems.[2] The Indo-European independent nominative is a suppletive form but different from that of Altaic, namely *e-g(h)o-m*, whose most closely related from in Eurasiatic is Chukchee *i-gəm/e-gəm* (vowel-harmony variants) 'I' (cf. *i-gət/e-gət* 'thou'. Forms without the initial vowel occur as bound objects).

Returning to Altaic, it is clear that the probability of an irregular alternation such as *bi/men* occurring three times by accident is infinitesimal. That it should be borrowed twice is also utterly improbable. One has literally to scour the earth to find a few instances of a borrowed pronoun, much less an entire irregular alternation in pronouns. By itself it is enough to show that the Altaic languages are related, moreover the specific innovation of *bi* in the nominative is confined to these languages. Therefore it can be considered a shared common innovation within Eurasiatic that contributes to the establishment of traditional Altaic as a valid genetic entity.

How is this evidence treated by Clauson and Doerfer, the two leading exponents of the anti-Altaicist position? It is ignored where possible. In Clauson (1969: 38), which applies glottochronology to the Altaic problem, discussion is unavoidable since 'I' is part of the glottochronological list. He seeks to argue away the threefold resemblances, indicated by italicized entries, among Old Turkish, Old Mongolian, and Manchu, the three languages he utilizes in his study as follows:

It is known (but has not been explained up to now) that there are phonetic resemblances between personal pronouns in languages which are completely unconnected with each other, e.g. between *mine*, German *mein* and the Turkish genitive *menin* (from *ben*) and Mongolian *minö* [sic!] from *bi*; between Latin *tu* and Mongolian *či* (**ti*). The phonetic resemblances between Turkish, Mongolian, and Tungus-Manchurian in regard to these lexical items cannot be therefore recognized as probative.

This reasoning, which is very common, is to deny the significance of a resemblance because it is found somewhere else. This was used by Michelson against Sapir in regard to *n* first person, *m* second person in Algic because it occurs in so many other Amerind languages. It would be just as logical to deny the significance of the resemblance between English *mine* and German *mein* because it also occurs in Mongolian. One has to pursue the full

---

[2] The oblique -*n*, and indeed all the grammatical elements here were discovered by the Nostraticists. See especially the tables in Illich-Svitych (1971: 6–18). I discovered these independently at a time when I was not aware of Nostratic. In some instances, of course, I have found additional support, especially in languages not included in 'classical Nostratic' but often accepted now as Nostratic, e.g. Chukchi-Kamchatkan and Eskimo-Aleut.

distribution of these forms. As soon as one gets to Sino-Tibetan or Nilo-Saharan, or many others, it ceases. Both the Nostraticists and I include Indo-European and Altaic in the same group.

In addition, Clauson, by simply using the nominative as the translation form for the glottochronological list, fails to consider the agreement between Mongolian and Tungusic in the *bi/min-* alternation, and by not including Chuvash does not have to account for the threefold agreement in an irregularity among the three branches of Altaic.

And what of the second person singular pronouns? They are not discussed at all. Clauson unaccountably does not italicize Old Turkish *sen* and Manchu *si* as resemblances to be explained, or rather explained away, in spite of their complete parallelism with Old Turkish *ben* and Manchu *bi.* Old Mongolian *tere*, Manchu *tere* 'this' are italicized but passed over without comment.

Doerfer in general fails to discuss grammatical resemblances, but in his *Mongolo-Tungusica* (1985: 2), he says the following about the first person singular pronoun:

Indeed, even such an apparently clear comparison as Mongolian *bi*—Tungus *bi* is not convincing on closer examination, since the Mongolian forms (on account of the plural *bi-da*, cf. *e-de* 'these', *te-de* 'those') goes back to *bɨ*. A typical case of sound symbolism (*Elementarverwandschaft*), surface resemblance, but without the possibility of a connection by sound correspondence.

What Doerfer is saying is that Mongolian *i*, which has two sources in a system of back-front vowel harmony, must derive from a high *back* vowel, not a high front vowel, because of the vowel of the second syllable -*da* which is a back vowel.

What Doerfer fails to point out is that Mongolian *bida* is a first person inclusive plural. Now it is a worldwide typological fact that where there is a first person inclusive/exclusive distinction in the plural, the exclusive, when analyzable, is the plural of the first person. This is so in Mongolian, in which the first person is *ba*, with a perfect parallelism between the first and second persons, *bi:ba* = *či* < **ti:ta.*

On the other hand the first person inclusive is either a separate form unlike either the first or second person singular, or it is a combination of the two like Tok Pisin *yu-mi.* Hence *bi-da* is very likely a compound of singular *bi* with *ta* second plural. In compounds vowel harmony need not apply. A parallel situation is found in Tungusic, in which most languages have a first person plural inclusive/exclusive distinction in which the exclusive is the plural of the singular. The same parallelism reigns here as in Mongolian between the

first person and the second person, e.g. Evenki *bi:bu* = *si:su*. The first inclusive is here even more obviously a compound, e.g. Evenki *mi-ti*, *mi-t* (Tsintsius 1949: 270–1).

Note also that Doerfer fails to mention the striking parallelism between the nominative and oblique stems in the first person among Mongolian, Tungusic, and Chuvash. We are to believe that Mongolian *bi* here is not cognate with the Tungusic and Turkic forms in spite of the agreement between them in parallel irregularities. Characteristic also is Doerfer's resort to sound symbolism. This is done without any supporting evidence. Surely *b-* is not particularly frequent as a first person singular in languages of the world, nor is there any plausible support in sound imitation or other sources of *Elementarverwandschaft*.

Finally, it should be noted that violations of back-front vowel harmony are not uncommon in Uralic, a universally accepted family, and in etymologies which are obviously valid on other grounds. As late as 1910, Szinnyei, in his reconstruction of Proto-Finno-Ugric, resorted to a kind of majority rule to determine whether back or front vocalism was the original type in Proto-Finno-Ugric. Even now there are uncertain instances. A parallel situation exists in Turkic. As noted by Radloff (1882: 84) there are variations in stem vowels without any demonstrable cause. In fact there is an article by Dmitrijev on this topic, in which he observes that sporadic alternations in the same root of vowels of the front and back series is frequent in individual Turkic languages (Dmitrijev 1955: 115).

Another one of the very few grammatical etymologies in Doerfer (1985: 27) is his no. 66, the interrogative stem *ya-* of Mongolian and Tungus. He admits that it 'behaves like a genetically related word'. Once more he resorts to 'sound symbolism' and again his only support is Indo-European **jo*. But this is a widespread Eurasiatic interrogative (cf. Greenberg 1990). Once more we have the *ad hoc* resort to a highly implausible sound symbolic argument without any serious documentation.

Finally, what of the second person pronouns? They are passed over in complete silence. Doerfer, like Clauson, believes that Mongolian borrowed massively from Turkic, and then Tungusic from Mongol. He is clearly disturbed by the existence of certain etymologies common to Turkic and Tungusic and devotes a section to them (1985: 238–41), but he fails to mention the most glaring instance of all, the agreement of Turkic and Tungusic in an *s* second person as against Mongol *t*. Of course, if I am right in my discussion of the Mongol and Tungusic first person inclusive pronoun, *t* would also occur in Tungusic, but in a quite different context. Both *s* and *t* are widespread second person Eurasiatic pronouns. For example, we find Indo-European

t in the independent pronoun and plural verb endings and *s* as a singular verb suffix.

In general there are a considerable number of other grammatical markers common to all the Altaic branches, most of them entirely ignored by Doerfer. However, virtually all these are found in other branches of Eurasiatic. The number of these as well as the lexical evidence makes the relationship of the Altaic languages a certainty. However, the distinctness of Altaic as a valid subgroup, which is most conspicuously supported by the *bi/min* alternation in the first person singular pronoun requires further assessment, a task not undertaken here.

References

Clauson, G. 1969. 'Leksikostaticheskaja otsenka altajskoj teorii', *Voprosy Jazykoznanija* 5: 22–51. Moscow.

Dmitrijev, N. K. 1955. 'Cheredovanie glasnyx zadnego i perednogo rjada v odnom i tom zhe otdel'nyx tjurkskix jazykov', in N. K. Dmitrijev (ed.), *Issledovanija po sravitel'noj grammatike tjurkskix jazykov, chast' I: fonetika*, 115.

Doerfer, Gerhard. 1985. *Mongolo-Tungusica*. Wiesbaden: Harrassowitz.

Greenberg, J. H. 1987. *Language in the Americas*. Stanford: Stanford University Press.

——1990. 'Relative pronouns and P.I.E. word order type in the context of the Eurasiatic hypothesis', in Winfred Lehmann (ed.), *Language Typology 1987: Systematic Balance in Language*. Amsterdam: Benjamins, 123–38.

Illich-Svitych, V. M. 1971. *Opyt sravnenija nostraticheskix jazykov*. Moscow: Nauka.

Miller, Roy Andrew. 1991*a*. 'Genetic connections among Altaic languages', in Sydney M. Lamb and E. Douglas Mitchell (eds.), *Sprung From Some Common Source: Investigations into the Prehistory of Languages*. Stanford: Stanford University Press, 293–327.

——1991*b*. 'Anti-Altaicists contra Altaicists', *Ural-Altaische Jahrbücher*, 63: 5–62.

Poppe, Nikolaus. 1960. *Vergleichende Grammatik der altaischen Sprachen, Teil I: Lautlehre*. Wiesbaden: Otto Harrassowitz.

——1973. 'Über einige Verbalstammsbildungsaffixe in den altaischen Sprachen', *Orientalia Suecana*, 21: 119–41.

Radloff, W. 1882. *Phonetik der nördlichen Türksprachen*. Leipzig: Weigel.

Ramstedt, G. J. 1912. 'Zur Verbalstammbildungslehre der mongolisch-türkischen Sprachen', *Journal de la Société Finno-Ougrienne*, 28/3: 1–86.

Szemerényi, Oswald. 1970. *Einführung in die Vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Szinnyei, József. 1910. *Finnisch-ugrische Sprachwissenschaft*. Leipzig: Göschen.

Tsintsius, V. I. 1949. *Sravitel'naja fonetika tunguso-man'chzhurskix jazykov*. Leningrad: Uchpedgiz.

16

The convergence of Eurasiatic and Nostratic*

16.1 Introduction

In Greenberg (1987: 332) a linguistic stock named Eurasiatic is outlined with the following membership: 1. Indo-European; 2. Uralic-Yukaghir; 3. Altaic (Turkic, Mongolian, Tungusic); 4. Korean; 5. Japanese; 6. Ainu; 7. Gilyak; 8. Chukotian; 9. Eskaleut. Of these, 4–6, i.e. Korean, Japanese, and Ainu are stated as probably forming a subgroup. Since then in Greenberg (1990a, 1990b, and 1991), a series of Eurasiatic grammatical markers has been proposed. In addition Greenberg (1992) contains an extensive exposition of sixty-three grammatical items, which is intended as a preliminary version of one of the chapters of Greenberg (to appear) [Greenberg 2000-Ed.], which will present detailed evidence for this family.

The present writer arrived at the Eurasiatic hypothesis some time in the mid 1960s as part of the task of determining the genetic affiliations of languages in the Americas, in this instance, that of the Eskimo and Aleut group. At that time, the Moscow School, which has recently become prominent, was hardly known in the United States and I arrived at my own hypothesis in complete independence of their results. The Nostratic hypothesis is most commonly stated in terms of the genetic relationship among the following six families: Indo-European, Uralic, Altaic, Afroasiatic, Kartvelian, and Dravidian.

Stated in this manner, there are on the surface major differences, the inclusion of Yukaghir, Gilyak, Eskimo, Korean, [Ainu,] and Japanese in Eurasiatic and the exclusion of Afroasiatic, Kartvelian, and Dravidian. There is also, however, a common core consisting of Indo-European, Uralic, and Altaic. Because of this common core the term Eurasiatic/Nostratic is sometimes

employed. However the differences have understandably led some historical linguists, e.g. Watkins (1990: 295) to shed some doubt on both. He (Watkins) asks rhetorically whether these are just trivial differences and whether he is 'nit-picky' in pointing them out. The answer, of course, is that he is not.

However, as will be discussed in detail in the following section, there have been significant changes in the views of Nostraticists in recent years, as the result of which the difference between Eurasiatic and Nostratic has been very greatly reduced. A further point of significance, which will be set forth in detail in the second section, is that the construct of Nostratic as consisting of just these six particular families has never corresponded to the actual views of the Nostraticists, even Holger Pedersen who coined the term 'Nostratic'. Hence, the second section will be devoted to an historical account of Nostratic theory insofar as it relates to the actual membership of Nostratic as conceived by important individual members of this school. Since the position of Afroasiatic in relation to Indo-European, as compared with other families, particularly Uralic, has been a key point in recent developments, the third section is devoted to that topic.

The fourth section of the paper will discuss briefly some points of methodology. What might be called the six canonic branches of Nostratic are based on Illich-Svitych's dictionary (1971), especially the initial table on the correspondences of grammatical elements. The reasons for the exclusion of other languages, some recognized by all Nostraticists even though they figure marginally or not at all in the dictionary and not at all in the tables of grammatical elements, will be set forth. The reasons for the exclusions will lead to a critical discussion of methodological differences between the present writer and the Moscow School, the contributions which each can make, and the affirmation of the identity of goals in the two approaches.

16.1.1 *Recent changes in the conception of Nostratic*

Perhaps the best summary of the overall significance of recent developments in Nostratic theory is that of Lamb and Mitchell (1991: 123), namely, that in recent years Russian comparativists have revised their classification so that it is now closer to my Eurasiatic in two important respects. One is that Afroasiatic is now generally viewed as a sister superstock to Nostratic, rather than part of it. The other is that they have added further stocks to the northeast extending as far as Eskimo-Aleut. They note that an important question still to be resolved is that of Dravidian, to which I would add Kartvelian. The exclusion of Afroasiatic as being on the same level as the rest is likewise noted in Starostin (1989: 43) who refers to his 'cautious approach' to Afroasiatic which he now prefers to exclude from Nostratic comparison.

Fleming (1990: 3) takes note of Starostin's opinion and adds that Bomhard (personal communication) now feels that Afroasiatic is distinct from the rest of Nostratic.

Shevoroshkin and Manaster Ramer (1991: 179) along similar lines note that Afroasiatic may be coordinate with Nostratic, a 'sister' and not a 'daughter'. They add that the Russian investigators Golovastikov and Dolgopolsky (1972) and Mudrak (1984) have presented evidence for the Nostratic affiliations of Chukotko-Kamchatkan and Eskimo-Aleut respectively.

A comprehensive recent statement is that of Bomhard (1992) in which Eurasiatic is given status as a valid genetic group within a wider Nostratic family (cf. also Bomhard 1998). His Eurasiatic is identical with mine except for the omission of Ainu, Japanese, and Korean. However Korean, with some reservation, is included in his Altaic, so that the only real difference is the omission of Japanese and Ainu. The former of these is admitted to have Altaic elements but in view of Benedict (1990), a work which connects Japanese with Tai-Austronesian, judgement is reserved. Regarding Ainu Bomhard reserves judgement because he has not studied the question in any detail (Bomhard, personal communication).

16.2 History of views regarding the membership of Nostratic

As noted earlier, most of the recent developments regarding the membership of Nostratic have historical antecedents. It is particularly striking that in recent statements concerning the separate status of Afroasiatic, no mention is made of the fact that Pedersen, the founder of Nostratic theory, repeatedly stated similar views. I will consider here in turn the views of Pedersen, Illich-Svitych, and Dolgopolsky, who independently of Illich-Svitych developed very similar ideas during the same period.

16.2.1 *Holger Pedersen*

The term Nostratic was first introduced in 1903 as a designation of language and language families related to Indo-European (< Lat. *nostrates* 'our countrymen'). In subsequent writings, the following points are emphasized. The relationship of Indo-European to Uralic is much closer than that of Indo-European to Semitic (i.e. Afroasiatic, of which it is a part, in present terminology). The resemblance to Uralic is especially stressed. For example, in regard to pronouns he writes that you will not find such a thorough agreement a second time (1935: 330). Again in relation to Uralic and Indo-European (1931: 337) he maintains that after the work of Wiklund and

Paasonen, it is unnecessary to doubt the relation further, or again that to deny the relationship between these families would be overbold (1931: 331).

In connection with the closer relationship of Indo-European to Finno-Ugric than to Semitic he makes the interesting remark (1933: 309) that Finno-Ugric is to be compared to Indo-European in its post-Ablaut stage but should be compared to Semitic in its pre-Ablaut stage. He goes on to draw once more the conclusion that the separation of Indo-European and Semitic was at a more ancient period than that between Indo-European and Finno-Ugric. He seems to be referring to Indo-European qualitative Ablaut usually stated as *e ∼ o*. In Greenberg (1990*a*), for which see further details, an attempt is made to show that this Ablaut is part of a larger system of vowel alternations that has left traces in Indo-European itself and which corresponds quite clearly with a system of vowel height harmony (or more accurately originally ± ATR) found in Tungusic, Gilyak, Chukotian and Korean.

A second major point is that Nostratic contains a whole series of languages in Northern Asia besides Finno-Ugric. In this connection Pedersen (1931: 337) mentions similar, though fainter, resemblances throughout all northern Asia, in Turkish, Mongolian, Manchu, Yukaghir, and Eskimo. In fact, as early as 1908, after discussing first person pronoun *m* and second person *t*, Pedersen mentions Indo-European, Uralic, and Altaic and then indicates, referring to Uhlenbeck's Indo-European and Eskimo hypothesis, that this pattern also is found in Eskimo (Pedersen 1908: 342–3). Here as elsewhere he states that he cannot draw definite bounds to the extent of Nostratic and says that this should not cause concern (Pedersen, *loc. cit.*).

We may sum up Pedersen's views as follows. Indo-European is related to Semitic and to a whole series of languages in northern Asia, including Finno-Ugric and Samoyed (now grouped together as Uralic), Yukaghir, Altaic, and Eskimo. The relationship of Indo-European to Finno-Ugric is closer than that to Semitic. Nowhere does Pedersen mention Dravidian or Kartvelian, but for that matter he does not mention Japanese, Korean, or Ainu either.

With regard to Japanese, Starostin (1991), a leading Nostraticist, has recently devoted an entire volume to the connection between Japanese and Altaic. I believe that this accords with the opinion of almost all Nostraticists, Bomhard being an exception in this respect.

16.2.2 *Vladislav Illich-Svitych*

Illich-Svitych is generally regarded as the founder of modern Nostratic theory. His earliest comprehensive statement was published in 1966 in the form of a series of etymologies containing forms from the six families which 'at the least' according to him make up Nostratic. It is interesting that in

Illich-Svitych (1964), called significantly 'Oldest Indo-European–Semitic Language Contacts', the author considered that the case for a relationship between Semitic and Indo-European was weak and that most of the resemblances were due to borrowing from Semitic by Indo-European. In the 1971 work, however, Afroasiatic is included. We may conjecture that his views concerning Semitic and Indo-European were modified by putting them into a broader context, on the one hand with the replacement of Semitic by Afroasiatic of which it is but one branch—and, on the other, with the inclusion of other groups beyond Indo-European such as Uralic. Just so, if one takes English and French in isolation their genetic connection is less evident than if one compares the more extensive Germanic and Romance groups to which they belong respectively.

In 1971, there appeared the first volume of the Nostratic dictionary which has become the standard source for Nostratic etymologies. It was edited by Dybo after Illich-Svitych's untimely death in 1966.

Although the dictionary continues the tradition of comparing only six main groups, Korean is included as a member of Altaic, following Ramstedt and Poppe. In the introduction written by Illich-Svitych (1971: 61) he notes in reference to Yukaghir that the work of Collinder, Angere, and Tailleur, while not allowing us to consider Yukaghir a Uralic language does allow us to consider its Nostratic character. I believe that this agrees with the conclusions of the scholars mentioned by Illich-Svitych who did not consider Yukaghir a Uralic language but rather thought that it was related to Uralic as a whole. Dybo, in an editorial footnote to the passage just cited, adds that the same remarks probably apply to Korean and Japanese in relation to Altaic.

The question of the status of Yukaghir has an interesting history. Traditionally it has been considered along with Ket, Chukotian, and Gilyak as belonging to Paleo-Siberian, a nongenetic grouping. In 1907, Paasonen in an article designed to refute the hypothesis of a relationship between Finno-Ugric and Indo-European, stated that this was untenable because not only were many of the resemblances between Finno-Ugric and Indo-European found in Yukaghir but in addition there were further similarities between Finno-Ugric and Yukaghir which did not occur in Indo-European. Since Yukaghir was an isolated language not known to be related to any other, *a fortiori* Finno-Ugric was not related to Indo-European.

It is of course very risky to assume that two languages are not related. Following Paasonen, Collinder (1940), Ankeria (1956) and Tailleur (1959) all drew the conclusion that Yukaghir was related to Finno-Ugric (by then called in more modern terms Uralic and consisting of Finno-Ugric and Samoyed), and that the Uralic relationship was even closer to Yukaghir than

to Indo-European, assuming the latter was valid. Among items common to all three were pronouns of the first and second persons, several interrogative and demonstrative pronouns, and a marker for the ablative case.

16.2.3 *Aron Dolgopolsky*

About the same time that Illich-Svitych began to publish regarding Nostratic, Dolgopolsky independently developed a theory linking Indo-European with Afroasiatic, Kartvelian, and a series of languages in Northern Asia that includes Uralic, Altaic and Eskimo-Aleut (Dolgopolsky 1964, 1965). He called his hypothesis Sibero-European and criticized the name Nostratic (no doubt because of its ethnocentric character). Nevertheless, he later adopted the more widely known term. In a later work on personal pronouns (1984), he included Gilyak and Chukotian along with Elamite and Dravidian.

From this historical review, it is clear that the Nostraticists, beginning with Pedersen, never restricted their notion of Nostratic to the six groups usually mentioned. The special, more distant position of Afroasiatic is already insisted on by Pedersen. Moreover, every group I include in Eurasiatic, with the exception of Ainu, is included in some of their enumerations, and often in their comparisons.

16.3 The position of Afroasiatic

From the foregoing historical exposition it should be clear that the key problem is the position of Afroasiatic in regard to Indo-European as compared to that of the latter in relation to a series of languages in Northern Asia, beginning with Uralic.

The inclusion of Afroasiatic at the same level as such groups as Uralic clearly has its roots in the earlier attempts to link Semitic and Indo-European. This was obviously motivated by nonlinguistic considerations as shown most vividly and in terms that would today probably be characterized as racist by Cuny (1937: 142) when he asserts that Pedersen 'did not hide his faith in the single origin of the languages of the white race'.

Now one could, of course, be accidentally right. However after the establishment of Afroasiatic, the larger family in which Semitic has a place, it became obvious that the comparison must be with Afroasiatic as a whole and this we see in fact in Illich-Svitych's dictionary.

In regard to languages to the east of Indo-European, most of the earlier discussion centered on the connection between Indo-European and Uralic, largely because of the size and importance of the latter, its geographical proximity to Indo-European and the existence of a considerable body of

specialists who have produced important comparative work. However my own work as well as recent reconsideration of the problem as a whole by a number of Nostraticists has led to the conclusion that, as shown most strikingly perhaps in grammar, Indo-European, as compared to Afroasiatic, shows significantly greater resemblance not only to Uralic but to a whole group of languages in Siberia extending as far as Eskimo.

16.4 Some problems of methodology

Since such languages and language groups as Yukaghir, Chukotian, and Eskimo-Aleut are repeatedly mentioned as Nostratic beginning with Pedersen, why were they not included in the Nostratic dictionary? The answer is given in a statement by two Nostraticists, Chejka and Lamprecht (1984: 86). After discussing Pedersen and Illich-Svitych, they remark, regarding the six groups almost exclusively cited in the latter's comparative dictionary: 'Obviously this does not mean that the number of Nostratic families in the world is confined to the six mentioned. Illich-Svitych in his generalization used only those language families for which the proto-linguistic bases have progressed to a satisfactory level.'

   This, incidentally, helps to explain the absence of Ainu from Nostratic comparisons, since it has not yet figured in any comparative reconstruction and, unlike Korean and Japanese, does not have a long period of documented attestation. This, of course, does not accord with the practice of Indo-Europeanists, who have not hesitated to include such poorly attested languages as Phrygian in their comparative study. Patrie (1982) compares Ainu with Korean and Japanese, considering them to form a genetic grouping, and then compares Ainu to Altaic. Even Refsing, obviously conservative in these matters, in his grammar of Ainu cautiously endorses Patrie's results.

   The result of these restrictions is that 'classical' Nostratic is an arbitrary group of related languages which do not have an exclusive common ancestor and do not therefore constitute a linguistic family in the ordinary sense.

   This issue has now become a topic of discussion among the Nostraticists themselves. It is broached by Bomhard (1992: 62) and in his 'Nostratic, Eurasiatic and Indo-European' (this volume) [Bomhard 1998-Ed.]. He states that there are now two main approaches which might be termed 'taxonomy first' and 'reconstruction first'. He himself favors the former but believes that the two can inform and further one another. In giving the edge to taxonomy he notes that, after all, one cannot successfully reconstruct until one has first established which languages have a reasonable chance of being genetically related; that is to say, one must know which languages to compare.

Finally, a word should be said about inspection. After all, it is a synonym for observation and observation is the first step in any empirical science. The Nostraticists themselves must have done some inspection. They did not first compare just any language groups for which reconstruction had been carried out. For example, the study of Proto-Austronesian was well advanced but it was not included in Nostratic. Presumably this requires that one would have noticed resemblances among the Nostratic languages as a background to the rejection of Austronesian, before actually proceeding to the reconstruction of Proto-Nostratic.

When one inspects one notes not only similarities and, of course, degrees of similarity, but also differences. In my own work and that of the Nostraticists, there is not only substantial agreement on the existence of a valid genetic group almost identical with Eurasiatic but also on the exclusion of Ket, a language traditionally labeled Paleosiberian, whose affiliations are rather with Sino-Tibetan and North Caucasian (cf. Starostin 1984).

I agree with Bomhard regarding the value of the interaction of the two approaches. The Nostraticists have discovered a large number of widespread lexical etymologies and grammatical markers which will be of great value in attaining the common goal of a taxonomically correct and comprehensive classification of the languages of the world and extending vastly the reach of the comparative method of reconstruction and our knowledge of linguistic change.

References

Ankeria, J. 1956. *Die Uralo-jukagirische Frage: Ein Beitrag zum Problem der sprachvergleichenden sprachlichen Urverwandschaft*. Stockholm: Almqvist & Viksell.

Benedict, Paul K. 1990. *Japanese/Austro-Tai*. Ann Arbor: Karoma.

Bomhard, Allan R. 1992. 'The Nostratic macrofamily (with special reference to Indo-European)', *Word* 43: 61–84.

——1998. 'Nostratic, Eurasiatic, and Indo-European', in Joseph C. Salmons and Brian D. Joseph (eds.), *Nostratic: Sifting the Evidence*. Amsterdam: John Benjamins, 17–49.

Chejka, M. and Lamprecht, A. 1984. 'Nostratichnata ipoteza, sŭstojanie i perspektivi', *Sŭpostavitelno Ezikoznanie* 9: 86–92. Sofia: Sofia Universitet.

Collinder, Björn. 1940. 'Yukagirisch und Uralisch', *Uppsala Universitets Årsskrift* 1940: 8.

Cuny, Albert. 1937. 'Chamito-Semitique et Indo-Européen', *Melanges de linguistique et de philologique offerts à Jacqu. van Ginneken*. Paris: Klincksieck, 141–7.

Dolgopolsky, Aaron. 1964. 'Gipoteza drevnejshego rodstva jazykovyx semej severnoj Evrazii s verojatnostnoj tochki zrenija', *Voprosy jazykoznanija* 2: 53–63. (English

translation: 'A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia', in V. Shevoroshkin and Thomas L. Markey (eds.), 1986. *Typology, Relationship, and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists.* Ann Arbor, Mich.: Karoma, 27–50.

——1965. 'Metody rekonstrukcii obshcheindoevropejskogo jazyka i sibir-oevropejskaja gipoteza', *Ètimologija* 1964: 259–70.

——1984. 'On personal pronouns in the Nostratic languages', in O. Gschwantler, K. Rédei, and H. Reichert (eds.), *Linguistica et Philologica, Gedankschrift für Björn Collinder.* Vienna: Braumüller, 65–112.

Fleming, Hal, (ed.). 1990. *Mother Tongue: Newsletter of the Association of Language in Prehistory* 12, December 1990.

Golovastikov, A. N. and Dolgopolsky, Aaron B. 1972. 'Rekonstruktsija chukotsko-korjakskix kornej i nostraticheskie etimologii', *Nauko-konferentsija po srasvitel'noj-istoricheskoj grammatike.* Moscow: Nauka.

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford: Stanford University Press.

——1990a. 'The Prehistory of the Indo-European Vowel System', in Shevoroshkin 1990, 77–136.

——1990b. 'Relative pronouns and P.I.E. word order in the context of the Eurasiatic hypothesis', in Lehmann 1990, 123–38.

——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: Stanford University Press, 127–40.

——1992. *The Grammatical Evidence for Eurasiatic.* Stanford University MS.

——2000. *Indo-European and its Closest Relatives: the Eurasiatic Language Family, vol. 1: Grammar.* Stanford: Stanford University Press.

Illich-Svitych, V. M. 1964. 'Drevnejshije indoevropejsko-semitiskije jazykovyje kon-takty', in V. N. Toporov (ed.), *Problemy indoevropejskogo jazykoznanija.* Moscow: Nauka, 3–12.

——1971. *Opyt sravnenija nostraticheskix jazykov*, vol. IV. ed. A. Dybo. Moscow: Nauka.

Lamb, Sidney M. and Mitchell, E. Douglas. 1991. 'Editorial introduction to Part III', in Sydney M. Lamb and E. Douglas Mitchell (eds.), *Sprung From Some Common Source: Investigations into the Prehistory of Languages.* Stanford: Stanford University Press, 123–4.

Lehmann, Winfred P. (ed.). 1990. *Language Typology* 1987*: Systematic Balance in Language.* Amsterdam & Philadelphia: John Benjamins.

Mudrak, O. A. 1984. 'K voprosu o vneshnyx svjazax eskimoskix jazykov', in I. F. Vardul' (ed.), *Lingvisticheskaja rekonstruktsija i drevnejshaja istorija vostoka.* Moskow: Nauka, 64–9.

Paasonen, Heiu. 1907. 'Zur Frage von der unverwandschaft der Finnisch–ugrischen und indoeuropäischen Sprachen', *Finnisch–Ugrische Forschungen* 17: 13–31.

Patrie, J. 1982. *The Genetic Relationship of the Ainu Language.* Honolulu: University of Hawaii Press.

Pedersen, Holger. 1903. 'Türkische Lautgesetze', *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 57: 535–61.

Pedersen, Holger. 1908. 'Die idg-semitische Hypothese und die idg. Lautlehre', *Indogermanische Forschungen* 22: 341–65.

—— 1931. *Linguistic Science in the Nineteenth Century.* Cambridge, Mass.: Harvard University Press.

—— 1933. 'Zur Frage der Urverwandschaft des Indoeuropäischen mit dem Ugrofinnischen', *Mémoires de la Société Finno-Ougrienne* 67: 308–25.

—— 1935. 'Il problema delle parentele tra i grandi gruppi linguistici', *Atti del Congresso Internazionale dei Linguisti, Roma* 1933, 328–33. Florence: Felice de Monnier.

Refsing, K. 1986. *The Ainu Language: The Morphology and Syntax of the Shizunai Dialect.* Aarhus: Aarhus University Press.

Shevoroshkin Vitalij (ed.). 1989. *Explorations in Macro-Families.* Bochum: N. Brockmeyer.

—— 1990. *Proto-Languages and Proto-Cultures: Materials from the First International Interdisciplinary Symposium on Language and Prehistory.* Bochum: N. Brockmeyer.

Shevoroshkin Vitalij and Manaster Ramer, Alexis. 1991. 'Some recent work on the remote relations of languages', in Lamb and Mitchell 1991, 178–99.

Starostin, Sergei. 1984. 'Gipoteza o geneticheskix svjazax sinotibetskix jazykov s enisejskimi, severnokavkazskimi jazykami', *Lingvisticheskaja rekonstruktsija i drevnejshaja istorija vostoka* 4 (Moscow): 19–38.

—— 1989. 'Nostratic and Sino-Caucasian', in Shevoroshkin 1989, 42–66.

—— 1991. *Altajskaja problema i proisxozhdenije japonskogo jazyka.* Moscow: Nauka.

Tailleur, O. G. 1959. 'Plaidoyer pour Youkaghir branche orientale de la famille ouralienne', *Lingua* 8: 403–23.

Vardul', I. F. (ed.). 1984. *Lingvisticheskaja rekonstruktsija i drevnejshaja istorija vostoka.* Moscow: Nauka.

Watkins, Calvert. 1990. 'Etymologies, equations and comparanda: types and values and criteria for judgement', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology.* Berlin: Mouton de Gruyter, 289–304.

Part IV

Genetic linguistics and human history

This page intentionally left blank

Linguistic typology and history: Review of *Linguistic Diversity in Space and Time* by Johanna Nichols*

Linguistic Diversity in Space and Time is a book with a highly complex set of arguments and, to this reader at least, a frequently obscure style which hinders comprehension. Moreover, it is provided with no fewer than ninety-six tables, mostly with numerical entries regarding either the absolute or relative frequency of certain typological properties as distributed over various areas of the globe or the statistical significance of differences among such distributions. In addition there are twelve maps containing worldwide plots of the occurrence of the specific typological features on which Nichols has based her arguments. She deserves credit for being willing to attack such large-scale historical problems in linguistics at a time when neither historical nor broad-scale efforts are the fashion.

Obviously, in a brief review it will not be possible to consider all her arguments in detail. However, the heart of the argument is that the comparative method on which historical inferences have been centrally based can by its nature give results only for relatively recent time depths. Nichols wishes to use instead population typology, which 'gives us the heuristic method that standard comparative-historical method lacks for great time-depths' (p. 3). Another leitmotiv alongside of population typology is that diversity as such is a worthy subject of study. Regarding this she notes that linguistics needs a theory of diversity (p. 231):

If we are to have a theory of diversity, we must have ways of describing diversity, information on its distribution throughout the world, and ways of explaining variations in degree and types of diversity. We then apply the comparative method

to sheer diversity and draw inferences about the relative chronology and the mechanism of the spread of human languages over the earth.

Her third major emphasis is on geography. For example, there are certain zones which might be called spread zones and distinguished from residual ones in regard to the distribution of linguistic diversity. An example of a spread zone is the western Eurasian steppe, of a residual zone the Caucasus. Rather more significant, she divides the languages of the world into ten principal areas, for example, Africa (but omitting North Africa), Northern Eurasia, New Guinea, and Mesoamerica. For her major conclusions, however, she groups these ten areas into just three macroareas: the Old World, the Pacific, and the New World. She believes that it is important to keep her areas discrete, hence the omission of North Africa (between sub-Saharan African and the Near East) and of Eskimo-Aleut (between North America and Northern Eurasia). 'Northern Eurasia is kept discrete from South and Southeast Asia by not taking any languages for the North Eurasian area from its Southern periphery' (p. 26). (This I find vague and hard to interpret.) Finally, 'Mesoamerica is kept discrete from North and South America by arbitrary use of political boundaries.' Elsewhere, she mentions the omission of Indonesia, no doubt to separate South and Southeast Asia from Oceania (Melanesia, Micronesia, Polynesia). Thus languages from the Philippines, Java, etc., do not figure in her sample.

Evidently diversity, linguistic population typology, and geography are connected in the following way: in each area or macroarea, the populations, which consist of languages classified by certain typological criteria, are examined for diversity within and between areas in regard to the relative frequencies of the typological properties which have been selected. From this we will deduce the ways in which languages have spread over the world at times too remote to be amenable to the comparative method.

Clearly, we cannot look at all the languages of the world (even with the exclusion of the areas mentioned earlier), so we must sample. Nichols's sample is 176 languages. She tries to get one language from each lineage, by which is meant either a genetic isolate or a family, ideally a subdivision of a stock. (For example, Indo-European is a stock, but Germanic is a family.) She is aiming here at a time depth of 2,500–4,000 years (p. 24). Here her view that valid linguistic stocks can be detected at very shallow chronological depths is essential. The resultant extreme splitting in classification has a profound effect on her sampling.

For example, she evidently considers Northwest and Nakh-Dagestan in the Caucasus unrelated. Their distinction from Kartvelian in the southern

Caucasus is not at issue. However, the relationship of Northwest and Nakh-Dagestan seems evident, and Catford (1977: 254), probably the world's leading expert on Caucasian, simply states that there are two major groups of Caucasian languages, Northern and Southern. Here as elsewhere, whatever view gives the larger number of stocks is accepted without referring to dissenting opinions. Another example is Altaic, which we learn 'is now abandoned'. The reference here is to the summary report of the Altaic panel by Unger (1990). This is a mere four-page effort which does not cite any linguistic facts, and a significant note appended to it shows that it is not fully representative: 'Changes in the makeup of the committee which were necessitated by defection of the original members made the final group somewhat less sympathetic to Altaic than the original.' It is to her credit that she is troubled by what is no doubt the strongest evidence for Altaic, namely, that 'the pronominal roots are just too similar, in both basic consonantism and patterns of suppletion, to be the product of chance' (p. 6). She believes, however, that the lexical resemblances are too few and that they have (all?) proved invalid—an evident exaggeration.

We are faced, then, with a paradox. Either the Altaic languages are related—and very closely, because the pronominal resemblances are comparable, according to her, to those within a single branch of Indo-European such as Germanic—or they are not, because lexical evidence is absent. (Actually, resemblances of the kind found in Altaic pronouns occur between different branches of Indo-European, for example, 'I', 'me', 'thou' = Russian *ja, menja, tu* = Italian *io, mi, tu*, etc.) However, according to Nichols the population-typological approach offers other explanations 'for the striking similarities in genetically unrelated languages'. These resemblances are most obvious in personal pronouns but are found elsewhere. Her solution is that pronouns are sound-symbolic. How this derives from the population-typological approach is hard to see. In regard to *m* 'first person', *t* 'second person', which she acknowledges occurs in a dozen or so Eurasian stocks (to which one should add Eskimo-Aleut) (p. 313 n. 3), Wundt offered a sound-symbolic explanation according to which *m* stands for the nearer and *t* for the farther.

To her the sound-symbolic properties of personal pronouns are comparable to those of 'mama' and 'papa' words but even stronger (pp. 261–2):

Specifically, personal pronoun systems the world over are symbolically identified by a high frequency of nasals in their roots, a strong tendency for nasality and labiality to co-occur in the same person form, and a tendency to counterpoise this form to one containing a dental. In the Old World, the labial and nasal favor the first person; in the New World, they favor the second person. The Pacific is intermediate, with

a distribution of dentals like that of the Old World and nasals like that of the New World.

What we have here is an involved way of saying that the Old World has *m* in the first person and *t* in the second person and the New World *n* in the first person and *m* in the second person. The Pacific I cannot interpret. But the *m/t* pattern occurs only in Europe and Northern Asia in the Old World, all in members of what I call the Eurasiatic family. Elsewhere, the patterns are quite different (e.g. *a/i* in Nilo-Saharan). In the New World, Na-Dene again is quite different and Eskimo-Aleut actually goes with the Old World. But why should there be this shifting between labial/dental in the Old World and dental/labial in the New World (subject to the strong genetic and areal limitations I have already noted)? The answer is that their meanings are 'shifters' (Jakobson's term) or deictics (now the more usual term) and in this way are unlike the *mama-papa* terms, but this is at best a bad metaphor. To equate Eurasiatic *m* 'first person' with Amerind *m* 'second person' requires that first to second person or vice versa be a well-attested meaning change. I do not know of any examples.

We are told that pronouns are subject to phonetic wear, etc., but the important process of replication is omitted. An independent pronoun can be affixed to verbs to indicate subject or object and to nouns to indicate possession and continue in its original form, thus multiplying the chances for survival. The real proof that this is all special pleading to get rid of inconvenient evidence is the coherent distribution of these forms along with other resemblances defining deeper groupings than Nichols will allow.

But what does it matter, one may ask, if she is sampling at depths of 2,500–3,000 years consistently? The problem is that, for all of her evident statistical sophistication, she seems unaware of Galton's problem, that of the independence of cases, because she proceeds purely typologically. For example, she calculates the percentage of languages in her sample with the distinction between inclusive and exclusive first person plural pronouns in each of the three macroareas but never asks whether the phonetic forms are similar and explainable by genetic inheritance so that they are in effect a single case. This is a matter to which I will return.

The conclusions that Nichols develops depend mainly on a 'global cline' among the macroareas which she finds for a number of typological characteristics. For example, 22 per cent of her sample in the Old World, 48 per cent in the New World, and 57 per cent in the Pacific have the inclusive/exclusive distinction. From this and a few other similar distributions she draws the conclusion that 'an early typological bifurcation took place in Southeast Asia

or the Western Pacific and the New World underwent multiple colonization by a circum-Pacific population with an Old World admixture that increased over time' (p. 207). This is about the closest we come to any actual historical hypothesis—one without geographical detail or chronology. In the Pacific, however, five of the twenty-eight languages that have the inclusive/exclusive distinction are Austronesian, and the distinction occurs in an additional Austronesian language, Acehnese, in Southeast Asia—in the Old World macroarea. All of these go back to the same proto-Austronesian forms. Sixteen out of the eighteen Australian languages in the group have the distinction, and these can, at least for the non-Pama-Nyungan languages, be reconstructed for the protolanguage. It is striking that hypotheses regarding global clines and temporally remote circum-Pacific movements should be drawn from Austronesian forms reconstructable for probably less then 2,000 years ago, all counted as independent cases and even occurring in two different macroareas.

Other problems abound. One is that, using another of her key concepts, genetic density (i.e. number of genetically distinct stocks per unit area), Nichols notes that 'the Old World shows low genetic density even under circumstances which should favor high genetic density.' It is a long-established principle that earlier-settled areas show greater diversity—but if, instead of 150-odd stocks in the Americas, there are only three, the difficulty disappears. Again, she finds problems reconciling Beringia as the entry point for the settling of the New World, a notion she accepts alongside her theory of circum-Pacific colonization (p. 228). Related to this is her problem with the physical evidence (pp. 224–5): 'the typological affinities of the linguistic population of the New World are with Melanesia (although the physiological affinities of its human population are clearly with Siberia).'

The basic fallacy of the book is the notion that we can use statistics concerning the relative frequencies of typological features in different areas to reconstruct remote prehistory. It is rather the distribution of such typological features (which themselves normally allow very limited possibilities, for example, the presence versus absence of the inclusive/exclusive distinction) that itself requires historical explanation. It can be inherited within small or large families, the result of areal contact, or a quite recent independent innovation. Thus, from the historical point of view, typological distributions are *explananda*, not explanatory principles.

Although, as should be evident, I consider this book to be fundamentally flawed, one cannot but admire the author's willingness to work on a vast scale. She clearly has an extensive knowledge of the world's languages, and

in the course of her work she has noted some important typological implicational universals.

References

Catford, J. C. 1977. 'Mountain of tongues: the languages of the Caucasus', *Annual Review of Anthropology* 6: 283–314.

Unger, J. Marshall. 1990. 'Summary report of the Altaic panel', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin: Mouton de Gruyter, 479–82.

18

Are there mixed languages?*

Thomason and Kaufman, in their widely influential volume on language contact, creolization and genetic linguistics (1988), while not denying that the genetic model describes the differentiation of earlier languages into subsequent groups of languages related by common descent in the vast majority of instances, emphasize two particular cases in which they state that the genetic model does not fit and for which they posit language mixture. One of the two languages is Ma'a in Tanzania, which they state has two components, one of Southern Cushitic provenience and hence a part of Afroasiatic and the other Bantu, hence a part of Niger-Kordofanian. The other instance is Mednyj Aleut (i.e. the Aleut of Copper Island), in which there are two sources, Aleut and Russian. The greater part of this paper is devoted to the former case, both because of the writer's background as an Africanist and the larger documentation in the published literature. The problem of creole languages, which is of a quite different nature, is not within the scope of this paper.

The case of Ma'a and Mednyj Aleut, since they are probably the most frequently cited instances of mixed languages, will serve as *a fortiori* proofs of the universal applicability of the genetic model. Both Ma'a and Mednyj Aleut are cited as instances in which, to state the facts in summary and over-simplified fashion, the mass of the vocabulary is of one provenience (Cushitic, Aleut) while the grammars, specifically major portions of the morphology, have a different origin, Bantu and Russian respectively.

It is not difficult to see that at the present time, the genetic model which has been the foundation of comparative linguistics is under a broader attack than merely the positing of relatively rare cases of mixed languages. The genetic interpretation of widespread resemblance in both sound and meaning in large groups of languages, e.g. in the Americas, as somehow due to language contact over vast areas and in the most stable elements of language, however implausible on common-sense grounds, is basically a reaction by specialists to any evidence of wider connections which would sometimes put

in jeopardy their own historical explanations based on a restricted group of languages and/or falsify already published negative statements.

We now turn to Ma'a, known in the earlier literature as Mbugu. The basic descriptive facts about Ma'a are known from Shaw (1885), Meinhof (1906), Copland (1934), Tucker and Bryan (1957) and Green (1963). Of those who have written about the genetic status of Ma'a, the fullest account of the facts concerning the language based on the sources just enumerated is Goodman (1971). It is not my purpose to recount all of these facts in detail. The most basic may be summarized as follows. The nominal vocabulary, except for a few nonbasic terms which are obvious loans from Bantu sources, is non-Bantu. There are a number of obvious resemblances in basic vocabulary to Southern Cushitic languages. Ehret (1976: 85), who has along with Fleming done the largest amount of comparative work on Cushitic and on Southern Cushitic in particular, notes that the highest percentage of lexical cognates of Ma'a is with Oromo (Galla) in Eastern Cushitic, namely 14 per cent. In general, the proportion of lexical cognates among Cushitic languages is low. I have long suspected that there is an especially close relationship between Southern and Eastern Cushitic. As noted above, there are some obvious loans from Bantu among nominal stems. In the verbs, which are generally far more impervious to borrowings from foreign sources in languages, there are no Bantu stems but a few have cognates in Cushitic languages.

To what extent does the morphology have a Bantu source? The most conspicuous Bantu elements are the well-known noun-class prefixes which are generally paired as singular and plural. These paired classes have semantic correlates to varying degrees. For example, the members of the classes traditionally numbered 1 and 2 designate human beings in the singular and plural respectively e.g. Proto-Bantu *mu-ntu* 'person', *ba-ntu* 'people', the latter being the source of the designation Bantu for the linguistic group.

However unlike Bantu languages in general, possessives and demonstratives in Ma'a do not agree with the noun in class and in fact have single forms which do not vary. The pronominal subject and object are, however, part of the verb complex, are Bantu in form and in the third person agree with the subject and/or object of the verb in noun class. The independent pronouns are clearly non-Bantu and a number of them have good Cushitic etymologies.

Before considering some facts about the history of the Ma'a and certain pertinent details about how the Bantu and Cushitic elements function in the language, we may consider the conclusions regarding the genetic affiliations of Ma'a which have been advanced by various writers who have either themselves compiled linguistic data from the Ma'a or have written about the problem of its linguistic affiliations.

The two main positions are that Ma'a is to be classified as a Cushitic language which has been strongly influenced by Bantu or is a mixed language. Members of the former group include Copland (1934), the present writer (Greenberg 1953)[§1-Ed.], Welmers (1973: 9), Ehret (1976, 1980), Elderkin (1976) and Sim (1994: 799). Those who have considered Ma'a a mixed language included Whiteley (1960), Tucker and Bryan (1974) and Thomason and Kaufman (1988). The reason given by these writers is that basic vocabulary derives from Cushitic and the grammar from Bantu and because of these two sources the language is to be considered mixed. A fuller treatment of Thomason and Kaufman's views is reserved for a later section after a consideration of Mednyj Aleut since these writers are the only ones who consider the two cases and present analogous treatments of both.

The only writer who has considered Ma'a to be unequivocally Bantu is Dolgopolsky (1973: 24). His reason is that languages are to be classified by their grammar and since most of the grammar in Ma'a has a Bantu source, the language should be classified as belonging to that group.

Two views which deserve separate consideration as not falling into the trichotomy Bantu, Cushitic, or mixed are those of Meinhof (1906) and Goodman (1971). Meinhof, the founder of modern comparative Bantu studies, believed that Ma'a had started as a Sudanic language. It then was absorbed into a Cushitic language. Meinhof must be credited with having first noted important vocabulary resemblances to Mbulunge, the best known Southern Cushitic language at that time. Ma'a then fell under Bantu influence. While Meinhof's overall ideas regarding classification are of course not acceptable in present-day African linguistic studies, he is correct in noting (1906: 295) that the Bantu prefixes and vocabulary are borrowed elements.

Goodman, after his careful and elaborate consideration of the evidence, first notes that (1971: 253) 'A comparison of the Bantu and non-Bantu portions of the Mbugu vocabulary shows that the former is more likely to have been borrowed than the latter.' However, his final conclusion is simply one of puzzlement, namely that 'the development which Mbugu has undergone defies easy categorization; it remains a unique linguistic specimen.' Further '. . . it clearly challenges the presupposition that one can unambiguously determine the linguistic antecedents of every language.' However he draws back from the unequivocal conclusion that Ma'a is a true hybrid, saying that 'this would depend on one's definition'. However, he never states what the definitional difference is to which he alludes.

I believe that the problem with the hypothesis of Ma'a as a mixed language is that those who espouse it have treated it as a synchronic problem. There are two elements at present in Ma'a. One is Cushitic and the other is Bantu.

Hence Ma'a is a mixed language. Just so, if one says that English is related to German it *seems* to be a statement about the present time and hence synchronic but, of course, one is positing a sequence of events through time: namely that an original unity, the Germanic languages, differentiated into separate languages over the course of time by geographical separation and differential changes.

The question is then whether, as is the case with the Romance elements in English, they arose from a contact phenomenon of later date. Now, as one goes back in time, the French and other Romance elements are sloughed off and older English is closer to older Germanic than the present language. Thus Anglo-Saxon and Old High Germanic are certainly more similar to each other than modern English and modern German. The historical linguist simply extends this trajectory to the prehistoric period and posits an original unity Proto-Germanic as the source of modern English, modern German and, of course, other Germanic languages.

In the case of Ma'a the Bantu elements can, it appears clearly to me as it has to others, be considered later intrusions into an originally Cushitic language. Note that in the instance of English and German a fundamental assumption is one of genetic continuity. Modern English is the result of a changed transmission through time of Anglo-Saxon into English, and German that of Old High German into modern German.

There is no evidence of interrupted transmission of Ma'a from an earlier period into the present, nor has anyone claimed that there is. The case of Bantu elements in Ma'a is then really like that of Romance elements in English. The only difference is that grammatical morphemes are involved to a rather startling degree while vocabulary has been far less affected than in the case of Romance influence in English.

We have not yet considered a number of further indications that Bantu elements in Ma'a are relatively late and intrusive. In regard to the class prefixes, Meinhof already remarked (1906) that in recordings by different observers the same noun sometimes appears with and without a prefix. Thus alongside of *mu-asa* 'fire' (class 3, a singular class) we find *asa*. This is particularly true in the singulars, so that the plurals, in accordance with the marked nature of this category are more likely to have overt expression, cf. English *hand, hand-s*. Later Tucker and Bryan (1957) noted that some nouns have no prefixes and that in others the prefix is sometimes omitted, especially in the singular.

That this use of noun prefixes was to some extent, as it were, an affectation used when speaking to the more numerous Bantu speakers or to other outsiders was reported by Meinhof (1906: 403) as a strong impression of his

resident German missionary acquaintance Röhl. Röhl could not escape the suspicion that Ma'a speakers only used Bantu prefixes freely in talking to others but not among themselves in order as he put it, to 'make the matter plausible'. This impression of Röhl becomes a virtual certainty through the report by Tucker and Bryan (1974: 192) that Bryan remembered that one informant gave the word for 'river' as *mu-haraza* when it was elicited as part of a word list and that the same informant used *haraza* in conversation. It is noteworthy that Meinhof (1906) had earlier elicited *haraza* as part of a word list. Thus the historical direction of movement was towards the adoption of the Bantu affix system is further attested by Whiteley (1960: 96) who noted that it was certainly true that the affix system is better established among the younger generation than the older generation.

To my knowledge it is unexampled in Bantu languages that nouns can have variants with and without the noun class prefixes and is a further indication of their recency [in Ma'a]. Moreover their source is here well known. Two historically based and genuine Bantu languages Pare and Shambala, closely related but distinct languages, have been the successive sources of Bantu elements in Ma'a.

There is a widespread drift in Niger-Congo languages towards the loss of noun classes which reaches its climax in languages like Yoruba and Ibo. However, such loss and acquisition by borrowing are sharply differing processes. In the former, thousands of years are required. The chief mechanisms are phonological changes of the class markers leading to merger and the gradual loss of certain classes through analogical shifts into other classes which thereby become more productive. These have acquired new semantic correlates via ambiguous meanings. Thus a word meaning 'fire, wood', a common semantic equation in Africa as elsewhere in the world, can come to be used with tree names and thus acquire a new class meaning while other classes become depleted and ultimately lost. A similar phenomenon can be noted also in the verb system. Copland (1934: 243), the earliest systematic observer, saw in the final *-a* of *va-kwekw-a* 'they fought' an attempt to Bantuize a verbal root by the *-a* ending of the Bantu infinitive and most tense markers as against the use of various vowels in Cushitic. He also noted (ibid.) the variants *gululu* and *gulula* 'to run away'.

In addition to the independent personal pronouns, possessives, and demonstratives, there appears to be a single Cushitic survival in morphology, a suffixed plural *-no* in *ma-lare-no* 'clouds' (Green 1963: 289). Here *ma-* is the prefix of Bantu class 6. The suffix *-no* may be compared to the widespread plural suffixes with *-n* found in both nouns and adjectives in Cushitic, and frequently elsewhere in Afroasiatic.

Finally, a most significant piece of evidence is furnished by the numerals. The set 1–6 are non-Bantu and of these 'one' and 'two' can be identified as Cushitic. Likewise 'ten' and its multiples are non-Bantu. It is a well-attested phenomenon that when a language borrows numerals from a foreign source, it is the lowest and most unmarked numerals which will be of indigenous origin as well as the numeral for the base of the system, most commonly ten. This marking hierarchy in numerals is described in Greenberg (1978: 288–9) and includes its exemplification in borrowed numerals.

We can see then that the most plausible scenario is that Ma'a is a Southern Cushitic language which underwent considerable and, in certain respects, remarkable changes induced by contact with Bantu languages. That this is the natural way of accounting for the present situation is shown by a remarkable statement in Thomason and Kaufman, the strongest proponents of the mixed language theory when they state (1983: 208): 'We can establish that Ma'a arose from a Cushitic language'. Probably the best statement remains that of Welmers (1973: 9) when he states that Ma'a appears to have had a continuous genetic history which is Cushitic with remarkably extensive but entirely explicable foreign influence.

The case of Mednyj Aleut and its genetic affiliation is even clearer and may be dealt with more briefly for reasons discussed in the opening section of this paper. Our knowledge of this dialect of Aleut is derived from two discussions by Menovshchikov (1968, 1969). As basic linguistic background, it should be noted that the relationship of Aleut as a whole to Eskimo in an Eskimo-Aleut grouping is uncontested. The main body of Aleut speakers, in the Aleutian chain, politically part of Alaska, are divided into three dialect groups, Western, Central, and Eastern. Their differences are not great. In 1826 the Russians, who at that time controlled Alaska and the Aleutians, moved a number of Aleutians to the Commander Islands, which are close to Kamchatka. One group, from Attu, in the extreme west end of the Aleutians, was moved to Mednyj (Copper) Island. The others were transported from Atka, whose dialect is the sole member of Central Aleut, to Bering Island.

The speakers on Bering Island have continued to speak their dialect of Aleut with some relatively minor vocabulary borrowings from Russian and no grammatic influence at all. The story on Copper Island is, however, different presumably because of the numerous marriages between Russian men and Aleut women. Here also, except for obvious Russian loans, the vocabulary is Aleut. In the verb, however, in the indicative tenses, the present has the Russian endings *-ju/-is/-it/-im/-iti/-jut* and the past has the Russian form in *-l* derived, of course, from a Slavic participle and agreeing like Russian with the subject in number and gender but not in person. To indicate the persons,

in this tense Mednyj Aleut uses the Russian nominative pronouns, but nowhere else. Menovshchikov here mentions no Aleut feminine or neuter forms but he does cite a plural form *aguli* 'they built'. Moreover, in the listing of categories of the verb he includes number, but not gender. Aleut does not have the category of gender. It is likely that the verb inflections cited above all reflect the Russian first conjugation, which is probably the most frequent one. However, Menovshchikov also cites one form in the third person plural in -*at* so this point remains somewhat uncertain. The Mednyj future is formed from *budu* followed by the Aleut verb stem and then the Russian infinitive -*it'*. Thus Mednyj for 'I will sit' has *budu unguchit'*. The negative of the verb does not use the Aleut suffix -*laka* on the verb stem followed by the Aleut personal subject marker, but *ni-* as a negative prefix followed by the Aleut stem and then the person markers -*ju/-is,* etc. Thus 'we do not work' is *ni-aba-im*. Finally, both the positive and negative imperatives follow the Russian model e.g. *aba-j* 'work!', *ni-aba-j* 'do not work'. No information on the imperative plural is given. However all of the subordinate forms of the verb based on participles are Aleut in formation.

Thomason and Kaufman on the basis of the foregoing facts consider Mednyj Aleut, as they do Ma'a, to be a mixed language. In the case of Mednyj Aleut, they use the strange expression '(ex-?)dialect of Aleut' (Thomason and Kaufman 1988: 105). However, as would seem to be obvious, a historical fact cannot be annulled. A language which is Germanic cannot 'become' Romance. Its speakers, however, may undergo a language shift and begin speaking a Romance language, abandoning a Germanic one which they formerly spoke.

Perhaps we can get to the heart of the matter if we consider the two quotations with which Thomason and Kaufman begin their book. One is from Max Müller and states that there are no mixed languages. The other is from Hugo Schuchardt and states that all languages are mixed. However, this 'contradiction' rests on an ambiguity in the use of the word 'mixed'. In Schuchardt's sense all languages are mixed because they contain elements of diverse historical origin. In regard to this Schuchardt is correct. However if we use the word 'mixed' in this sense English is mixed language. On a dictionary count there are surely more words of present-day English of Romance than of Germanic origin. For that matter, English is then .003 per cent Aztec because of *cacao* (Aztec *cacaoatl*), *atlatl* (spear thrower), *chocolate* (*chocolatl*) and a handful of others.

Of course, no Indo-Europeanist will agree that English is a mixed language and Thomason and Kaufman in deference to the unanimous opinion that English is Germanic shy away from the notion that English is a mixed language. Their two prize examples, however, are in principle no different from

English. There has been no break in linguistic continuity in either case and that the earlier elements of Ma'a and Mednyj are Cushitic and Aleut respectively is freely admitted by Thomason and Kaufman. The only thing peculiar about Ma'a and Mednyj Aleut is that the latter's borrowed elements are predominantly grammatical. In the case of Ma'a it is strange that Thomason and Kaufman failed to notice the numerous hints of the secondary and still largely unassimilated nature of the Bantu elements as discussed earlier in this paper.

Note that a Bantu language which underwent Cushitic influence and a Cushitic language which underwent Bantu influence would look very different. Perhaps this difference is shown most starkly in the numerals. If the former were the case, the lowest numerals would be Bantu and the higher ones Cushitic in provenience, whereas the opposite is the case.

It is indeed hard to imagine how a truly mixed language in Müller's (which is the usual) sense, could arise by a natural process. Suppose someone had a dictionary and grammar of two quite distinct languages. He or she could then take alternate words and grammatical morphemes first from one and then the other. This would truly be a mixed language but, of course, not arising by any natural process.

The closest I can come is the case of Russenorsk which seems to have roughly equal amounts of Russian and Norwegian. It is, however, a pidgin with very limited vocabulary and syntax, formerly used by traders and not anyone's first language. If it had become creolized and the first language of a population it might be considered a truly mixed language. However, as it stands, no authentic example of a mixed language exists. The normal and apparently exclusive way in which new languages arise is by transmission of an earlier language which has undergone sufficient change not to be mutually intelligible with different forms of the same earlier language transmitted in other areas by other populations. This is precisely the genetic model. *Omnis lingua ex lingua.*

References

Copland, B. D. 1934. 'A note on the origin of the Mbugu, with a text', *Zeitschrift für eingeborenen Sprachen* 24: 241–4.

Dolgopolsky, Aaron. 1973. *Sravitel'no-istoricheskaja fonetika kushitskix jazykov.* Moscow: Nauka.

Ehret, Christopher. 1976. 'Cushitic prehistory', in M. Lionel Bender (ed.), *The Non-Semitic Languages of Ethiopia.* East Lansing: African Studies Center, Michigan State University, 85–96.

—— 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary.* (Kölner Beiträge zur Afrikanistik, vol. 5.) Berlin: Reimer.

Elderkin, E. Derek. 1976. 'Southern Cushitic', in M. Lionel Bender, *The Non-Semitic Languages of Ethiopia.* East Lansing: African Studies Center, Michigan State University, 278–97.

Goodman, Morris. 1971. 'The strange case of Mbugu', in Dell Hymes (ed.), *Pidginization and Creolization of Languages.* Cambridge: Cambridge University Press, 243–54.

Green, E. C. 1963. 'The Wambugu of Usambara (With Note on Kimbugu)', *Tanganyika Notes and Records* 61: 175–89.

Greenberg, Joseph H. 1963. *The Languages of Africa.* The Hague: Mouton.

—— 1978. 'Generalizations about numeral systems', in Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of Human Language, vol. 3: Word Structure.* Stanford: Stanford University Press, 249–96.

Meinhof, Carl, 1906. 'Linguistische Studien in Ostafrika 10: Mbugu', *Mitteilungen des Seminars Orientalische Sprachen* 9: 293–323.

Menovshchikov, G. A. 1968. 'Aleutsky Jazyk', *Jazyki Narodov SSSR* 5: 386–406.

—— 1969. 'O nekotoryx social'nyx aspektax èvoljucii jazyka', in A. V. Desnitskaja, V. M. Zhirmunskiji and L. S. Koutun (eds.), *Voprosy social'noj lingvistiki*, 110–134. Leningrad: Nauka.

Shaw, A. D. 1885. *A Pocket Vocabulary of the Ki-Swahili, Ki-Nyika and Other Languages.* London: Society for Promoting Christian Knowledge.

Sim, R. J. 1994. 'Cushitic', in R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics.* Oxford: Pergamon Press, 798–9.

Thomason, Sarah G. 1983. 'Genetic relationship and the case of Ma'a (Mbugu)', *Studies in African Linguistics* 14: 195–271.

Thomason, Sarah G. and Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley: University of California Press.

Tucker, A. N. and Bryan, M. A. 1957. *Linguistic Survey of the Northern Bantu Borderland*, vol. IV, London: Oxford University Press, 72–4.

—— 1974. 'The Mbugu anomaly', *Bulletin of the School of Oriental and African Languages* 37: 188–207.

Welmers, William E. 1973. *African Language Structures.* Berkeley: University of California Press.

Whiteley, W. H. 1960. 'Linguistic hybrids', *African Studies* 19: 95–7.

This page intentionally left blank

Language and archaeology: Review of *Archaeology and Language: the Puzzle of Indo-European Origins* by Colin Renfrew and *A Guide to the World's Languages, vol.* 1: *Classification*, by Merritt Ruhlen*

As Renfrew states at the beginning of his work (p. 1), archaeologists of late have had little to say about language origins. With regard to Indo-European, the focus of his book, it is indeed remarkable that the standard textbook on European archaeology (Champion *et al.* 1984) does not even contain the word Indo-European in the index. It is Renfrew's view that with recent advances in processual archaeology and comparative linguistics the time is ripe for a new attempt to investigate Indo-European origins. The whole enterprise had, of course, fallen into disrepute because of the excesses of the Aryan theory in the 1920s and 1930s, which was the immediate background of Boas's well-known pronouncements regarding the independence of race, language, and culture. However, to me at least, it is clear that Boas could not, as an empirical scientist, object to researches into such connections. Moreover, the conclusion that one group is superior and has the inevitable destiny of conquering other groups has no logical connection with the endeavor to find correlations between linguistic and nonlinguistic evidence.

In point of fact, taking a broader world perspective as well as one deriving from Indo-European studies, the situation is not quite as stark as would appear from Renfrew's initial statement and similar ones elsewhere in his book. Any widespread linguistic stock whose unity is not to be doubted—e.g. Austronesian or Bantu—practically invites the attempt to employ archaeological

---

data to account for the present distributions by dynamic processes in space and time. Both these examples are cited by Renfrew as support for his own methodology (pp. 277–84), a matter to which I will return later in the discussion.

Indo-European has been something of a special case. Interest in the topic, prominent already in the nineteenth century, has never really ceased but it has been pursued throughout more actively by linguists than by archaeologists.

Moreover, apparently even while Renfrew was preparing his volume, two contributions originally published in Russian in 1981 and 1982, and then translated into English in *Soviet Anthropology and Archaeology* in 1984, presented new theories of Indo-European origins. The authors of these theories were once more linguists: Gamkrelidze and Ivanov, who propose eastern Anatolia as the homeland, and Diakonov, who, in reaction to their theory, opts rather for the Balkans.

In addition to the interest in the problem of linguistic origins evidenced by the appearance of three theories (including Renfrew's), to challenge what had begun to attain virtually an orthodox position regarding Indo-European origins, (namely, the identification of Indo-European spread with that of the Kurgan culture originating north of the Black Sea) there have been other indications of a reawakening of interest in the possibility of correlating archaeological and linguistic evidences. One such indication is the recent joint paper of Greenberg, Turner, and Zegura (1987), linking language, population genetics, and dentition in relating to the settlement of the New World, a paper reviewed in detail by Frederick West (1987). A second evidence of this renewed interest is an important paper by Cavalli-Sforza *et al.* (1988) which shows a surprising degree of congruence between the genetic structure of populations and language groupings on a worldwide scale and incidentally provides strongly confirmatory evidence for the division of New World population into the three groups proposed in Greenberg *et al.* (1987).

The reason for the present interest in Ruhlen's book is that the genetic classification of language is the foundation on which all other historical inferences from language are based and, in fact, the 1988 Cavalli-Sforza paper uses Ruhlen's work as its basic source for language divisions. It seems therefore useful to bring it to the attention of archaeologists who, in general, would not be expected to follow the linguistic literature.

Ruhlen's book is worldwide in its scope. It contains an initial discussion of the methodology of linguistic classification followed by chapters covering each major world area and reviewing the history and current states of classification of each major family; a basic bibliography is also included for each.

The book is plentifully illustrated with maps of linguistic stocks. Ruhlen does not hesitate to draw conclusions himself but all the major viewpoints are presented. An appendix contains a detailed classification of all areas of the world down to the level of individual languages. It is clearly the most up-to-date and complete source on the classification of the world's languages.

With regard to Renfrew's book, readers of this journal will doubtless be acquainted with his basic thesis since it received a double review in this journal [*The Quarterly Review of Archaeology*, 9:1-Ed.]. The reason for my discussion is that Renfrew's argument depends in crucial ways on his handling of linguistic data as a source of historical inferences, and neither of the two previous reviewers is a linguist.

To begin with, it may be said in extenuation of Renfrew, who has courageously set forth views which he knows both archaeologists and linguists will disagree with, that it is not easy for an outsider to comprehend clearly a field basically different from his own. This is certainly true of historical linguistics. The outsider with his own pet theory, naturally enough, looks for that which will support his own point of view, but he will necessarily lack the first-hand knowledge of an alien field, including the languages themselves, which will allow him to make informed judgements. Linguists, like other scientists, sometimes say foolish things and there is no certified psychiatrist to say which views are rational and which are not. I have felt the same way in regard to archaeology. In regard to the implications of my Amerind classification, how am I to judge between the conflicting claims of the majority who claim that the oldest settlement in the area is not much more than 12,000 years and those who claim that it is far earlier?

Any discussion of Indo-European origins is just a single application, however important, of a set of general principles that have evolved regarding the methodology of such inferences. By now we know that ideas often expressed in the nineteenth century about 'primitive languages'—based primarily on then-popular evolutionary typology of an isolating, followed by an agglutinative, and finally an inflective stage (with Indo-European and Semitic as the only true embodiments of this last and highest form) simply evaporate in the face of facts. Basically all languages share the same fundamental structural characteristics and, what is most relevant here, go through basically similar processes of change. It is an important corollary that since the processes are basically similar, the methods of utilizing linguistic data for historical inferences in different parts of the world must be also essentially the same.

In fact, Sapir's famous essay on time depth in American aboriginal culture (1916), applied with two major more recent additions—one relating to

glottochronology and the other to areal distributions within the protolanguage—even now could be recommended as probably the best introduction to this whole area of investigation.

What are the basic questions to be answered and how does Renfrew answer them in relation to the Indo-European data? The crucial questions are *what, where*, and *when*. What was Proto-Indo-European culture like, where was its homeland, and when was it spoken? For all of these questions, methods have been developed which depend crucially on the genetic classification of languages in terms basically of family tree, and this paradigm is also the foundation of comparative-historical linguistics as it developed in the nineteenth century.

Put very briefly, the *what* of cultural content is based on reconstruction of the cultural vocabulary of the protolanguage, in which occurrence of terms that, on linguistic grounds, give evidence of cognacy in two or more genetic subgroups is considered adequate grounds for assigning it to the proto-vocabulary (the principle of two independent witnesses, as apparently first explicitly stated by Brugmann, but actually earlier). The extent to which this method needs some modification on the basis of areal considerations (related to wave theory) will be considered later.

The *where* involves the principle of 'center of gravity', first (to my knowledge) enunciated by Sapir, a method brilliantly illustrated by his inference concerning the northern origin of the Navaho-Apache branch of Athapaskan. In this method the evidence of linguistic groupings of the same genetic rank is accorded equal weight regardless of the geographical spread of those groupings, population size, or internal differentiation.

Finally, the *when* is answered by glottochronological methods in which percentages of cognate vocabulary between various genetically determined branches are evaluated on the basis of a decay function derived from historically attestable and datable cases. As we can see, these are all based on family tree (cladistic) classification.

All the methods mentioned have their difficulties and provide answers of varying degrees of precision and certainty depending on the circumstances in specific cases. Among these circumstances are the clarity of subgrouping within the linguistic stock and the extent of internal ramification, i.e. the number of branches.

By now Renfrew's basic theory will be familiar to the reader: the spread of Indo-European is coincident with that of agriculture through Europe from Greece, ultimately deriving from western Anatolia, about 6000 BC and covering all of Europe within two millennia. The mechanism involved is the wave of advance model of Cavalli-Sforza and Ammerman.

To contrast his theory with what is probably the most widely accepted, that of Gimbutas, for Renfrew Proto-Indo-European (PIE) must be at least two thousand years earlier than usually thought, to originate in Greece rather than north of the Black Sea, and to be agricultural rather than pastoral. In support of it, as we have seen, he claims to be integrating modern processual archaeology with recent advances in linguistic theory. As one who is not an archaeologist I hesitate to express an opinion on processual archaeology. However, it appears that his version of it is one which seems to deprecate migrations altogether as a mechanism of culture spread in favor of diffusion of artifacts by trade and peer interaction. It would seem to me that both migration and diffusion* occur and possess linguistic correlates. The former results in language differentiations according to the model of genetic split, while the latter would more likely show up as borrowing for words denoting the objects themselves. Both processes certainly occur but a whole school of archaeology should presumably not be founded on the assumption that one process is at work and the other is not, or, at least, is operating only to a very minor extent. Renfrew is certainly correct in noting that the past equation, *archaeological culture = linguistic stock* is often too simple and leads in some instances to the positing of wholesale migrations backed up by very little evidence. Still, there are many historically attested migrations and linguistic evidence often leads us to posit them with great confidence, as in the Polynesian example discussed by Renfrew.

Returning to the linguistic question, just what are the recent advances which can lead to a new and more convincing answer to the problem of Indo-European origins? The most important appears to be the linguistic wave theory which Renfrew opposes to the commonly accepted genetic model, which, as we have seen, figures in a fundamental way in the usual reasoning from language to cultural history. Renfrew is perhaps here subconsciously influenced by the use of the term 'wave of advance' in regard to the spread of agriculture. However, wave theory in linguistics as Renfrew indicates is hardly a recent development. It is usually considered to date from the fundamental monograph of Schmidt (1872) which is, in fact, cited by Renfrew. In regard to Renfrew's version of intellectual history, however, it is definitely an error to assume that the notion of the family tree, whose resemblance to evolutionary theory in biology is obvious, is merely an imitation by linguistics of biology (p. 102). In fact it was accepted earlier in linguistics, and it was later that both biologists and linguists became aware of the parallel. So Max Mueller, a well-known nineteenth-century

---

* The two terms are reversed in the original-Ed.

linguist, could say, 'In linguistics I was a Darwinian before Darwin' (Greenberg 1959).

Neither the family tree nor the wave theory in its extreme form can account for all the facts. I am talking about those facts whose explanation logically precedes the *what, where*, and *when* mentioned earlier. This is the *how*. Unless we understand how linguistic divergence and convergence proceed we cannot develop systematic ways of answering the other questions.

There are three basic phenomena to be accounted for. In the world we see many different, distinct, mutually unintelligible languages. At the same time we see that their differences are not random. German and English, though distinct languages, are obviously much more like each other than either is like French. It was the attempt to explain these facts in dynamic terms that represented the breakthrough that was the birth of modern linguistics in the nineteenth century.

English resembles German more than it does French because there was once a language, Proto-Germanic, which developed into English, German, and a whole group of other Germanic languages over time, while French did not share in the common ancestry of Germanic—although, at a more distant remove, it does do so as a descendant of Latin, a member of the Italic branch of Indo-European. This latter, along with a whole series of other languages, e.g. Proto-Slavic, shares a common ancestor with Germanic, namely, Proto-Indo-European. The third question is that of variation within individual languages, namely, geographical and social dialects. How do they arise and how are they related to the process of differentiation just sketched in family tree terms? If languages always had clearly demarcated dialects as shown by large numbers of coinciding isoglosses, there would be no problem. The dialect of today is the language of tomorrow.

Here is where the wave theory comes in. It really arose from dialect geography in the 1860s. It was found that innovations do not always spread uniformly over identical areas producing neat divisions. Renfrew (p. 103) cites Bloomfield as saying that the method of comparative reconstruction assumes that the parent community was completely uniform in language and that the parent community split suddenly and sharply into two or more daughter communities which lost all contact with each other. However, no comparative linguist believes this, and Bloomfield was certainly not expressing this as a necessary condition for comparative linguistics and linguistic reconstruction to be feasible.

Now let us consider the wave model, also in an extreme form. Let us suppose we have a pond representing the original homogeneous speech community. We throw in pebbles and rocks of various sizes. Each represents

a linguistic innovation. Where each pebble falls is completely random, the spread of the resulting wave is concentric, and the extent of its spread depends on the size of the stone. The results will, of course, be gradual transitions so that dialects will resemble each other in proportion to their distances. There will be no branching at all.

This picture, which is indeed close to Schmidt's original idea, is even farther from reality than the extreme family tree branching model discussed above. The medium must be completely homogeneous. For example, if there is a mountainous region the spread must take no notice of it. Nor is it possible for any group to migrate, thus cutting off communications with the remaining speakers and preventing the transmission of innovations.

Schmidt himself needed to explain why, if the wave theory really worked, there were no transitions between recognized subfamilies. The distinction, for example, between a Celtic and a Germanic language is completely clear; there are no transitions. In fact, Schmidt operated with distinct branches such as Slavic and Baltic and his discussion was simply oriented towards higher groups such as Balto-Slavic within Indo-European. He sought to explain the absence of transitional dialects by assuming that they once existed but were displaced by the dialects of various distinct centers which had become dominant, presumably for sociolinguistic reasons. However, no such transitional dialects have ever been discovered. When we find a new Indo-European language, as since Schmidt's time has been the case with Hittite and Tokharian, they turn out to be different, often drastically so, from the languages we already know.

However, wave theory did make one important contribution. It modified Brugmann's law of two independent witnesses. It was recognized that the breakup of the original stock was not a sudden process. There were even very early, definite lines of cleavage, 'protodialects', but contact was often maintained and early borrowings, largely indistinguishable by linguistic analyses, were inherited from an early period before the cleavage had become pronounced. Moreover, where dialects were still very similar, sound changes might for a time at least be propagated across such boundaries.

It is only in Indo-European that studies regarding this problem have been actively pursued. Careful studies do allow us to reconstruct at least approximately the relative geographical positions of the dialects of the protostock which are ancestral to the later branches, e.g. Porzig (1954).

Note that wave theory in its pure form makes no real distinction between resemblances resulting from borrowing in the usual sense (between distinct languages) and the spread of innovations from dialect to dialect within a particular speech community. In this extreme form it is, once again, untenable.

If a word passes from present-day English into French, which are not mutually intelligible forms of speech, it is surely a borrowing.

It would seem that the preceding discussion is a digression from our consideration of Renfrew. It is not, however. Renfrew seems to believe that borrowing was not recognized before the wave model and that it can only be understood as a consequence of it. Neither of these two assumptions can be maintained.

Renfrew (p. 104) gives a list of early loanwords from Latin into West Germanic, e.g. Latin *vinum* 'wine' as the source of the Old English and Old High German *win* 'wine'. But these and similar borrowings, known well before wave theory, were taken from a Latin dominant in political and intellectual life into the Germanic languages long after Latin had itself differentiated from other Italic dialects such as Oscan, Umbrian, and Falisean. If it were a wave phenomenon we would have to imagine imperfectly separated Germanic and Italic neighboring dialect areas and forms spreading from Italic dialects to adjacent Germanic dialects. But Latin, by the time it was Latin, was surely not in such a geographical relation to Germanic. In fact it is only with a cladistic model, modified areally in the way noted earlier, that the distinction between internal dialect contacts and true borrowing can be made.

Incidentally, borrowing often between a protolanguage and some other genetically diverse language is a second major source of information regarding geographical locations. Thus, the numerous loanwords from what is very close to, if not identical with, Proto-Germanic as we understand it, in the Balto-Finnic languages indicates that they must have at one time inhabited geographically contiguous areas. The absence of loanwords from Balto-Finnic into Germanic suggests a culturally dominant position for the Germanic speakers and this ought to show up in some way in the archaeological record.

Renfrew's obvious aversion to the genetic (cladistic) model derives from the fact that it gives results that disagree with his basic hypothesis. Very strange is Renfrew's 'flirtation' with Trubetskoy's 1939 paper, which is mentioned repeatedly. Renfrew realizes that it must be rejected and in fact one cannot name a single Indo-Europeanist who accepted it. It was convincingly refuted, especially by Thieme (1953). Briefly stated, Trubetskoy's theory held that there was no Proto-Indo-European speech community. A variety of unrelated languages were spoken and by linguistic contact they became so similar over time that we can now speak of an Indo-European group of languages.

This cannot be so for numerous and weighty reasons. Just one of these is the following. If the present Indo-European languages are a product of convergence they should be more and more different as one goes back in time. But contemporary English and German are far more different from

each other than are Old English and Old High German, and a similar state of affairs holds wherever we look. It is thus a simple projection that if we had still earlier records they would in fact represent variants of the same language just as the present Romance languages differentiated from late Vulgar Latin and, consequently, Old French, Old Provençal, and Old Spanish were far more similar to each other than are the present-day Romance languages.

Oddly enough, Renfrew is attracted to Trubetskoy's theory just as he is to wave theory, but these are exact opposites. Schmidt started with a primitive unity which became more and more diverse by changes spreading in a wave-like fashion, while for Trubetskoy the starting point was a primitive diversity which by multifold contacts produced the resemblances found in the present Indo-European languages. If this view were to be adopted, the problem to which Renfrew is proposing an answer would simply disappear. There would be no Proto-Indo-European speech community to which his hypotheses of place, time, and cultural characteristics could apply.

In regard to the general problem discussed here, as to how different languages arise, Renfrew in fact has proposed no clear mechanism. How, for example, in spreading across Europe such linguistically highly diverse groups as the Celts, the Germans, and the Slavs originate is not clear and is hidden in obscurity by his use of such terms as Celtic 'ethnogenesis'.

As was noted earlier, the usual types of inferences, all based ultimately on a modified cladistic model, militate against Renfrew's thesis. Two of these, from protolexicon and glottochronology, he seeks to discount often by invalid arguments. It is true, for example, that protolexicon must be used with care. There are two major problems. One is the possibility of parallel semantic changes producing a specious protomeaning. For example, one might suppose that the language ancestral to the Kanuri and Teda-Daza of the Central Sudan (a subgroup of Nilo-Saharan) had a word for 'to write' and hence the people had writing, but they are parallel changes from a word originally meaning 'to scratch'. Still a careful consideration of a large number of other words such as 'read', 'pen', etc., all from Arabic, would be sufficient to refute this.

Renfrew, however, wishes to throw out all protovocabulary evidence. 'If there are no common words today among the Indo-European languages [for wheat, barley, etc.] then other explanations must be found' (p. 84), but none are. This contrasts with the fairly numerous obviously cognate words for domestic animals. There is here, as in the case of Central Saharan words relating to writing, safety in numbers.

As has been often pointed out, such distinctively Greek flora as the cypress and the olive tree have no Indo-European etymology and their internal linguistic structure suggests that they are borrowings.

With regard to glottochronology, since it results in a date much later than that posited by Renfrew, it is simply rejected *in toto*. While its weaknesses are well known, the assumption that the retention rate of basic vocabulary, which Renfrew does not distinguish clearly from vocabulary as a whole, has almost limitless variability is clearly untenable.

In regard to place of origin, Renfrew seems unaware of the basic principle described earlier by Sapir as the center of gravity method, which corresponds to Vavilov's principle in the study of domesticated plants. Since it takes time for variation to develop we expect the greatest genetic diversity near the point of origin. In fact, both of Renfrew's examples (pp. 277–84), Polynesian and Bantu, illustrate this perfectly if Renfrew had carefully examined the linguistic family trees which he presents.

There is just one point at which Renfrew shows some awareness of the general principle that diversification takes time and that highly similar forms of speech must have diverged over a very short period. This relates to the problem of the existence of the Indo-Aryan languages to the *east* of the hypothetical western Anatolian homeland. For these he proposes two hypotheses. The first (hypothesis A) involves a gradual spread of Indo-European farmers eastward in accordance with the Cavalli-Sforza and Ammerman model. His B model would involve a migration from the north of the Caucasus. He naturally prefers A to B but, in addition to the absence of Indo-Aryan loanwords in the languages of the early Near East, which he notes, he for once admits that the high degree of similarity between the earliest Iranian (the languages of the *Avesta*) and the Vedas is an obstacle. Just how similar they are he apparently does not realize. Even Schmidt, the wave theorist, admitted an Indo-Aryan branch of Indo-European. Whole passages of the *Avesta* could be turned into Vedic Sanskrit by the application of a few rules of phonetic replacement.

Renfrew considers one further development in recent historical linguistics, namely, word order typology (p. 111). Linguistic typological classification is a kind of structural classification, logically independent of genetic classification, and while it can make a contribution under certain circumstances to historical problems when treated carefully, it is not, in fact, utilized by Renfrew and is, I gather, merely mentioned to show that there are a lot of recent developments in historical linguistics which lead to a reassessment of the traditional evidence.

In fairness to Renfrew, one should mention two considerations: one is in regard to chronology. As noted earlier, glottochronology *is* a treacherous guide. I suspect strongly that because of the greater retention rate of the more stable words, (the so-called dregs phenomenon), long dates are probably

underestimated. Moreover, again and again, as in the case of Mycenean Greek and Hittite, the dates of Indo-European speakers in particular areas outside of their putative homeland have been underestimated.

The other observation is that a word should, I think, be put in regarding the merits of Renfrew's sixth chapter in which he has sought to classify and analyze the distinct mechanisms of language spread and to provide a suitable terminology. I do not recall any similar previous attempt although it may well exist. I have seen recent references to this chapter and I consider it a thought-provoking treatment which also provides a useful terminology.

It should be noted that the present reviewer does not endorse any of the existing answers to a very difficult question. One of them may well be right. It may also be true that we will never be utterly sure that the correct theory is indeed valid.

References

Cavalli-Sforza, Luca L., Piazza, Alberto, Menozzi, Paolo, and Mountain, Joanna. 1988. 'Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data', *Proceedings of the National Academy of Sciences* 88: 6002–6.

Champion, Timothy, Gamble, Clive, Shennan, Stephen, and Whittle, Alasdair. 1984. *Prehistoric Europe.* London: Academic Press.

Greenberg, Joseph H. 1959. 'Language and evolution', in Betty J. Meggers (ed.), *Evolution and Anthropology: A Centennial Appraisal.* Washington DC: Anthropological Society of Washington, 61–75.

Greenberg, Joseph H., Turner, Christy G., II, and Zegura, Steven. 1987. 'The settlement of the Americas', *Current Anthropology* 25: 477–97.

Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets.* Heidelberg: Winter.

Sapir, Edward. 1916. 'Time perspective in aboriginal American culture: a study in method', *Canada Department of Mines Geological Survey Memoir* 90, *Anthropological Series* no. 13. Ottawa: Government Printing Office.

Schmidt, Johannes. 1872. *Die Verwandschaftsverhältnisse der indogermanischen Sprachen.* Weimer: Böhlau.

Thieme, Paul. 1953. 'Die Heimat der idg. Grundsprache', *Akademie der Wissenschaften und der Literatur, Abhandlungen*, no. 11. Wiesbaden: Steiner.

Trubetskoy, Nikolai S. 1939. 'Gedanken über das Indogermanenproblem', *Acta Linguistica* I: 81–9.

West, Frederick H. 1987. 'Migrationism and New World origins', *Quarterly Review of Archaeology* 8/1: 11–14.

This page intentionally left blank

20

Beringia and New World origins: the linguistic evidence*

What can the languages of Native Americans tell us about how the New World was settled? In seeking an answer to this question, I will as far as possible rely on linguistic evidence only and, not being an archaeologist or geneticist by profession, and therefore not capable of independent evaluation in these other fields, I will refrain as far as possible from using hypotheses which derive from them. Conclusions which derive jointly from a number of scholarly disciplines are reliable and convincing only insofar as they are based on independent evidence from each field.

Two preliminary observations are in order. One is that in considering linguistic evidence, there are first the languages themselves, that is, their vocabulary, sound systems, morphological and syntactic structures, and their similarities and differences, which lead to hypotheses of common origin and of language contacts over time. But one major apparently external factor does inevitably play a role: the present and presumed former geographical distribution of languages and their speakers.

Moreover, the geography plays a similar role as a formally external, yet basic, intrinsic factor in considering archaeological and genetic evidence also. It forms, as it were, the universal background in all of these fields, yet what is presumed to have happened within these spatial configurations is based on different and independent evidence. If the results converge to produce reasonable and harmonious conclusions, we are naturally gratified. The other preliminary observation is one that presumably just about every linguist and nonlinguist at all acquainted with historical linguistics takes for granted, yet it deserves some consideration. This is the role of the so-called genetic classification of languages. In orthodox historical linguistics this is constituted by a taxonomy of inclusion basically similar to that of biological classification. Just as in biology, this basic family-tree approach has its problems. For

example, at the lowest taxonomic level, difficulty in distinguishing separate species from varieties of the same species has its parallel in distinguishing dialect from language. This is because we are dealing with a dynamic process, speciation and language formation. Ultimately we have what are indubitably separate languages and separate species. Moreover, as has long been seen, it is the assumption that this dynamic process of the development of ever more internal variation and ultimate separation takes place in a similar fashion in the present and past that leads to an evolutionary interpretation in biology on the one hand, and in genetic linguistics on the other.

Not all historical deductions from language are based directly on the genetic hierarchy of languages, e.g. conclusions regarding language contact and those regarding the cultural implications of protovocabularies, but they all require a genetic classification as a prerequisite. These may be illustrated from examples. When a group of languages presently or formerly adjacent have had intrinsic contacts with each other we talk of areal factors. One well-known example is that of the Balkans. Not only borrowed words, but typological characteristics such as the existence of a suffixed definite article and a future tense formed from a verb meaning 'to wish' are among the Balkan linguistic characteristics (Sandfeld 1968). These latter items are often called loan translations or calques.

In regard to the suffixed article, Bulgarian, Romanian, and Albanian all share the structural similarity to which we have alluded. But in each language the article itself is based on inherited material, different in each case. It is only on the basis of previous genetic classification that we can identify its sources, and only by comparison with other languages of the same genetic level that we can identify it as a change in inherited forms. Thus, the Romanian suffixed article -*l*, as in *calu-l* 'the horse' has the same origin as the French article *le/la/les* which occurs before the noun, and they have the same immediate source in the Latin *ille* 'that'. Similarly, standard Bulgarian has a suffixed article -*t* which has a common Slavic origin with non-Balkan Slavic languages and can be seen to be a convergent structural feature like that of the Romanian postposed article. Without a background, then, in genetic classification, such conclusions could not be reached.

So also for cultural vocabulary. Here comparison and reconstruction within a genetically defined group is of the essence. Without Proto-Indo-European, we have no right to posit a former linguistic community. Without the existence of reconstructed forms for numerals at least as high as 100 and of a reconstructible vocabulary which includes 'horse' and a number of other domesticated pastoral species, we would not be able to make statements about the economic subsistence type of Indo-European speakers. Genetic

classification is, then, the indispensable background, directly or indirectly, for all historical inferences drawn from languages in the absence of direct, written, historical records.

The chief types of inference with which we are concerned in regard to the settlement of the Americas is the number of such settlements, the relative—and, if possible, the absolute—chronology of their arrival, and the area of first settlement and that of subsequent groups derived by successive splits from the original group.

We may note that at the very beginning of the paper the question raised, naturally enough, was what the languages of Native Americans can tell us about the settlement of the Americas.

From this it might be thought that were it not for the abundance of direct historical evidence regarding European settlement, we would not be able to determine that the European languages—which for the sake of simplicity will be confined here to Spanish, Portuguese, French, English, and Dutch—were recent in the New World, nor would we be able to determine, on purely linguistic grounds, that the corresponding languages in Europe and elsewhere in the world originated in Europe only a few centuries ago.

If we consider any one of the European languages earlier mentioned in relation to its present differentiation and the geographical location of these variant forms, we always arrive at the same conclusions. For example, the still-surviving local dialects of French that are spoken in France show far more drastic and fundamental differences from each other than from the French spoken in Quebec, Louisiana, Martinique, and, for that matter, the French spoken in other parts of the world, such as, for example, Tahiti. In fact, when a language spreads it is normally only one variant, frequently some version of the 'standard' language, that is involved.

This leads us naturally to the conclusion that French spoken in the Americas is the result of recent migration from France. Similar results will follow from an examination of Spanish, Portuguese, English, and Dutch.

There are additional linguistic facts which greatly strengthen this already well-established conclusion. If we proceed to compare all of the five languages mentioned, we find a second-stage application of the same principle. It is almost immediately obvious that these five languages fall into two groupings: Spanish, Portuguese, and French, which are Romance languages, and English and Dutch, which are Germanic languages.

If we look at the geographical distribution of the remaining Romance languages, e.g. Catalan, Italian, Romansch, and Romanian, we see here a solid block of languages, all spoken in Europe and all showing considerable internal differentiation into local dialect forms. On the other hand, English

and Dutch go together as members of what is usually called the Germanic linguistic stock. Once more we find a solid block of languages, all spoken in Western Europe and each with considerable antiquity and internal differentiations.

In fact, we can divide Germanic languages genetically into North Germanic (Scandinavian, Icelandic, Faroese) and West Germanic (English, Frisian, Dutch, Low German, and High German). Dutch was also carried into South Africa where in its changed form it is called Afrikaans. The same kind of reasoning described earlier will lead to the conclusion that Dutch/Afrikaans originated in the Dutch/Flemish area of Europe and not in Africa. Within West Germanic, Anglo-Frisian forms a separate subgroup, while Dutch goes with Low German (Plattdeutsch). The general distribution of Germanic to the east of England on the continent, and the location of the Frisian Islands off the coast of Holland, Germany, and Denmark, suggests once more a continental European origin for English from the coasts and islands closest to the coast.

We can carry this kind of reasoning one or, probably, two steps further. Both the Germanic and Romance languages are members of the Indo-European family of languages which stretches, with interruptions, through the presence of Turkish and Semitic languages as far as India. Hence Germanic and Romance are among the western outliers of a vast distribution that centers further east. We can probably extend this analysis yet one more step in that Indo-European appears to be the westernmost branch of a vast family, Eurasiatic. Closest to Indo-European on the east are the Uralic languages which apparently have a special relationship to Yukaghir, still further east in the central and eastern parts of northern Siberia, and after that to the Altaic family with its three branches: Turkic, Altaic, and Tungus-Manchu. A further discussion of this vast stock which I have called Eurasiatic is postponed until later for its bearing on the settling of the Americas.

It will be noted that each successive higher genetic node in linguistic classification strengthens the case for the European rather than American origin of the Western European languages which now are spoken by the vast majority of the population of the New World. We have seen how, even in the absence of written attestation, it would have been possible to deduce the extreme recency of these languages in the Americas, which would lead us back to the numerous languages of the Native Americans in our search for linguistic evidence regarding the peopling of the New World.

Throughout this reasoning, a number of principles have been tacitly employed. One of these is that the existence of a hierarchic taxonomy in genetic linguistic classification leads us to evaluate each level separately in

terms of the distribution of the constituent languages as language groups of the next lower levels. The second is that we make no assumption regarding any inherent tendency of peoples to move in one direction or the other. This leads to the general notion that the most central area of the distribution of the component members is the most likely area of origin. This is based on 'the principle of least moves' (Dyen 1956).

Of course, external nonlinguistic reasons, including geographic conformations in terms of land and water and considerations of climate, as well as linguistic factors regarding earlier locations and evidence from contact shown through loanwords, play a role in modifying or specifying more exactly the original area and subsequent spread of a linguistic grouping.

For example, the existence of a stratum of words in Proto-Finno-Ugric, which are undoubtedly Indo-European and which point specifically to the Indo-Iranian branch of Indo-European, show that these peoples must once have been in contact, and our historical scenario must somehow account for this. Finno-Ugric is one of the two branches of Uralic, Samoyed being the other. There are no Indo-European loanwords in Samoyed, and there are a sufficient number of words common to Finno-Ugric and Samoyed, as well as other, often basic, terms and grammatical markers in either Finno-Ugric or Samoyed which agree with those of Indo-European to indicate that Indo-European and Uralic are related and are probably both members of a larger grouping, in fact the one that I have called Eurasiatic.

What might be called the 'center-of-gravity method,' which has just been sketched, is most effective and convincing when one stock of low genetic rank is peripheral in distribution and its fellow members are all located in some other area. This method was effectively used by Sapir in his well-known paper regarding the northern origins of the Navaho (Sapir 1936). Another example is that of the Bantu, who occupy almost all of the southern third of Africa and whose origin is to the northwest in the border area of Nigeria and the Cameroons (Greenberg 1963).

One further principle should, in fact, not even need mention from the scientific point of view, but in practice is often a major barrier. This principle might be called linguistic democracy. The number of speakers and the geographical expanse of a grouping are irrelevant. What counts are sheer linguistic differences and similarities. Thus, at the same genetic level, English, a world language with hundreds of millions of speakers, and Frisian, a language confined to small islands with a small and dwindling number of speakers, are of equal weight in determining the Anglo-Frisian homeland.

Guthrie, a well-known Bantuist, was outraged at the notion that Bantu should have any external connections or that it should be ranked at a very low

level genetically within Niger-Kordofanian and have, as its closest relations at its own level, minor languages of Nigeria with small numbers of speakers.

One major factor that has been considered up to now in the exposition only in an incidental way is chronology. We wish to know not only where and whence, but also when. It is clear that a valid, genetically based taxonomy of languages contains an inherent relative chronology. By definition, Proto-Germanic must be earlier than the present Germanic languages. This correlates with the degree of genetically relevant similarity, of course, precisely because such degrees of similarity are themselves the basis of the hierarchies incorporated in the genetic classification.

For absolute dating, the only purely linguistic method we have is glottochronology, which has many weaknesses, especially for long dates. Further discussion is here deferred to the point at which it becomes applicable to the analysis of the Native American linguistic situation, to which we now turn.

In 1960 the present writer outlined in brief form a linguistic classification of the native languages of the Americas [§4-Ed.], and this was published in much fuller detail in the book *Language in the Americas* (1987). According to this analysis there are three basic linguistic stocks of very unequal size: Eskimo-Aleut in the extreme north; Na-Dene, spoken exclusively in North America, for the most part in the northwestern part of that continent; and the large Amerind stock, including all of South and Central America and most of North America.

Before proceeding to the evidence from each of these three stocks regarding the settling of the Americas, I shall use a set of linguistic assumptions at slight variance with those adduced earlier to indicate that, with an extremely high probability, the Americas must have been settled from Asia. This is contrary to the view that extralinguistic data are required for such a statement.

Geographically, there are only three possibilities. The Americas were settled from Asia, from islands in the Pacific, or from Africa or Europe. It has already been shown that languages from Europe represent recent branches of Indo-European. The same is true for languages from Africa. Interestingly enough, earlier enumerations of indigenous language stocks of South America included a language called Arda, spoken in Colombia. Rivet (1925), alerted by the resemblance of Arda to Ardra, the language spoken in the slave-trading port of Dahomey, discovered that Arda, evidently spoken by a community of runaway slaves, was virtually identical with Ewe, a West African Niger-Congo language spoken in and around that area. Note that, even in the absence of other evidence, the same kind of reasoning illustrated repeatedly above would show the recent African provenance of Arda,

a member of the Niger-Congo subgroup of Niger-Kordofanian, the most widespread of the four major African language families.

In the Pacific islands, there are three major linguistic groups: Indo-Pacific, Australian, and Austronesian. Of these three, the Austronesian that is spoken on the islands closest to the Americas, e.g. Easter Island, the Hawaiian Islands, etc., has the least internal depth. Moreover, all of the islands that are closest to the Americas have languages of the Polynesian subgroup. There is much difference of opinion regarding the subgrouping of Austronesian, but there is unanimity that the Polynesian languages, which are very similar to each other, are at the lowest genetic level among the major subgroupings of Austronesian. According to Pawley and Green (1985) Polynesian would be at an intermediate genetic level, coordinate with Fiji and Rotuma; the center of genetic diversity within Austronesian as a whole is in the area comprising Taiwan and the Philippines. The next step in the analysis concerning Austronesian brings us directly to the mainland of Southeast Asia since, as shown by Benedict (1942), its closest relative is the Thai-Kadai group of languages of which Thai, the standard language of Thailand, is the only one with a large group of speakers.

The Indo-Pacific family shows far deeper internal differences than Austronesian and is doubtless much earlier in the Pacific. Its major groupings are almost all found on New Guinea and neighboring islands. The farthest eastern extension is in Melanesia, in the Solomon Islands, a vast distance indeed from the mainlands of North and South America. Australian is confined to the continent of Australia.

This leaves Asia as the only plausible source for the pre-European settling of the Americas. In first discussing this topic I stated it in terms of very high probability, not certainty. For example, it is not *impossible* that groups that left no linguistic relatives in their homeland emigrated at a very early date from Europe, Africa, or the Pacific, and these died out in the Americas. However, there is no positive evidence in the form of actual languages. The same holds for Asia, so that the three-migration theory enunciated in the remainder of this paper is really a '3 + n' theory.

Assuming the theory of three linguistic groups among the speakers of Native American languages, does this ensure that all three of these came from Asia separately and at different times? Once again it is possible to give an answer based on language alone. For there to have been only two or even one population movement with subsequent differentiation within the New World, some two or all three of these linguistic stocks must be shown not only to be related but to form a complete stock—that is, to have formed a valid linguistic entity without other members—thus presupposing a single

population which then differentiated into two or three branches. To answer the question, then, we must consider the external relations of these three families.

It is clear that Eskaleut, Na-Dene, and Amerind are not branches of the same stock, much less the *only* branches of such a stock. If this is true, then there cannot have been just one migration followed by the subsequent differentiation of this family into three branches.

It will be convenient to consider Na-Dene first. Genetically, Na-Dene is the most divergent of the three stocks. Sapir believed that Na-Dene was related to Sino-Tibetan. This idea has been taken up in a wider context by several Russian linguists, notably Starostin (1984) and Nikolaev (1989). Much of their evidence is as yet unpublished, but they have already made a plausible case for a widespread family of 'leftovers' which they call Sino-Caucasian. The language families they have connected are Ket, an isolated language in northeastern Siberia; Sino-Tibetan; North Caucasian (i.e. the non-Kartvelian Caucasian languages); and Na-Dene. If we assume the validity of this family, three of its four branches are in the Eastern Hemisphere, and the center-of-gravity type of reasoning already discussed suggests an Old World origin, perhaps in northern China or Manchuria.

With regard to Eskaleut and Amerind, an ultimate connection is highly probable, but not a direct genetic one in the sense that there is a single family which has Eskaleut and Amerind as branches.

In Greenberg (1987) it was hypothesized that Eskimo-Aleut was the easternmost branch of a Eurasiatic family for which the evidence will be presented in Greenberg [2000, 2002-Ed.]. Its membership consists of (1) Indo-European, (2) Uralic-Yukaghir, (3) Altaic, (4) Ainu-Korean-Japanese, (5) Gilyak, (6) Chukotian, (7) Eskimo-Aleut. Since Eskimo-Aleut is at the eastern end of a vast extension that centers in Asia, we are once more led to postulate an origin in Asia, presumably in central or western Siberia and to the north of the ancestral area of Na-Dene.

Amerind as a whole shows clear indication of a closer relation to Eskimo-Aleut than to Na-Dene. However, Amerind is, for a number of linguistic reasons not discussed here, to be viewed as coordinate with Eurasiatic as a whole, not as one of its branches. Once more Asia is indicated as the ultimate source, but at a greater linguistic time depth than Eskaleut, which was presumably not yet a distinct linguistic entity at the time the Amerindian population entered the New World.

The external linguistic evidence thus indicates an Asian provenance for all three stocks, and a more recent one for Eskimo-Aleut than for Amerind. This is reinforced by the internal linguistic evidence. The Amerind stock

has the greatest internal diversity of the three, followed by Na-Dene and Eskimo-Aleut.

For the moment, however, we can see that the geographical position of the three stocks reinforces the hypothesis that is beginning to emerge from our consideration of the data up to this moment: namely, that the Western Hemisphere was successively occupied by the Amerind, the Na-Dene, and the Eskaleut branch of Eurasiatic. This might be called the 'sock argument'. If we take out Christmas presents from a sock that hangs over the mantelpiece on Christmas morning, we will deduce that the present on the bottom was first put in, and so on. The geographical location of the three stocks conforms quite closely to a southern-northern progression: Amerind, Na-Dene, Eskaleut.

We may now consider the internal divisions of each of these stocks. Eskaleut divides into two branches, Eskimo and Aleut. While the difference between Eskimo and Aleut is not trivial, Rask recognized this grouping in the early nineteenth century and it has not been called into serious doubt since. Eskimo, in turn, divides into two groups based on the word for 'people', which became a kind of shibboleth in this matter. On this basis we have the Inuit and Yuit. There is a sharp division at Unalakleet on the central coast of western Alaska. Everything to the north, occupying not only northwestern Alaska and Canada, but also the west and east coasts of Greenland, is hardly more than one language. Its spread must therefore be very recent, within the last 1,500–2,000 years. It has been identified with the archaeological Thule culture.

The Yuit include not only the population in southwestern Alaska but also a few communities in Siberia, not far from the Bering Strait, which are considered to be a recent reflux. The position of Aleut and the small internal diversity of Inuit suggest southwestern Alaska as the area in which Proto-Eskaleut began to divide into Eskimo and Aleut.

The Athabaskans are by far the largest population speaking a Na-Dene language. They dominate the northwestern interior of Canada, with outliers in California and the American Southwest (Navaho-Apache). The greatest number of distinct branches of Athabaskan are in the interior of Alaska, and the supposition of their northwestern origin is strengthened by the fact that the language most closely related to Athabaskan is that of the virtually extinct small group of the Eyak in southeastern Alaska. At the next genetic remove is Tlingit, which ranks with Athabaskan-Eyak. Finally, most distant of all—in fact, not recognized by some as Na-Dene at all—is Haida, spoken on Queen Charlotte Island. The distribution of Na-Dene therefore suggests a central or insular origin in the southeastern extension of Alaska adjacent to Canada.

It should be interjected at this point that the earlier arrival of the Na-Dene, as compared to Eskaleut, is disputed by some who see Eskaleut settlement as somewhat earlier. The linguistic evidence suggests, however, the explanation described above: basically, the greater internal linguistic differentiation of Na-Dene than of Eskaleut and a more southerly geographical center of gravity.

The internal genetic divisions of Amerind and the historical inferences to be drawn from them present some complex problems. In Greenberg (1987) the etymologies confined to single subgroups are present in terms of eleven subgroups. Following this, etymologies found in two or more of these eleven are presented. However, of these eleven, two in South America, Macro-Ge (1) and Macro-Panoan (2), are especially close, and then at a further remove Macro-Carib (3), thus forming a Ge-Pano-Carib group. Since there were no previous internal comparisons within these groups and they had never before been defined fully in the literature, individual etymologies for each group were provided. In addition to the lexicon, grammatical evidence points to a special relationship among these three groups. The most important are the pronominal possessive prefixes: *i-* first person singular palatalizing; *a-* second person singular; *i-* third person nonpalatalizing and recurring before consonants, alternating with *t-* before vowels. Parts of the pattern are pan-Amerindian in origin, but the particular configuration is Ge-Pano-Carib. No special Ge-Pano-Carib section is contained in the book.

Two further South American groups, Equatorial (4) and Macro-Tucanoan (5), are suggested as having a special relationship. In addition, there is an Andean group in South America (6), and Chibchan-Paezan (7) which itself falls into two parts, Chibchan and Paezan, extending into Central America with one Paezan outlier, Timucua, found even in Florida.

Further, there is a Central Group (8) (Oto-Mangue, Uto-Aztecan, and Kiowa-Tanoan), extending from Central America to the American Southwest. Finally there are three groups, Penutian (9), Hokan (10), and Almosan-Keresiouan (11), found chiefly in North America with a few outliers further south, collectively called Northern Amerind. Once more, in Greenberg (1987) the intermediate groupings Equatorial-Tucanoan and Northern Amerind do not have separate etymological sections. Groupings of this level seemed evident to me in the course of working through the whole classification.

In the book, however, there is a matrix in Appendix C which shows the distribution of each etymology in relation to the eleven groups mentioned above. In Ruhlen (1991), this matrix is analyzed mathematically. It supports strongly all the intermediate groups mentioned earlier: (Ge-Pano-) Carib, Equatorial-Tucanoan, and Northern Amerind. It also, though somewhat

less clearly, suggests two further conclusions. The first is that the four most southern groups, found mainly, or in some instances exclusively, in South America, form a southern division of Amerind consisting of (1) Ge-Pano-Carib, (2) Equatorial-Tucanoan, (3) Andean, (4) Chibchan-Paezan. The second main conclusion is that Central Amerind (Oto-Mangue, Uto-Aztecan, Tanoan-Kiowa) stands apart from the rest and therefore probably presents the first cleavage within Amerind. Both these conclusions seem plausible.

If, as we assume, the Amerind stock came from Asia, we would expect its deepest internal diversity to be in the north. While this is not so, since the unity of Northern Amerind is strongly supported, the probable existence of a single southern group is gratifying as it shows that the deepest divisions are not within southern Amerind languages.

It would seem that the spread of Amerind must have been fairly rapid, i.e. within one or two thousand years, since numerous separate northern sub-branches of highest genetic rank within Amerind did not develop. The first split between Central and the rest seems to indicate an initial movement of Central Amerind into the American Southwest while the main body moved separately, leaving behind a Northern Amerind branch while the southern groups split up within South America. This topic is not pursued here further. For a more detailed reconstruction of the presumed movements of Amerind peoples, the reader is referred to Ruhlen (1991).

The general picture of three genetic groups moving separately from Asia into their present locations is supported by the existence of an almost identical threefold division based on fossil teeth and population genetics (Greenberg, Turner, and Zegura 1986). The correlation of language with population genetics is greatly strengthened by the monumental work of Cavalli-Sforza and his associates on the world distribution of genes in populations whose languages are spoken worldwide. Insofar as these results concern the Americas, they are strongly confirmatory. Recent work on mitochondrial DNA does not always give a clear picture, some of it agreeing and some disagreeing, but the technique is new, and I believe that we shall have to wait some time before this evidence can be reliably assessed.

What is the alternative to the view presented here in relation to the linguistic evidence? It would be to accept the view of what is probably the majority of linguists working on American Indian languages, according to which there are somewhere between 100 and 200 pre-Columbian linguistic stocks, among which no affinity can be traced. An approximate figure can only be given. There is no precise classification accepted by everyone. To take but one example, some Americanists accept part or all of the wider Hokan grouping, while some break it down into a dozen or more families.

Something like an 'official' list is found in Campbell and Mithun (1979) which, in spite of its title, *The Languages of Native America*, does not include South America or indeed parts of Central America. Here we find, with some hedging—for example regarding the relationship of Tlingit to Athabaskan-Eyak, said to be 'perhaps distantly related' (ibid.: 39), but included within Na-Dene in the listing—62 independent stocks compared to Powell's 1891 listing of 58 north of Mexico. Regarding South America, the closest thing to an 'official' listing I can find is Voegelin and Voegelin (1965: 146–50), with 93 independent stocks. In addition, one would have to add Central American [families] not found in Campbell and Mithun, of which twelve are listed in Voegelin and Voegelin. Six of these are now universally accepted as forming a single Otomanguean family.

Assuming that this represents an approximately true picture—namely, about 150 separate families with no known connection to each other—what are the historical implications for the settlement of the Americas?

A recent discussion, based on the assumption that my classification is wrong and the 'official' doctrine is true, is that of Nichols (1990). She refers at the very outset of her paper to Austerlitz (1980) as strikingly original, and since her paper is basically a continuation of the same approach, though with some important differences which will be noted, it will be discussed here first.

Austerlitz's paper appeared, of course, before Greenberg (1987) and as a matter of course assumes a very large number of families in the Americas. Actually Austerlitz only compares Eurasia and North America, omitting Africa, Oceania, Australia, and South America. He believes that 'there is something like an ideal density of language families ... differences among continents are not likely to be dramatic.' He proposes a measure called the GUDR (genetic unit density ratio): the number of linguistic genetic units in a continent divided by its area. Even with his highly conservative views, there are only 37 linguistic genetic units in Eurasia. This is then compared to 71 in North America, which has roughly half the area of Eurasia. It is assumed that genetic classification is based on the comparative method and that linguists apply it in the same manner in each area. It turns out then that there are about four times as many linguistic stocks per unit area in North America as in Eurasia. How to account for this 'dramatic discrepancy'? The reason he offers as the most plausible is that it would appear that the Old World, admittedly a much older inhabited area, was drained of more than half of its original stocks by migration to the New World. These groups migrated *in toto*, not leaving related languages behind. The reason for this potent *Drang nach Osten* is not given. Nichols in fact rejects this explanation (1990: 487) on the grounds that 'available evidence suggests that it is most typical for migrations

into new territory to produce distributions where part of the group moves and part stays behind.' Regarding South America, Austerlitz notes that it is enigmatic and that as many as seventy genetic units have been proposed, which would produce an even higher genetic density than North America and make the discrepancy between Eurasia and the Western Hemisphere 'still more dramatic'.

Nichols's paper continues the basic approach of Austerlitz but is more inclusive geographically. In her appendix she gives a linguistic survey of the entire world in which languages are classified on two levels: stocks and families. Thus, Indo-European is a stock, but Germanic and Celtic are among its constituent families. Many stocks have no major internal branching, e.g. Basque, and are therefore isolates. Stocks (ibid.: 477) are the oldest groupings reachable by the comparative method, are mostly in the vicinity of about 6,000 years old, and display regular phonological correspondences. The Niger-Congo stock is one of the major families which I was the first to distinguish as part of my African work (Greenberg 1963). In my classification of the languages of native America, Amerind, and even 'Hokan,' a subgroup of Amerind, were grouped above the stock level. These higher groups, which Nichols obviously views as speculative and uncertain, are not reckoned in her calculations, which, along the lines laid down by Austerlitz, consist of ratios of linguistic stocks and isolates to land areas.

Accepting a large number of different stocks in the New World, she therefore confronts the same problem as Austerlitz: namely, how to account for the great discrepancy between the far greater density of 'lineages' in the Eastern as opposed to the Western Hemisphere while rejecting the emptying-of-the-Old-World hypothesis of Austerlitz as unrealistic? Her answer can best be given by a citation from the abstract at the beginning of her paper:

The unmistakable testimony of the linguistic evidence is that the New World has been inhabited nearly as long as Australia or New Guinea, perhaps some 35,000 years. Genetic unity for 'Amerind' is incompatible with the chronology demanded by the linguistic facts.

The denial of the existence of an Amerind family is thus asserted by a proof *per impossibile*. Historical linguistics can only attain the stock level of about 6,000 years. Since such a vast number of separate stocks are found in America, the conventional chronology of the archaeologists is wrong. '"Amerind" presents us with a chronological paradox . . . If "Amerind" is a single genetic lineage, it is at least 50,000 years old.' On the other hand, the colonization of the New World by numerous independent lineages, which she of course favors, would also go well beyond the usually accepted chronology

as we have seen. Hence linguistic conservatism becomes allied with archae-
ological radicalism in chronology!

The reader may have noted by now a major contradiction in Nichols's
exposition. A stock is about 6,000 years old and attainable by the comparative
method. Niger-Congo and even the larger Niger-Kordofanian (of which
Niger-Congo is a constituent) are listed as stocks together with the other
three families that I distinguished in Africa: Nilo-Saharan, Afroasiatic, and
Khoisan. However, Afroasiatic, for which we have written attestation for
Semitic and Egyptian, is surely well beyond the 6,000-year limit.

According to footnote 1 on page 477 of Nichols's article, regarding stocks,
'Greenberg (1987) makes clear that he believes that such groupings [that is,
those higher than stocks] cannot be reached by the standard comparative
method; a wholly different method, mass comparison, is required.'

But I reached my African classification by mass comparison, and I
employed an identical method of classification in the Americas (and else-
where). If she accepts my conclusions for the Americas, the pseudo-paradox
she is seeking to explain simply disappears. No doubt more than 6,000 years
are required, but from what reasoning does this limit of 6,000 years come? It
has never been explained or justified.

If one looks at the table of common words in the languages of Europe in
Greenberg (1987: 12) [p. 95-Ed.], one sees that by the time one gets to the third
word, the division among Indo-European, Finno-Ugric, and Basque is clear
and confirmed again and again by further lexical items. If a correct hypothesis
can be generated by such a small part of the evidence, it can clearly go further
and generate chronologically deeper classifications. Indeed, if one were to
continue with such highly stable items as the first- and second-person pro-
nouns, the interrogative pronoun, and the word for 'water', it would even be
clear that Indo-European and Finno-Ugric group against Basque. Indo-
European and Uralic (which includes Finno-Ugric) belong to a large group of
languages, Eurasiatic, which includes Eskaleut, as noted earlier.

Finally, the question of methodology raised by Nichols's statement quoted
above should be discussed since it involves widely shared but false assumptions
about the relation between mass (or multilateral, as I now call it) comparison
and the comparative method. To either my credit or discredit, I am reputed to
have abandoned the comparative method in making language classifications.
Although Nichols states that I 'make clear that such groupings cannot be
reached by the comparative method', she cites no specific statement of mine.
The whole topic is treated at some length in Chapter 1 of Greenberg (1987)
[§6-Ed.]. It is, I believe, important to discuss it here, if only briefly, because
linguists and nonlinguists alike have misunderstood the whole question.

It is pardonable, therefore, for an archaeologist like Renfrew, repeating what he has heard from linguists, to make inaccurate statements. Thus in a recent paper (1991) Renfrew makes the following statement about multilateral comparison, after outlining my African classification:

The methodology employed proved controversial among linguists depending on what Greenberg terms 'multilateral comparison', that is to say, on lexical similarities studied in a number of languages at the same time. Previously, most linguistic comparisons had been made two at a time, but Greenberg claims to reach greater time depths with his multiple comparisons.

The first misunderstanding has to do with the notion that only lexical (and therefore not grammatical) items are compared. As early as my first essay on African linguistic classification (1949), in which I was seeking to exclude the irrelevant typological criteria which had confused earlier African classification, I sought to distinguish typological from genetically relevant criteria. Genetic criteria involve resemblances in form and meaning simultaneously, whether lexical, e.g. German *Nase*, English *nose*, or grammatical, as the German and English adjective comparatives in the suffix -*er*. These contrast with typological resemblances involving sound only, e.g. having glottalized stops, and meaning only, e.g. having a future tense. In fact, the resemblance in noun class systems, a grammatical set of criteria, played a major role in my first work on Niger-Congo, and the relation to Kordofanian was based more on these resemblances than on lexical items, few of which were available at the time. In my volume *Language in the Americas* (1987) there is an entire chapter on grammatical resemblances, and they play an important role in the book.

Secondly, and of at least equal significance, comparing only two languages at a time can never lead to a taxonomy of languages. It is no doubt widely practiced by American Indianists but does not represent orthodox historical linguistics. The work which is universally recognized as inaugurating historical linguistics is Bopp (1816), which compares Germanic, Latin, Greek, Sanskrit, and Persian. It was by this very fact that it was novel and epoch-making. All general treatments of Indo-European compare all the branches simultaneously. When Buck (1933) wrote a comparative grammar of Latin and Greek, he explained in his preface that while treating these two languages together was linguistically unjustified, the cultural relations between their speakers and the fact that they were the common concern of a specific body of scholars, the Classicists, made such a treatment convenient. However, all the forms in the two languages are explained by reference to Proto-Indo-European, which is, of course, reconstructed with the aid of many other

languages and thus involves tacit reference to multilateral comparison. Grammars of Balto-Slavic are different, and no apologia for them is given since they are generally presumed to stem from a common intermediate unity between Baltic and Slavic on the one hand and Proto-Indo-European on the other.

Finally, what is the relationship between multilateral comparison and the comparative method? There is no contradiction between the two. By the comparative method is meant the comparison of lexical and grammatical forms between members of the same genetic group of languages in order to reconstruct as far as possible the ancestral forms and the changes these forms underwent in becoming the forms of the later languages. Such a theory may be deemed explanatory in the historic sense. Later forms are explained from the earlier forms and the manner in which they gave rise to them. The most important fact about languages that makes this possible is the fact that most sound change is regular.

But to do this one must first determine which languages belong to the same genetic units at various levels. Since the number of ways of classifying n objects rapidly becomes astronomical, one needs a classification to begin with, and, as in any empirical science, one must first observe the objects to be classified. It is in this preliminary stage that multilateral comparison is clearly the only viable method.

In general, forms which descend from a more immediate common ancestor will be more alike both in sound and meaning than those from languages with a more remote ancestry. Phonetic and semantic change is, on the whole, from similar to similar. There are literally hundreds, perhaps thousands, of well-attested examples of 'similar' changes from *p*, a labial sound, to *f*, in contrast to strange or exotic changes: from 'nose' to 'nostril,' in contrast to the change from 'bead' to 'prayer.' Hence the observation of 'surface resemblances' leads to the correct classification.

In fact, it makes sense to see in classification based on multilateral comparison the first step in the comparative method itself. Sound correspondences do not spring like Athena from the head of Zeus. They are based at the beginning on the resemblances found in the initial stage of classification. At this stage, if we compare English *four* with German *vier*, the meaning is identical, and *f* corresponds to German *v* (pronounced *f*), *r* to *r*, and the vowels to each other.

At a later stage we posit the ancestral form and the regular recurrent correspondences. But whether we do this or not, the classification is valid, and the detailed application of the comparative method in the usual sense only becomes possible on its basis.

This point has now become clear at least to some comparative linguists. In a recent publication, Watkins, a distinguished Indo-Europeanist, cites with approval the remarks of Newman, an Africanist linguist whom I had previously cited on this matter (Watkins 1990). Newman asserts that the comparative method is only applied to languages already presumed to be related. But how do we presume them to be related except by looking at a broad array of them and noting how they group genetically? This, precisely, is multilateral comparison.

Note that, because in biology nothing corresponds to sound laws (Greenberg 1987: 34; Dyen 1987: 101–8), reconstruction of such entities as the proto-feline or proto-mammal is not feasible. But who will claim that biology is less taxonomically advanced than linguistics?

References

Austerlitz, Robert. 1980. 'Language family density in Asia and north Eurasia', *Ural-Altaische Jahrbücher* 52: 1–10.

Benedict, Paul. 1942. 'Thai, Kadai and Indonesian: a new alignment in southwestern Asia', *American Anthropologist* 44: 576–601.

Bopp, Franz. 1816. *Über das Conjugationssytem der Sanskrit-sprache in Vergleichung mit dem der griechischen, lateinischen, persischen und germanischen Sprache.* Frankfurt: Andreäische Buchhandlung.

Buck, Carl Darling. 1933. *Comparative Grammar of Latin and Greek.* Chicago: University of Chicago Press.

Campbell, Lyle and Mithun, Marianne (eds.). 1979. *The Languages of Native America: Historical and Comparative Assessment.* Austin: University of Texas Press.

Dyen, Isidore. 1956. 'Language distribution and migration theory', *Language* 32: 11–26.

——1987. 'Genetic classification in linguistics and biology', in George Cardona and Norman H. Zide (eds.), *Festschrift for Henry Hoenigswald.* Tübingen: Gunter Narr, 101–8.

Greenberg, Joseph H. 1949. 'Studies in African linguistic classification: I. The Niger-Congo family', *Southwestern Journal of Anthropology* 5: 79–100.

——1963. *The Languages of Africa.* Bloomington: University of Indiana Press.

——1987. *Language in the Americas.* Stanford: Stanford University Press.

——2000. *Indo-European and its Closest Relatives: the Eurasiatic Language Family, vol. 1: Grammar.* Stanford: Stanford University Press.

——2002. *Indo-European and its Closest Relatives: the Eurasiatic Language Family, vol. 2: Dictionary.* Stanford: Stanford University Press.

Greenberg, Joseph H., Turner, Christy G., II, and Zegura, Stephen. 1986. 'The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence', *Current Anthropology* 25: 477–97.

Nichols, Johanna. 1990. 'Linguistic diversity and the first settlement of the New World', *Language* 66: 475–521.

Nikolaev, S. 1989. 'Eyak-Athabaskan—North Caucasian sound correspondences', in Vitaly Shevoroshkin (ed.), *Reconstructing Languages and Cultures*. Bochum: Brockmeyer, 63–5.

Pawley, A., and Green, R. C. 1985. 'The Proto-Oceanic language community', in R. Kirk and E. Szathmary (eds.), *Out of Asia*, Canberra: Journal of Pacific History, 147–60.

Renfrew, Colin. 1991. 'Before Babel: speculations on the origins of linguistic diversity', *Cambridge Archaeological Journal* 1: 3–23.

Rivet, P. 1925. 'La langue arda ou une plaisante méprise', *Proceedings of the 21st International Congress of Americanists* 2. Lichtenstein: Nendeln, 388–90.

Ruhlen, Merritt. 1991. 'The Amerind phylum and the prehistory of the New World', in S. M. Lamb and E. D. Mitchell (eds.), *Sprung From Some Common Source: Investigations into the Prehistory of Languges*. Stanford: Stanford University Press, 328–50.

Sandfeld, K. 1968. *Linguistique balkanique.* Paris: Klincksieck.

Sapir, E. 1936. 'Internal linguistic evidence suggestive of the northern origin of the Navaho', *American Anthropologist* 38: 224–35.

Starostin, Sergei. A. 1984. 'Gipoteza o genetiskechix svjazax sinotebetskix jazykov s enisejskimi i severnokavkazkimi jazikami', *Konferencija: Nostratichiskije jazyki i nostraticheskoje jazykoznanije* 4: 74–94. Moscow.

Voegelin, C. F., and Voegelin, F. 1965. 'Native America', *Anthropological Linguistics* 7: 7, fascicle 2.

Watkins, C. 1990. 'Etymologies, equations and comparanda: types and values and criteria for judgment', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology.* Berlin: Mouton de Gruyter, 289–304.

Bibliography of Works Related to Joseph H. Greenberg's Theory and Methods for Genetic Linguistics

A. Works by Joseph H. Greenberg on Genetic Linguistics

This bibliography follows the format of the bibliography published in *On Language* (pp. 723–37; with emendations) and the corrected supplemental bibliography in *Language* 78: 560–4 (2002). I have included publications on genetic classification in general; publications on specific language family proposals; and publications containing data and discussion of specific proposals, including proposals by others and criticisms of Greenberg's proposals. The asterisked items are included in this collection.

1948

 8. The classification of African languages. *American Anthropologist* 50: 24–30.
10. The tonal system of Proto-Bantu. *Word* 4: 196–208.

1949

15. Studies in African linguistic classification: I. Introduction, Niger-Congo family. *Southwestern Journal of Anthropology* 5: 79–100. Reprinted in *Studies in African Linguistic Classification*, 1–23. New Haven, Connecticut: Compass Press, 1955.
16. Studies in African linguistic classification: II. The classification of Fulani. *Southwestern Journal of Anthropology* 5: 190–98. Reprinted in *Studies in African Linguistic Classification*, 24–32. New Haven, Connecticut: Compass Press, 1955.
17. Studies in African Linguistic Classification: III. The position of Bantu. *Southwestern Journal of Anthropology* 5: 309–17. Reprinted in *Studies in African Linguistic Classification*, 33–42. New Haven, Connecticut: Compass Press, 1955.
18. Review of Malcolm Guthrie, *The Classification of the Bantu Languages*. *Word* 5: 81–3. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 361–4. Stanford: Stanford University Press, 1990.

1950

21. Studies in African linguistic classification: IV. Hamito-Semitic. *Southwestern Journal of Anthropology* 6: 47–63. Reprinted in *Studies in African Linguistic Classification*, 43–61. New Haven, Connecticut: Compass Press, 1955.

22. Studies in African linguistic classification: V. The Eastern Sudanic Family. *Southwestern Journal of Anthropology* 6: 143–60. Reprinted in *Studies in African Linguistic Classification*, 62–79. New Haven, Connecticut: Compass Press, 1955.

23. Studies in African linguistic classification: VI. The Click languages. *Southwestern Journal of Anthropology* 6: 223–37. Reprinted in *Studies in African Linguistic Classification*, 80–94. New Haven, Connecticut: Compass Press, 1955.

24. Studies in African linguistic classification: VII. Smaller families; index of languages. *Southwestern Journal of Anthropology* 6: 388–98. Reprinted in *Studies in African Linguistic Classification*, 95–105. New Haven, Connecticut: Compass Press, 1955.

25. Review of L. D. Turner, *Africanisms in the Gullah Dialect. Journal of American Folklore* 63: 381–2.

26. Review of M. A. Bryan, *The Distribution of the Nilotic and Nilo-Hamitic Languages of Africa. Language* 26: 173–5.

28. Review of L. Homberger, *The Negro African Languages. Language* 26: 170–3.

1951

32. Vowel and nasal harmony in Bantu languages. *Zaïre, Revue Congolaise* 8: 813–20.

33. Review of Carl Meinhof, *Grundzüge einer vergleichenden Grammatik der Bantusprachen. Journal of the American Oriental Society* 71: 94–5.

1952

36. The Afro-Asiatic (Hamito-Semitic) present. *Journal of the American Oriental Society* 72: 1–9.

1953

38. An application of New World evidence to an African linguistic problem (Hausa). *Mémoires de l'Institut Français d'Afrique Noire* 27: 129–31.

*39. Historical linguistics and unwritten languages. *Anthropology Today*, ed. A. L. Kroeber, 265–86. Chicago: University of Chicago Press. Reprinted in *Language, Culture, and Communication: Essays by Joseph H. Greenberg*, selected and introduced by Anwar S. Dil, 11–46. Stanford: Stanford University Press, 1971.

40. (With Morris Swadesh). Jicaque as a Hokan language. *International Journal of American Linguistics* 19: 216–22.

44. A reply to W. Leslau: The imperfect in South-East Semitic. *Journal of the American Oriental Society* 73: 167–8.

45. Comments in *An Appraisal of Anthropology Today*, ed. Sol Tax *et al.*, 59–60, 232–3, 266–7, 292. Chicago: University of Chicago Press.

1954

48. Studies in African linguistic classification: VIII. Further remarks on method; revisions and corrections. *Southwestern Journal of Anthropology* 10: 405–15. Reprinted in *Studies in African Linguistic Classification*, 106–16. New Haven, Connecticut: Compass Press, 1955.

51. Review of Antoine Meillet and Marcel Cohen, *Les langues du monde. American Anthropologist* 56: 1133–4.

52. Review of D. Westermann and M. A. Bryan, *Languages of West Africa. Language* 30: 302–9. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 388–96. Stanford: Stanford University Press, 1990.

55. Review of W. Schmidt, *Die Tasmanischen Sprachen. Word* 10: 117–19.

1955

57. *Studies in African linguistic classification.* New Haven, Connecticut: Compass Press. Pp. 116. French translation in *Bulletin de l'Institut Français d'Afrique Noire*, Tome XVI.

58. Internal *a*-Plurals in Afroasiatic (Hamito-Semitic). *Afrikanistische Studien zum 80.en Geburtstag Diedrich Westermann gewidmet*, ed. Johannes Lukas. (Veröffentlichungen des Instituts für Orientforschung der Wissenschaften der DDR 26.) Berlin: Akademie-Verlag. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 397–404. Stanford: Stanford University Press, 1990.

59. Review of U. Weinreich, *Languages in Contact. American Anthropologist* 57: 167–8.

1956

63. Review of Oswin Köhler, *Geschichte der Erforschung der Nilotischen Sprachen. Language* 32: 563–7.

1957

*65. *Essays in Linguistics.* Chicago: University of Chicago Press. Pp. vii, 108. Also issued as *Viking Fund Publications in Anthropology, No.* 24. New York: Wenner-Gren Foundation for Anthropological Research, 1957. Pp. vii, 108. Chapter 3, Genetic relations among languages, reprinted in *Readings for the History of the English Language*, ed. E. T. Scott and J. L. Ericson, 35–47. Boston: Allyn and Badon, 1968. Chapter 5, Language and evolutionary theory, reprinted in *Readings in Anthropology, vol.* 1, ed. Morton H. Fried, 208–29. New York: Thomas Y. Crowell, 1968. Chapter 5, Language and evolutionary theory, and Chapter 6, Language, diffusion and migration, reprinted in *Language, Culture and Communication: Essays by Joseph H. Greenberg*, selected and introduced by Anwar S. Dil, 78–92, 93–105. Stanford: Stanford University Press, 1971. [Chapters 3–4 reprinted here; see also Chapter 6.]

67.   Nilotic, 'Nilo-Hamitic' and Hamito-Semitic: a reply. *Africa* 27: 364–78.

68.   Review of *A Bushman Dictionary*, by Dorthea F. Bleek, *Language* 33: 495–7.

1958

69.   The labial consonants of Proto-Afro-Asiatic. *Word* 14: 295–302. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 405–11. Stanford: Stanford University Press, 1990.

1959

74.   The origin of the Masai passive. *Africa* 29: 171–6. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 412–18. Stanford: Stanford University Press, 1990.

1960

78.   An Afro-Asiatic pattern of gender and number agreement. *Journal of the American Oriental Society* 80: 317–21.

79.   African tongues and tribes. *Rotarian* 96/4: 35, 61–2. Reprinted in *Language, Culture and Communication: Essays by Joseph H. Greenberg*, selected and introduced by Anwar S. Dil, 137–42. Stanford: University Press, 1971.

*80. The general classification of Central and South American languages. *Selected papers of the Fifth International Congress of Anthropological and Ethnological Sciences*, ed. A. Wallace, 791–4. Philadelphia: University of Pennsylvania Press.

81.   Linguistic evidence for the influence of the Kanuri on the Hausa. *Journal of African History* 1: 205–12. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 419–27. Stanford: Stanford University Press, 1990.

84.   Review of Henry M. Hoenigswald, *Language Change and Language Reconstruction. American Anthropologist* 62: 1108–10.

1962

90.   On the African affiliation of Hebrew and Semitic languages. *Jewish Social Studies* 24: 79–85.

1963

*97. *The Languages of Africa. International Journal of American Linguistics* 29:1 (Part II). Pp. vi, 171. (Publication of the Indiana University Research Center in Anthropology, Folklore and Linguistics, 25.) Bloomington: Indiana University and the Hague: Mouton. Revised edition, 1966. The Hague: Mouton and Company. Pp. vi, 180. Reprinted in *Problems in African History*, ed. Robert O. Collins: Niger-Congo 70–5 from 1963: 6–7, 30–3, 35–8; The languages of Africa 124–33 from 1963: 42–3, 49–51, 85–6, 89–94. Englewood Cliffs, New Jersey, 1968. [Chapter 1 reprinted here.]

98.   The Mogogodo, a forgotten Cushitic people. *Journal of African Languages* 2 (Part I): 29–43.

101. History and present status of the Kwa problem. *Actes du Second Colloque Internationale de Linguistique Africaine*, 215–7. Dakar: Université de Dakar, West African Languages Survey.

102. Langues et histoire en Afrique. *Présence Africaine* 45: 33–45.

1964

104. Historical inferences from linguistic research in Sub-Saharan Africa. *Boston University Papers in African History*, vol. 1, ed. J. Butler, 1–15. Boston: Boston University Press.

109. Reply to H. K. Schneider: Confusion in African linguistic classification. *Current Anthropology* 5: 56–67.

1965

113. The history and present status of African linguistic studies. *Proceedings of the First International Congress of Africanists, Accra, University of Ghana,* 1962, 85–96. London: Longmans. Reprinted in *Language, Culture and Communication: Essays by Joseph H. Greenberg*, selected and introduced by Anwar S. Dil. 212–27. Stanford: Stanford University Press, 1971. Polish translation, Historia i obecny stan Afrykanistycznych badan lingwistycznych, in *Problemy Afrikanistyki*, ed. Stefan Strelcyn, 120–33. Warszawa: Ponstwowe Wydawnictwo Naukowe.

114. The evidence for */mb/ as a Proto-Afroasiatic phoneme. *Symbolae Linguisticae in Honorem Georgii Kurylowicz, Bulietyn Towarzystwa Jezykoznawczego*, 88–92. Krakow: Polska Akademia Nauk.

115. Forward to David W. Crabb, *Ekoid Bantu, Languages of Ogoja, Eastern Nigeria, Part 1, Introduction, Phonology and Comparative Vocabulary*, ix-x. Cambridge, England: Cambridge University Press.

1966

116. Polyglotta evidence for consonant mutation in the Mandyak languages. *Sierra Leone Language Review* 5: 106–10.

118. Interdisciplinary perspectives in African linguistic research. *African Studies Bulletin* 9: 8–23. Reprinted in *Language, Culture and Communication: Essays by Joseph H. Greenberg*, selected and introduced by Anwar S. Dil, 228–48. Stanford: Stanford University Press, 1971.

1969

134. Review of I. Fodor, *The Problems in the Classification of the African Languages. Language* 45: 427–32. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 439–45. Stanford: Stanford University Press, 1990.

1971

148. Nilo-Saharan and Meroitic. *Current Trends in Linguistics, vol. VII: Linguistics in sub-Saharan Africa*, ed. Thomas A. Sebeok, 421–42. The Hague: Mouton.

*149. The Indo-Pacific hypothesis. *Current Trends in Linguistics, vol. VIII: Linguistics in Oceania*, ed. Thomas A. Sebeok, 808–71. The Hague: Mouton.

1972

151. Linguistic evidence regarding Bantu origins, *Journal of African History* 13: 189–216. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 446–75. Stanford: Stanford University Press, 1990.

152. On the identity of Jungraithmayr's Mimi. *Africana Marburgensia* 5/2: 45–9. Marburg: Universitätsbibliotek.

1974

157. Bantu and its closest relatives. *Studies in African Linguistics*, Supplement no. 5, 115–24.

1977

164. Niger-Congo noun class markers: prefixes, suffixes, both or neither. *Studies in African Linguistics*, Supplement 7, 94–104.

1978

166. Foreword to *L'Expansion bantoue* (Colloques Internationaux du Centre National de la Recherche Scientifique, Viviers 4–16 avril 1977), ed. Larry M. Hyman and Jan Voorhoeve, 9–12. Paris: SELAF.

173. How does a language acquire gender markers? *Universals of Human Language, volume 3: Word Structure*, ed. J. H. Greenberg *et al.*, 47–82. Stanford: Stanford University Press. Reprinted in *On language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 241–70. Stanford: Stanford University Press, 1990.

1979

176. The classification of American Indian languages. *Papers of the Mid-American Linguistic Conference at Oklahoma*, ed. Ralph Cooley *et al.*, 7–22. Norman, Oklahoma: University of Oklahoma Interdisciplinary Linguistics Program.

1981

180. African linguistic classification. *General history of Africa I. Methodology and African Prehistory*, ed. J. Ki-Zerbo, 292–308. Berkeley and Los Angeles: University of California Press.

181. Nilo-Saharan moveable *k-* as a stage III article (with a Penutian typological parallel). *Journal of African Languages and Linguistics* 3: 105–12. Reprinted in *On Language: Selected Writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 476–83. Stanford: Stanford University Press, 1990.

1985

184. (With Christy G. Turner II and Stephen Zegura). Convergence of evidence for peopling of the Americas. *Collegium Anthropologicum* 9: 33–42.

1986

187. (With Christy G. Turner II and Stephen Zegura). The settlement of the Americas: A comparison of the linguistic, dental and genetic evidence. *Current Anthropology* 25: 477–97.

1987

*192. *Language in the Americas.* Stanford: Stanford University Press. Pp. xvi, 438. [Chapter 1 and Appendix A reprinted here.]
193. *Current Anthropology* book review, *Language in the Americas*: précis and reply. *Current Anthropology* 28: 647–52, 664–6.

1988

*196. Language and archaeology [review of Colin Renfrew, *Archaeology and Language: the Puzzle of Indo-European Origins* and Merritt Ruhlen, *A Guide to the World's Languages, vol. 1: Classification*]. *The Quarterly Review of Archaeology* 9/3: 1, 8–9.
197. Review of Colin Renfrew, *Archaeology and Language: the Puzzle of Indo-European Origins. American Anthropologist* 90: 1029–30.

1989

*198. Classification of American Indian languages: a reply to Campbell. *Language* 65: 107–14.
199. Predystorija indoevropeyskoj sistemy glasnykh v sravnitel'noj i typologicheskoj perspektive. *Voprosy Jazykoznanija* 1989/4: 5–31. English original, The prehistory of the Indo-European vowel system in comparative and typological perspective. *Proto-languages and Proto-cultures*, ed. Vitaly Shevoroshkin, 77–136. Bochum: Brockmeyer, 1990.
*201. Proto-linguistic variation: a link between historical linguistics and socio-linguistics. *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, 91–101. Berkeley: Berkeley Linguistics Society.
202. Comments to Bateman *et al.*, Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language. *Current Anthropology* 31: 18–19.

1990

205. Relative pronouns and P.I.E. word order type in the context of the Eurasiatic hypothesis. *Language Typology 1987: Systematic Balance in Language*, ed. Winfred P. Lehmann, 123–38. Amsterdam: John Benjamins.
206. Correction to Matisoff, On megalocomparison. *Language* 66: 660.
207. The American Indian language controversy. *Review of Archaeology* 11/2: 5–14. Reprinted in *California Linguistic Newsletter* 22/3: 16–22, 1991.
208. Address to NACAL 18, Emory University, March 1990. *California Linguistic Newsletter* 20/3: 8–10.

1991

212. Some problems of Indo-European in historical perspective. *Sprung From Some Common Source: Investigations into the Prehistory of Languages*, ed. Sydney M. Lamb and E. Douglas Mitchell, 127–40. Stanford: Stanford University Press.
215. The last stages of grammatical elements: contractive and expansive desemanticization. *Approaches to Grammaticalization*, ed. Elizabeth Traugott and Bernd Heine, 301–14. Amsterdam: John Benjamins.
217. Ob utrate garmonii glasnykh v nekotorykh chukotsko-kamchatskikh jazykakh. (On the loss of vowel harmony in some Chukchi-Kamchatkan languages.) *Voprosy Jazykoznanija* 1991/3: 76–83.
218. [Paul Newman]. An interview with Joseph Greenberg. *Current Anthropology* 32: 453–467.

1992

219. [with Merritt Ruhlen]. Linguistic origins of Native Americans. *Scientific American* 267/5: 94–9.
220. Preliminaries to a systematic comparison between biological and linguistic evolution. *The Evolution of Human Languages*, ed. John A. Hawkins and Murray Gell-Mann, 139–58. Redwood City, CA: Addison Wesley.

1993

221. Observations concerning Ringe's *Calculating the Factor of Chance in Language Comparison*. *Proceedings of the American Philosophical Society* 137: 79–90.
*223. Linguistic typology and history: review of Johanna Nichols, *Linguistic Diversity in Space and Time*. *Current Anthropology* 34:.503–5.
224. The methods and purposes of linguistic genetic classification. *Proceedings of the Deseret Language and Linguistics Society* 1993 *Symposium*, ed. Dallin Oaks and Andrew Bay, 1–12. Provo: Deseret Language and Linguistics Society.

1994

226. On the Amerind affiliations of Zuñí and Tonkawa. *California Linguistic Notes* 24/2: 4–6.

1995

228. Indo-European *eg(h)hom/m*-suppletion in the context of Eurasiatic. *UCLA Friends and Alumni of Indo-European Studies Newsletter* 5/1: 14–15.

*230. The concept of proof in genetic linguistics. *Mother Tongue* 1: 207–16. Expanded version published in *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*, ed. Spike Gildea, 161–75. Amsterdam: John Benjamins, 2000. [Expanded version repinted here.]

1996

*231. In defense of Amerind. *International Journal of American Linguistics* 62: 131–64.

*232. Beringia and New World origins: the linguistic evidence. *American Beginnings: the Prehistory and Palaeoecology of Beringia*, ed. Frederick Hadleigh West, 525–36. Chicago: University of Chicago Press.

233. The genesis of multilateral comparison. *Mother Tongue* 2: 145–8.

*234. Review of Vladimir E. Orël and Olga V. Stolbova, *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction. Anthropological Linguistics* 38: 550–6.

235. Genes, languages and other things: review of *History and Geography of Human Genes* by Luca Luigi Cavalli-Sforza, Paolo Menozzi, and Alberto Piazza. *Review of Archaeology* 16/2: 24–8.

236. The 'Greenberg hypothesis'. *Science* (29 November) 274: 1447.

1997

*237. Does Altaic exist? *Indo-European, Nostratic and Beyond: A Festschrift for Vitaly V. Shevoroshkin*, ed. Irén Hegedus, Peter A. Michalove and Alexis Manaster Ramer, 88–93. Washington, DC: Institute for the Study of Man.

238. The Indo-European first and second person pronouns in the perspective of Eurasiatic, especially Chukotian. *Anthropological Linguistics* 39: 187–95.

239. Mary Haas, Algic and the scientific consensus. *Anthropological Linguistics* 39: 668–72.

240. Some grammatical evidence for Eurasiatic, especially pronominal. *The Twenty-third LACUS Forum* 1996, ed. Alan K. Melby, 59–65. Chapel Hill, NC: Linguistic Association of Canada and the U.S.

1998

*241. The convergence of Eurasiatic and Nostratic. *Nostratic: Sifting the Evidence*, ed. Joseph C. Salmons and Brian D. Joseph, 51–60. Amsterdam: John Benjamins.

1999

*242. On lumping and splitting in linguistics and biology. *Review of Archaeology* 20/1: 17–18.

*243. Are there mixed languages? *Essays in Poetics, Literary History and Linguistics Presented to Viacheslav Vsevolodovich Ivanov on the Occasion of his Seventieth Birthday*, ed. L. Fleishman *et al.*, 626–33. Moscow: OGI.

2000

245. *Indo-European and its Closest Relatives: the Eurasiatic Language Family, vol.* 1: *Grammar.* Stanford: Stanford University Press. Pp. xi, 326. French translation, *Les langues indoeuropéenes et la famille eurasiatique, vol.* 1: *Grammaire.* Paris: Belin.

246. From first to second person: the history of Amerind **k(i).* *Functional Approaches to Language, Culture and Cognition*, ed. David G. Lockwood, Peter H. Fries, and James E. Copeland, 413–25. Amsterdam: John Benjamins.

248. Review of Michael Fortescue, *Language Relations across Bering Strait: reap-Praising the Archaeological and Linguistic Evidence. (Open Linguistics Series).* *Review of Archaeology* 21/2: 23–4.

2001

249. The methods and purposes of linguistic genetic classification. *Language and Linguistics* 2: 111–36.

2002

250. *Indo-European and its Closest Relatives: the Eurasiatic Language Family, vol.* 2: *Lexicon.* Stanford: Stanford University Press. Pp. 216.

to appear

*252. Indo-Europeanist practice and American Indianist theory in linguistic classi-fication. *Language and Prehistory in the Americas: Assessments of Greenberg's Classification*, ed. Allan R. Taylor.

253. [with Merritt Ruhlen] *Amerind Etymological Dictionary.*

B. Reviews, Commentaries, and Discussion of Joseph H. Greenberg's Works on Genetic Linguistics

The literature on historical linguistic methodology is vast. This bibliography includes only scholarly publications that contain discussion of Greenberg's methods and results. Every attempt has been made to make this as complete a bibliography as possible. However, inevitably some sources will have been missed; I express my apologies to those authors whose work was unintentionally left out of this bibliography.

1. *General*

Baldi, Philip. 1990. 'Introduction: the comparative method', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin and New York: Mouton de Gruyter, 1–13.

Baxter, William H. and Manaster Ramer, Alexis. 2000. 'Beyond lumping and splitting: probabilistic issues in historical linguistics', in Colin Renfrew, April M. S. McMahon, and R. Lawrence Trask (eds.), *Time Depth in Historical Linguistics*. Cambridge: McDonald Institute for Archaeological Research, 167–88.

Campbell, Lyle. 1999. *Historical Linguistics: An Introduction*. Cambridge, Mass.: MIT Press.

Campbell, Lyle. 2003. 'How to show languages are related: methods for distant genetic relationship', in Brian D. Joseph and Richard D. Janda (eds.), *Handbook of Historical Linguistics*. Oxford: Blackwell, 262–82.

Cavalli-Sforza, Luca, *et al.* 1988. 'Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data', *Proceedings of the National Academy of Sciences in the USA* 85: 6002–6.

Chrétien, C. Douglas. 1969. 'Genetic linguistics and the probability model', *Language* 42: 518–30.

Comrie, Bernard. 1989. 'Genetic classification, contact and variation', in Thomas G. Walsh (ed.), *Synchronic and Diachronic Approaches to Linguistic Variation and Change*. (Georgetown University Round Table on Language and Linguistics 1988.) Washington, DC: Georgetown University Press, 81–93.

Dimmendaal, Gerrit Jan. 1993. 'Review of Keith Denning and Suzanne Kemmer (eds.), *On Language: Selected Writings of Joseph H. Greenberg*', *Language* 69: 796–807.

Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Doerfer, Gerhard. 2000. 'Greenberg und der Niedergang der diachronischen Linguistik', *Indogermanischen Forschungen* 105: 27–46.

Dyen, Isidore. 1959. 'Review of *Essays in Linguistics* by Joseph H. Greenberg', *Language* 35: 527–52.

Fleming, Harold C. 1987. 'Toward a definitive classification of the world's languages', *Diachronica* 4: 159–224.

—— 2000–1. 'Joseph Harold Greenberg: a tribute and an appraisal', *Mother Tongue* 6: 9–27.

Fox, Anthony. 1995. *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press.

Gordon, Matthew. 1995. 'The phonological composition of personal pronouns: implications for genetic hypotheses', in Jocellyn Ahlers, Leela Bilmes, Joshua S. Guenter, Barbara A. Kaiser, and Ju Namkung (eds.), *Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 117–28.

Hetzron, Robert. 1976. 'Two principles of genetic reconstruction', *Linguistics* 38: 89–108.

Hock, Hans Henrich. 1993. 'Swallow tales: chance and the "world etymology" MALIQ'A "swallow, throat"', in Katharine Beals, Gina Cooke, David Kathman, Sotaro Kita, Karl-Erik McCullough, and David Testen (eds.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society, vol. 1: The Main Session.* Chicago: Chicago Linguistic Society, 215–38.

Hock, Hans Henrich and Joseph, Brian D. 1996. *Language History, Language Change and Language Relationship: An Introduction to Historical and Comparative Linguistics.* Berlin: Mouton de Gruyter.

Hymes, Dell H. 1959. 'Genetic classification: retrospect and prospect', *Anthropological Linguistics* 1/2: 50–66.

——1960. 'Lexicostatistics so far', *Current Anthropology* 1: 3–44.

Justeson, John S. and Stephens, Laurence D. 1980. 'Change cognation: a probabilistic model and decision procedure for historical inference', in Elizabeth Closs Traugott, Rebecca LaBrum, and Susan Shepherd (eds.), *Papers from the Fourth International Conference on Historical Linguistics.* Amsterdam: John Benjamins, 37–46.

Kessler, Brent. 2001. *The Significance of Word Lists.* Stanford: Center for the Study of Language and Information.

Laks, Bernard. 2002. 'Le comparatisme: de la généalogie à la génétique', *Langages* 146: 19–45.

Lass, Roger. 1997. *Historical Linguistics and Language Change.* Cambridge: Cambridge University Press.

Lehmann, Winfred P. 1958. 'Review of *Essays in linguistics* by Joseph H. Greenberg', *International Journal of American Linguistics* 24: 142–5.

Lohr, Marisa. 1998. 'Methods for the genetic classification of languages'. Ph.D. dissertation, University of Cambridge.

Manaster Ramer, Alexis. 1993. 'Language in Prehistory: Western Eurasia and North America', in André Crochetière, Jean-Claude Boulanger, and Conrad Ouellon (eds.), *Actes du XVe Congrès International des Linguistes: Les Langues menacées/ Endangered Languages: Proceedings of the XVth International Congress of Linguists*, vol. 1. Sainte-Foy: Presses Universitaires Laval, 150–2.

Manaster Ramer, Alexis and Hitchcock, Christopher. 1996. 'Glass houses: Greenberg, Ringe, and the mathematics of comparative linguistics', *Anthropological Linguistics* 38: 601–19.

Manaster Ramer, Alexis and Sidwell, P. 1997. 'The truth about Strahlenberg's classification of the languages of northeast Siberia', *Journal de la Société Finno-Ougrienne* 87: 139–60.

McQuown, Norman A. 1954. 'Review of *Anthropology Today* (Linguistics), ed. by A. L. Kroeber', *American Anthropologist* 56: 475–80.

——1957. 'Review of *Essays in Linguistics* by Joseph H. Greenberg', *American Anthropologist* 59: 1132–3.

Michalove, Peter A., Georg, Stefan, and Manaster Ramer, Alexis. 1998. 'Current issues in linguistic taxonomy', *Annual Review of Anthropology* 27: 451–72.

Nehring, Alfons. 1958. 'Review of *Essays in Linguistics* by Joseph H. Greenberg', *Anthropos* 53: 291–3.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.

Nichols, Johanna. 1996. 'The comparative method as heuristic', in Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* Oxford: Oxford University Press, 39–89.

Peter, Steve. 1993. 'All in the family? Greenberg's method of mass comparison and the genetic classification of languages', in Katharine Beals, Gina Cooke, David Kathman, Sotaro Kita, Karl-Erik McCullough, and David Testen (eds.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society, vol. 1: The Main Session.* Chicago: Chicago Linguistic Society, 329–42.

Pierce, Joe E. 1965. 'On the validity of genetic classifications', *Linguistics* 13: 25–33.

Poser, William. J. 1993*a*. 'Book notice on *A Guide to the World's Languages, vol. I: Classification* (2nd ed.), by Merritt Ruhlen', *Language* 69: 220–2.

——1993*b*. 'Book notice on *On Calculating the Factor of Chance in Language Comparison*, by Donald A. Ringe, Jr', *Language* 69: 635–6.

Poser, William J. and Campbell, Lyle. 1992. 'Indo-European practice and historical methodology', in Laura A. Buszard-Welcher, Lionel Wee, and William Weigel (eds.), *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society.* Berkeley: Berkeley Linguistics Society, 214–36.

Ringe, Donald A., Jr. 1992. *On Calculating the Factor of Chance in Language Comparison.* (Transactions of the American Philosophical Society, vol. 82, part 1.) Philadelphia: American Philosophical Society.

——1993. 'A reply to Professor Greenberg', *Proceedings of the American Philosophical Society* 137: 91–109.

——1999*a*. 'How hard is it to match CVC-roots?' *Transactions of the Philological Society* 97: 213–44.

——1999*b*, 'Language classification: scientific and unscientific methods', in Bryan Sykes (ed.), *The Human Inheritance: Genes, Language and Evolution.* Oxford: Oxford University Press, 45–73.

——2000. 'Some relevant facts about historical linguistics', in Colin Renfrew (ed.), *America Past, America Present: Genes and Languages in the Americas and Beyond.* Cambridge: McDonald Institute for Archaeological Research, 139–62.

Ross, Malcolm and Durie, Mark. 1996. 'Introduction', in Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* Oxford: Oxford University Press, 3–38.

Ruhlen, Merritt. 1991. *A Guide to the World's Languages, vol. 1: Classification* (2nd edn.). Stanford: Stanford University Press.

——1992. 'An overview of genetic classification', in Murray Gell-Mann (ed.), *The Evolution of Human Languages.* Reading, MA: Addison-Wesley, 159–90.

Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press.

—— 2000. 'Some unanswered linguistic questions', in Colin Renfrew (ed.), *America Past, America Present: Genes and Languages in the Americas and Beyond*. Cambridge: McDonald Institute for Archaeological Research, 163–75.

Salmons, Joe. 1992. 'A look at the data for a global etymology: **tik* "finger"' in Garry W. Davis and Gregory K. Iverson (eds.), *Explanation in Historical Linguistics*. Amsterdam: John Benjamins, 207–28.

Tax, Sol *et al.* 1953. *An Appraisal of Anthropology Today*. Chicago: University of Chicago Press. [Comments on Greenberg's 'Historical linguistics and unwritten languages' (§1 published in this volume), by Haas, Held, Voegelin, Heizer, Greenberg, and Kroeber, pp. 56–61.]

Thomason, Sarah Grey. 1993. 'Coping with partial information in historical linguistics', in Henk Aertsen and Robert J. Jeffers (eds.) *Historical Linguistics 1989: Papers from the Ninth International Conference on Historical Linguistics*, Amsterdam: John Benjamins, 485–96.

Watkins, Calvert. 1990. 'Etymologies, equations and comparanda: types and values and criteria for judgement', in Philip Baldi (ed.), *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*. Berlin: Mouton de Gruyter, 289–304.

2. *Africa*

Bender, M. Lionel. 1976. 'Nilo-Saharan overview', in M. Lionel Bender (ed.), *The Non-Semitic Languages of Ethiopia*. East Lansing: African Studies Center, Michigan State University, 439–83.

—— 1996. 'Nilo-Saharan '95', *Afrikanistische Arbeitspapiere* 45: 1–25.

Bendor-Samuel, John T. 1989. *The Niger-Congo Languages: A Classification and Description of Africa's Largest Language Family*. Lanham, Maryland: University Press of America.

Berry, Jack. 1956. 'Review of *Studies in African Linguistic Classification* by Joseph H. Greenberg', *Bulletin of the School of Oriental and African Studies* 18: 395.

Bouquiaux, Luc. 1967. 'Le système des classes nominales dans quelques langages (birom, ganawuri, anaguta, irigwe, kaje, rukuba, appartenant au groupe "Plateau" (Nigeria centrale))', in Gabriel Manessy (ed.), *La classification nominale dans les langues négro-africaines*. Paris: Centre National de la Recherche Scientifique, 133–56.

Bynon, James. 1984. 'Berber and Chadic: the lexical evidence', in James Bynon (ed.), *Current Progress in Afro-Asiatic Linguistics*. Amsterdam: John Benjamins, 241–90.

Bynon, James and Bynon, Theodora. 1968. 'Review of *The Problems in the Classification of the African Languages* by István Fodor', *Journal of African Languages* 7: 219–20.

Cole, Desmond T. 1960. 'African linguistic studies, 1943–1960', *African Studies* 19: 219–29.

Crabb, David W. 1965. *Ekoid Bantu Languages of Ogoja, Eastern Nigeria, Part I: Introduction, Phonology and Comparative Vocabulary.* Cambridge: Cambridge University Press.

Dalby, David. 1970. 'Reflections on the classification of African languages', *African Language Studies* 11: 147–71.

——1971. 'A referential approach to the classification of African languages', in Chin-wu Kim and Herbert Stahlke (eds.), *Papers in African Linguistics.* Edmonton: Linguistic Research, 17–31.

Delord, J. 1965. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *La Linguistique* 1: 142–4.

Der-Houssikian, Haig. 1972. 'The evidence for a Niger-Congo hypothesis', *Cahiers d'Études africaines* 12: 316–22.

Diakonoff, Igor M. 1965. *Semitoxamitskii jazyki: opyt klassifikacii.* Moscow: Akademia Nauk SSSR, Institut narodov Azii. (Translated as *Semito-Hamitic Languages: an Essay in Classification.* Moscow: Nauka, 1965.)

Ehret, Christopher. 1986. 'Proposals on Khoisan reconstruction', *Sprache und Geschichte in Afrika* 7: 105–30.

——2000. 'Is Krongo after all a Niger-Congo language?' in R. Vossen, A. Mietzner, and A. Meissner (eds.), '*Mehr als nur Worte...*': *Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland.* Köln: Rüdiger Köppe, 225–37.

Fleming, Harold C. 1986. 'Hadza and Sandawe genetic relations', *Sprache und Geschichte in Afrika* 7: 157–87.

Flight, Colin. 1981. 'Trees and traps: strategies for the classification of African languages and their historical significance', *History in Africa* 8: 43–74.

——1988. 'The Bantu expansion and the SOAS network', *History in Africa* 15: 261–301.

Fodor, István. 1966. *The problems in the Classification of the African Languages; Methodological and Theoretical Conclusions concerning the Classification System of Joseph H. Greenberg.* Budapest: Center for Afro-Asian Research, Hungarian Academy of Sciences. (Fourth edition published as *A Fallacy of Contemporary Linguistics: J. H. Greenberg's Classification of the African Languages and his 'Comparative Method'.* Hamburg: Helmut Buske, 1982.)

Gleason, H. A. 1956. 'Review of *Studies in African Linguistic Classification* by Joseph H. Greenberg', *American Anthropologist* 58: 948–9.

Goodman, Morris. 1970. 'Some questions on the classification of African languages', *International Journal of American Linguistics* 36: 117–22.

Gregersen, Edgar A. 1972. 'Congo-Saharan', *Journal of African Languages* 11/1: 69–89.

——1977. *Language in Africa.* New York: Gordon and Breach.

Guthrie, Malcolm. 1964. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *Journal of African History* 5: 135–6.

Heine, Bernd. 1971. 'On genetic relationship and the case of the "Nilo-Hamitic" languages', *Journal of Eastern African Research and Development* 1/1: 19–28.

Heine, Bernd. 1972. 'Historical linguistics and lexicostatistics in Africa', *Journal of African Languages* 11/3: 7–20.

Heine, Bernd and Nurse, Derek (eds.). 2000. *African Languages: An Introduction.* Cambridge: Cambridge University Press.

Hodge, Carleton T. 1983. 'Afroasiatic: the horizon and beyond', *Jewish Quarterly Review* 74: 137–58.

Hohenburger, J. 1957. 'Comparative Masai word list', *Africa* 26: 281–7.

——1958. 'Some notes on "Nilotic, 'Nilo-Hamitic,' and Hamito-Semitic", by Joseph H. Greenberg', *Africa* 28: 37–42.

Houis, M. 1964. Review of *The Languages of Africa*, by Joseph H. Greenberg. *Bulletin de l'Institut Français de l'Afrique Noire* 26: 286–95.

Huntingford, G. W. B. 1956. 'The "Nilo-Hamitic" languages', *Southwestern Journal of Anthropology* 12: 200–22.

Kaye, Alan S. and Daniels, Peter T. 1992. 'Comparative Afroasiatic and general genetic linguistics (review article on *Proceedings of the Fourth International Hamito-Semitic Congress*, ed. Hermann Jungraithmayr and Walter M. Müller.)'. *Word* 43: 429–58.

Knappert, Jan. 1987. 'The Bantu languages: an appraisal', *Archives Européennes de Sociologie* 28: 177–91.

Köhler, Oswin. 1955. *Geschichte der Erforschung der nilotischen Sprachen.* Berlin: Dietrich Reimer.

——1960. 'Sprachkritische Aspekte zur Hamitentheorie über die Herkkunft der Hottentotten', *Sociologus* 10: 69–77.

——1963. 'Observations on the Central Khoisan group', *Journal of African Languages* 2: 227–34.

——1981. 'Les langues khoïsan', in Jean Perrot (ed.), *Les langages dans le monde ancien et moderne*, vol. 1. Paris: Éditions du CNRS, 455–615.

Lacroix, Pierre-Francis. 1966. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *Bulletin de la Société de Linguistique de Paris* 61/2: 208–14.

——1971. 'L'ensemble songhay-jerma: problèmes et thèmes de travail', in *Actes du Huitième Congrès International de Linguistique Africaine*, Société Linguistique de l'Afrique Occidentale. Abidjan: Université d'Abidjan, 87–99.

Leslau, Wolf. 1953. 'The imperfect in South-East Semitic', *Journal of the American Oriental Society* 73: 164–6.

——1958. 'Observations of a Semitist on recent etymologies proposed by Africanists', *Africa* 28: 324–8.

Meeussen, A. E. 1957. 'Hamietisch en Nilotisch', *Zaïre, Revue Congolaise* 11: 263–72.

——1963. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *Journal of African Languages* 2: 170–1.

Monteil, V. 1965. 'La classification des langues de l'Afrique', *Bulletin de l'Institut Français d'Afrique Noire* 27: 155–68.

Montgomery, Christine A. 1969. 'Review of *Linguistic Analyses: The Non-Bantu Languages of Northeast Africa*, by Archibald N. Tucker and Margaret A. Bryan', *Language* 45: 661–73.

Newman, Paul. 1980. *The Classification of Chadic within Afroasiatic*. Leiden: Universitaire Pers.

—— *On Being Right: Greenberg's African Linguistic Classification and the Methodological Principles which Underlie it*. Bloomington, Indiana: Institute for the Study of Nigerian Languages and Cultures, African Studies Program, Indiana University.

Newman, Paul and Ma, Roxanna. 1966. 'Comparative Chadic: phonology and lexicon', *Journal of African Languages* 5: 218–51.

Nicolai, Robert. 1983. 'Position, structure and classification of Songay', in M. Lionel Bender (ed.) *Nilo-Saharan Language Studies*. East Lansing, Michigan: Africa Studies Center, Michigan State University, 11–41.

Olivier, Roland. 1966. 'The problem of the Bantu expansion', *Journal of African History* 7: 361–76.

Sands, Bonny. 1998. *Eastern and Southern African Khoisan: Evaluating Claims of Distant Linguistic Relationships*. (Quellen zur Khoisan-Forschung, 14.) Köln: Rüdiger Köppe.

Schadeberg, Thilo C. 1981. 'The classification of the Kadugli language group', in Thilo C. Schadeberg and M. Lionel Bender (eds.), *Nilo-Saharan: Proceedings of the First Nilo-Saharan Linguistics Colloquium*. Dordrecht: Foris, 291–305.

Schafer, Robert. 1959. 'Phonétique comparée du Nigéro-Sénégalien (Mande)', *Bulletin de l'Institut Français de l'Afrique Noire* 21: 179–200.

Sebeok, Thomas A. (ed.). 1971. *Current Trends in Linguistic*s, *vol. VII: Linguistics in Sub-Saharan Africa*. The Hague: Mouton.

Shimizu, Kiyoshi. 1975. 'A lexicostatistical study of Plateau languages and Jukun', *Anthropological Linguistics* 17: 413–18.

Siertsema, B. 1964. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *Lingua* 13: 85–7.

Stahlke, Herbert F. W. 1974. 'Review of *Current Trends in Linguistics, vol. VII: Linguistics in Sub-Saharan Africa*, edited by Thomas A. Sebeok', *Language* 50: 195–205.

Stevenson, Roland C. 1991. 'Relationship of Kadugli-Krongo to Nilo-Saharan: morphology and lexis', in M. Lionel Bender (ed.), *Proceedings of the Fourth Nilo-Saharan Linguistics Conference*. Hamburg: Helmut Buske, 347–69.

Tucker, Archibald. N. 1957. 'Philology and Africa', *Bulletin of the School of Oriental and African Studies* 20: 541–54.

Welmers, William E. 1956. 'Review of *Studies in African Linguistic Classification* by Joseph H. Greenberg', *Language* 32: 556–63.

—— 1963. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *Word* 19: 407–17.

—— 1973. *African Language Structures*. Berkeley: University of California Press.

Westermann, Dietrich. 1952. 'African linguistic classification', *Africa* 22: 250–6.

Westphal, E. O. J. 1957. 'On linguistic relationship', *Zaïre, Revue Congolaise* 11: 513–24.

—— 1962. 'A reclassification of Southern African non-Bantu languages', *Journal of African Languages* 1: 1–8.

—— 1964. 'Review of *The Languages of Africa*, by Joseph H. Greenberg', *American Anthropologist* 66: 1446–9.

Winston, F. D. 1966. 'Greenberg's classification of African languages', *African Language Studies* 7: 160–71.

3. *Southeast Asia and Oceania*

Benedict, Paul K. 1994. 'Greenberg and I: the case of Southeast Asian "vulva" ', *Linguistics of the Tibeto-Burman Area* 17: 177–8.

Blust, Robert. 1978. 'Review of *Current Trends in Linguistics*, VIII: *Linguistics in Oceania*, ed. Thomas A. Sebeok', *Language* 54: 467–80.

Franklin, Karl. 1973. 'The Gulf area in the light of Greenberg's Indo-Pacific hypothesis', in Karl Franklin (ed.), *The Linguistic Situation in the Gulf District and Adjacent Areas.* Canberra: Research School of Pacific Studies, Australia National University, 377–409.

Usher, Timothy. 2002. 'The origins of the Tasmanian languages', *Mother Tongue* 7: 65–84.

Wurm, Stephen A. 1975. 'Possible wider connections of Australian languages: Papuan and Australian; Greenberg's Indo-Pacific hypothesis', in Stephen A. Wurm, (ed.), *Papuan Languages and the New Guinea Linguistic Scene.* Canberra: Research School of Pacific Studies, Australia National University, 925–32.

Wurm, Stephen A. 1982. *Papuan Languages of Oceania.* Tübingen: Gunter Narr.

4. *The Americas*

Adelaar, Willem F. H. 1989. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Lingua* 78: 249–55.

Bateman, Richard, Goddard, Ives, O'Grady, Richard, Funk, V. A., Mooi, Rich, Kress, W. John, and Cannell, Peter. 1990. 'Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language', *Current Anthropology* 31: 1–13.

Bateman, Richard *et al.* 1990. 'On human phylogeny and linguistic history: reply to comments', *Current Anthropology* 31: 177–82.

—— 1990. 'Reply to comments', *Current Anthropology* 31: 315–16.

Berman, Howard. 1992. 'A comment on the Yurok and Kalapuya data in Greenberg's *Language in the Americas*', *International Journal of American Linguistics* 58: 230–48.

Blackmer, Elizabeth R. 1988. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Canadian Psychology/Psychologie Canadienne* 29: 223–4.

Blount, Ben B. 1990. 'Comment on Bateman *et al.*, "Speaking of forked tongues" ', *Current Anthropology* 31: 15.

Bright, William. 1984. 'The classification of North American and Meso-American Indian languages', in William Bright (ed.), *American Indian Linguistics and Literature*. Berlin: Mouton, 3–29.

—— 1988. 'Review of Greenberg, *Language in the Americas*', *American Reference Book Annual* 88: 440. Englewood, Colorado: Libraries Unlimited.

Callaghan, Catherine A. 1990. 'Comment on Bateman *et al.*, "Speaking of forked tongues" ', *Current Anthropology* 31: 15–16.

Campbell, Lyle. 1986. 'Comment on Greenberg, Zegura and Turner, "The settlement of the Americas" ', *Current Anthropology* 27: 488.

—— 1988. 'Review article on *Language in the Americas* by Joseph H. Greenberg', *Language* 64: 591–615.

—— 1991. 'On so-called Pan-Americanisms', *International Journal of American Linguistics* 57: 394–9.

—— 1994. 'Problems with the pronouns in proposals of remote relationships among Native American languages', in Margaret Langdon (ed.), *Proceedings of the Meeting of the Society for the Study of Indigenous Languages of the Americas and the Hokum-Penutian Workshop* (Survey of California and Other Indian Languages, report 8). Berkeley: University of California, 1–20.

—— 1997a. *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press.

—— 1997b. 'Amerind personal pronouns: a second opinion', *Language* 73: 339–51.

Campbell, Lyle and Goddard, Ives. 1990. 'Summary report: American Indian languages and principles of language change', in Philip Baldi (ed.), *Linguistic Change and Reconstruction Methodology*. Berlin and New York: Mouton de Gruyter, 17–32.

Chafe, Wallace. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 652–3.

Darnell, Regina. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 653–6.

Fox, James A. 1986. 'Comment on Greenberg, Zegura and Turner, "The settlement of the Americas" ', *Current Anthropology* 27: 488–9.

Goddard, Ives. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 656–7.

—— 1990. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Linguistics* 28: 556–8.

Goddard, Ives and Campbell, Lyle. 1994. 'The history and classification of American Indian languages: what are the implications for the peopling of the Americas?' in Robson Bonnichsen and D. Gentry Steele (eds.), *Method and Theory for Investigating the Peopling of the Americas*. Corvallis, Oregon: Center for the Study of the First Americans, 189–207.

Golla, Victor. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 657–9.

—— 1988. 'Review of *Language in the Americas* by Joseph H. Greenberg', *American Anthropologist* 90: 434–5.

Haas, Mary R. 1958. 'Algonkian-Ritwan: the end of a controversy', *International Journal of American Linguistics* 24: 159–73.

Hamp, Eric. 1987. 'Comment on Greenberg, Zegura and Turner, "The settlement of the Americas"', *Current Anthropology* 28: 101.

Hymes, Dell. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 659–62.

Kaufman, Terrence. 1990. 'Language history in South America: what we know and how to know more', in Doris Payne (ed.), *Amazonian Languages*. Austin: University of Texas Press, 13–74.

Kimball, Geoffrey. 1992. 'A critique of Muskogean, "Gulf," and Yukian materials in *Language in the Americas*', *International Journal of American Linguistics* 58: 447–501.

Lamb, Sydney. 1987. 'Comment on Greenberg, Zegura and Turner, "The settlement of the Americas"', *Current Anthropology* 28: 101–2.

Liedtke, Stefan. 1989. 'Review of Greenberg, *Language in the Americas*', *Anthropos* 84: 283–5.

——1991. *Indianersprachen Vergleich und Klassifizierung: eine ethnolinguistische Einführung in die Grundlagen und Methoden*. Hamburg: Buske.

Manaster Ramer, Alexis. 1994. 'Languages in the Americas', *California Linguistic Newsletter* 24/1: 21–5.

——1996*a*. 'Tonkawa and Zuñí: two test cases for the Greenberg classification', *International Journal of American Linguistics* 62: 264–88.

——1996*b*. 'Sapir's classifications: Haida and the other Na-Dene languages', *Anthropological Linguistics* 38: 179–215.

Matisoff, James A. 1990. 'On megalocomparison', *Language* 66: 106–120.

McMahon, April M. S. and McMahon, Robert. 1995. 'Linguistics, genetics and archaeology: internal and external evidence in the Amerind controversy', *Transactions of the Philological Society* 93: 125–225.

Mithun, Marianne. 1990. 'Studies of North American Indian languages', *Annual Review of Anthropology* 19: 309–30.

Munro, Pamela. 1994. 'Gulf and Yuki-Gulf', *Anthropological Linguistics* 36: 125–222.

Muntzel, Martha G. 1987. 'Review of Greenberg, *Language in the Americas*', *América Indígena* 47: 701–4.

Newman, Paul. 1993. 'Greenberg's American Indian classification: a report on the controversy', in Jaap van Marle (ed.), *Papers from the Tenth International Conference on Historical Linguistics, Amsterdam 12–16 August 1991*. Amsterdam: Benjamins, 229–42.

Nichols, Johanna. 1990. 'Linguistic diversity and the first settlement of the New World', *Language* 66: 475–521.

Nichols, Johanna and Peterson, David. 1996. 'The Amerind personal pronouns', *Language* 72: 336–71.

Oswalt, Robert L. 1990. 'Comment on Bateman *et al.*, "Speaking of forked tongues"', *Current Anthropology* 31: 21–2.

Payne, David L. 1990. 'Some widespread forms in South American languages', in Doris L. Payne (ed.), *Amazonian Languages*. Austin: University of Texas Press, 75–88.

Pinnow, Heinz-Jürgen. 1990. *Die Na-Dene-Sprachen im Licht der Greenberg-Klassifikation*. Notorf: Völkerkundliche Arbeitsgemeinschaft. (Abhandlungen, Heft 64.)

Poser, William J. 1992. 'The Salinan and Yurumanguí data in *Language in the Americas*', *International Journal of American Linguistics* 58: 202–29.

—— 1995. 'Binary comparison and the history of Hokan comparative studies', *International Journal of American Linguistics* 61: 135–44.

Rankin, Robert L. 1992. 'Review of *Language in the Americas* by Joseph H. Greenberg', *International Journal of American Linguistics* 58: 324–51.

Rankin, Robert L. 1998. 'Grammatical evidence for genetic relationship and the Macro-Siouan problem', in Marie-Lucie Tarpent (ed.), *Papers from the 21st Annual Meeting of the Atlantic Provinces Linguistic Association/Actes du Colloque Annuel de l'Association de Linguistique des Provinces Atlantiques (PAMAPLA)*. Halifax, Nova Scotia: Atlantic Provinces Linguistic Association, 20–44.

Ringe, Donald A., Jr. 1996. 'The mathematics of "Amerind"', *Diachronica* 13: 135–54.

Ruhlen, Merritt. 1994. 'Linguistic evidence for the peopling of the Americas', in Robson Bonnichsen and D. Gentry Steele (eds.), *Method and Theory for Investigating the Peopling of the Americas*. Corvallis, Oregon: Center for the Study of the First Americans, 177–88.

Sapir, J. David. 1987. 'Review of *Language in the Americas* by Joseph H. Greenberg', *Current Anthropology* 28: 663–4.

Swadesh, Morris. 1963. 'Comment on "On aboriginal languages of Latin America"', *Current Anthropology* 4: 317–19.

Taylor, Allan P. (ed.). Forthcoming. *Language and Prehistory in the Americas: A Conference on the Greenberg Classification*.

Weiss, Kenneth M. and Woolford, Ellen. 1986. 'Comment on Greenberg, Zegura and Turner, "The settlement of the Americas"', *Current Anthropology* 27: 491–2.

5. *Eurasia*

Bengtson, John D. 2000–1. 'Review of *Indo-European and its Closest Relatives: the Eurasiatic Language Family*, vol. 1, by Joseph H. Greenberg', *Mother Tongue* 6: 131–5.

Bomhard, Allan R. 2002. 'Reflections on Greenberg's *Indo-European and its Closest Relatives: the Eurasiatic Language Family*', *Mother Tongue* 7: 85–114.

Georg, Stefan and Vovin, Alexander. 2003. 'From mass comparison to mess comparison [review article on *Indo-European and its Closest Relatives: The Eurasiatic Language Family*, vol. 1, by Joseph H. Greenberg]', *Diachronica* 20: 331–62.

Lehmann, Winfred P. 2000–1. 'What constitutes scientific evidence in paleolinguistics?' *Mother Tongue* 6: 75–8.

Perdicoyianni-Paleologou, H. 2002. 'Review of *Indo-European and its Closest Relatives: the Eurasiatic Language Family*, vol. 1, by Joseph H. Greenberg', *Canadian Journal of Linguistics/Revue canadienne de linguistique* 47: 99–100.

Ringe, Donald A., Jr. 2002. 'Review of Joseph H. Greenberg, *Indo-European and its Closest Relatives*', *Journal of Linguistics* 38: 415–20.

Vajda, Edward. 2003. 'Review of *Indo-European and its Closest Relatives: the Eurasiatic Language Family*, vols. 1–2, by Joseph H. Greenberg', *Journal of Indo-European Studies* 31: 241–5.

Author Index

This page intentionally left blank

Subject Index