Annu. Rev. Anthropol. 1997. 26:359–84

MODELING ANCIENT POPULATION STRUCTURES AND MOVEMENT IN LINGUISTICS

Johanna Nichols

Department of Slavic Languages and Literatures, University of California, Berkeley, Berkeley, California 94720-2979; e-mail: johanna@uclink.berkeley.edu

KEY WORDS: language family, language spread, linguistic diversity, language dispersal, linguistic geography

ABSTRACT

Linguistic population structure is described in terms of language families. Geographical distributions of language families respond to climate, latitude, and economic factors. Characteristic shapes of phylogenetic descent trees for language families reflect particular types, rates, directionalities, and chronologies of spread. Languages move in predictable ways in particular geographical, economic, and social contexts. In this chapter, the linguistic prehistories of four continents are surveyed with regard to linguistic spreads, linguistic diversity, and language family origins.

INTRODUCTION

Languages enter into descent lineages, areal groupings, and typological classes. On the map they bunch up densely in some places and spread out widely in others. Over time they expand or contract in range and move in space, sometimes to great distances. They are adopted or abandoned by speech communities. Some give rise to daughter languages, some die out. Some change rapidly, some slowly. These various processes are favored or constrained by geographical and other external circumstances, and identifying them is useful in tracing linguistic prehistory and human prehistory more generally.

The most recent period of ferment in linguistic phylogeny, when principles were reviewed together with actual classification, was midcentury, when

0084-6570/97/1015-0359$08.00

works by Lamb (1959), Haas (1969), Hoenigswald (1960), and Voegelin & Voegelin (1964–1966) were produced partly in response to the bloom in field description of Native American languages. The past two decades have seen remarkable progress in the description and historical analysis of previously undescribed languages and in computational linguistics and statistical methods—as well as in archeology and human genetics—and it is time to review principles, classification, and linguistic prehistory again. Because the main obstacles to interdisciplinary communication appear to be nonlinguists' lack of understanding of how languages move and what constitutes a valid genetic grouping, most attention is given to these issues here.

Basic claims of historical linguistics are not individually referenced here; recent introductions include Crowley (1992), Hock (1991), Hock & Joseph (1996), McMahon (1994), Thomason & Kaufman (1988), and Trask (1996). No general reference work on the world's languages reflects current scientific knowledge; Moseley & Asher (1994) probably come closest. Unless otherwise indicated, all statements about numbers of languages in areas, ranges of languages, and population densities refer to the linguistic world as it was, or as it can be reconstructed, as of about 1492.

LINGUISTIC POPULATION STRUCTURE

The day-to-day work of historical-comparative linguistics can be called Neogrammarian comparison, after the self-designation of the field's theoretical founding school. It comprises the search for cognate vocabulary and regular sound correspondences within descent groups, reconstruction of ancestral sounds and words on the basis of that evidence, and the definition of subgrouping relations within the family. Neogrammarian method is called "the comparative method" in linguistics, but there is also a broader field of comparative linguistics whose concerns are not only genealogical.

Phylogenetic Trees and Linguistic Descent

The basic building block of linguistic populations is the descent group or clade. Neogrammarian comparison describes linguistic clades whose genetic relatedness is already established. Common descent is shown by shared unique innovations—features so specific and so unlikely to be diffused that their independent innovation is virtually precluded, and languages sharing them can with near-certainty be assumed to descend from a common ancestor. Genetic markers, as descent-proving features can be called, can be of three types:

1. An ordered sequence or other patterned set of form-function pairings whose overall probability of occurrence is about two orders of magnitude less

than the probability of any one language turning up in a random draw. (There are some 5000–6000 languages, hence $1/5000 \times 0.01 = 0.000002$ or two in a million). A four-consonant sequence or ordered set meets this threshold. The person prefix paradigm of the Algonquian family of North America, for example, consists of first person **ne-,* second person **ke-,* third person **we-,* and indefinite **me-* (Goddard 1975). This small but highly structured paradigm meets the threshold and is a strong genetic marker. It was recognized by Sapir (1913) and Haas (1958) as diagnostic of genetic identity between the Algonquian family and two languages of coastal California—Yurok and Wiyot (see Goddard 1975). Another genetic marker is the gender-number suffixes masculine *-n,* feminine *-t,* plural (masculine = feminine) *-n* in certain pronominal paradigms, which is diagnostic of Afroasiatic genetic reality (Greenberg 1960). For the computation and other examples see Nichols (1995, 1996).

2. Significantly greater than chance frequency of phoneme matchings in semantically identical words from a standard 100-word list. Tests of this type are offered by Oswalt (1991), Bender (1969), and Ringe (1992); applications of the latter are Ringe (1995a,b, 1996, 1997). (The mathematics of the significance thresholds is under revision for Ringe's test, but the basic procedure is valid.) All three use the 100-word basic vocabulary list proposed by Morris Swadesh; the first two seek greater-than-expected numbers of matching identical segments, and the third seeks greater-than-expected numbers of correspondences (not necessarily of identity) for which expected thresholds are based on the actual frequencies of segments in the two wordlists. The test of Nichols (1995) uses naturalistic ranges of phonological and semantic variation. All these are binary tests that can show that two languages are related but do not circumscribe families or identify subfamilies.

3. Although this has not been attempted, in principle a set of features—none a genetic marker by itself but all of known diachronic stability and relatively low frequency worldwide, and all demonstrably independent of each other—could as a set reach the threshold of genetic markers and serve as a diagnostic of deep genetic connection.

It is often claimed that demonstrating regular systematic sound correspondences among a set of languages (i.e. among putative cognates) constitutes demonstration of genetic relatedness. It does not. Systematic correspondences can be found among any random set of languages. It is not that systematic correspondences can be found within fixed wordlists (such as the Swadesh 100- or 200-word list) from randomly chosen languages. They cannot. Rather, a statistical wild card lurks in the assembling of putative cognate sets by the analyst. In normal stocks and families, cognates range widely in their semantics. For example, descendants of PIE **leuk-* 'light, brightness' listed in Watkins (1985) include 'light,' 'moon,' 'purify,' 'meadow,' 'shine,' 'flame, fire,' 'lamp,' 'rabies,' and possibly 'lynx,' and these are only the cognates retrievable from the English vocabulary. The analyst trying to demonstrate genetic relatedness generally believes in the relatedness and is therefore motivated to pursue all plausible semantic connections. This means that the procedure for seeking cognates is a series of more or less open-ended searches each of which ends with the first success, and a search rapidly inflates the probability of finding a match (Nichols 1995, 1996). Proving genetic relatedness therefore requires that semantics be fixed or firmly controlled; hence the available heuristic methods seek out that small number of cognates whose meanings are unchanged and test whether that small number is significantly above the expected frequency. There are a number of proposed deep "families" that are said to be based on the comparative method but are actually based on premature Neogrammarian comparison and are therefore spurious (or at least unproved). Space limitations preclude reviewing them here.

All known tests give the occasional false negative when applied to known families. Greenberg's gender-number paradigm, for example (see above), is present in its entirety in three branches of Afroasiatic—Semitic, Chadic, Berber—but only partly in the others. Ringe's test fails to detect relatedness of ancient Greek to some of its sisters because the test seeks consonant correspondences and Greek has lost several consonants. For both of these families, the identity and bounds of the families can be established by other means. The false negatives are the consequence of language change, which has reshaped paradigms in Afroasiatic and eroded consonants in Greek. Because all languages constantly change and for any linguistic element there is a nonzero probability of radical mutation or complete loss, over time genetic markers become fewer in number among related languages, tests yield false negatives more often, and eventually genetic relatedness cannot be proven. Inherited elements remain in the languages, but they are statistically indistinguishable from chance resemblances, and their number also decreases over time.

This fade-out effect provides some salient thresholds in the decay of genetic markers and dissolution of families. The following are commonly recognized stages, though terminology varies. The terms used here are those of Lamb (1959), up to the level he considers to represent established relatedness.

FAMILY This is the standard general term for a proven clade of any age. A relatively young family has a clear grammatical signature with obvious sharings of cognate morphemes in similar functions, numerous lexical cognates, and plentiful genetic markers. Families in the age range of 3500–4500 years include Kartvelian, Dravidian, Athabaskan, and Mayan. Classical Indo-European as first discovered consists of a 3500–4000-year-old core of Sanskrit, ancient Greek, and Latin. For younger families such as Romance, Germanic, Slavic, Turkic, Polynesian, Athabaskan, Algonquian, Quechuan—all around 2000 years in age—relatedness is unmistakable even to the nonlinguist, and some mutual intelligibility exists. Families aged closer to 6000 years, such as Uralic, Austronesian, and Semitic or Indo-European as judged only from modern attestations, exhibit genetic markers and cognate etyma that are not always evident to the nonspecialist.

STOCK This is a commonly used term for a maximal reconstructable clade, i.e. the oldest families displaying regular sound correspondences and amenable to Neogrammarian comparative method. A stock usually displays several genetic markers. The oldest known stocks are about 6000 years old: e.g. Indo-European, Uralic, and Austronesian.

If a family or a language isolate has no demonstrated kin and does not enter into any demonstrable stock, then it also constitutes its own stock. For technical clarity it can be called a stock-level family or stock-level isolate. A standard reference work using terminology of this sort is Wurm & Hattori (1983). Any isolate lower group is counted as a stock when taking a census of stocks. For instance, two or three stocks are indigenous to Western Europe: Indo-European, Basque (an isolate), and, marginally, Semitic (represented by Maltese).

QUASI-STOCK A quasi-stock is a quasi-genetic or probabilistic grouping of more than one stock which shares one or more features that are valid or promising genetic markers but which have few clear cognates and for which systematic regular sound correspondences cannot be demonstrated. Hence a quasi-stock is a probable clade but not a fully describable one and is not amenable to reconstruction. Linguistic understanding of what constitutes a promising genetic marker is still evolving and changing the received view of what is a promising quasi-stock.

The type-defining example is probably Niger-Kordofanian, a set of families and stocks mostly of sub-Saharan Africa including the Bantu family (Bendor-Samuel 1989, Greenberg 1963), discussed below. The genetic marker of Niger-Kordofanian is its complex systems of generally prefixal genders (also called concord classes), in which there are particular prefixes for particular classes and systematic correspondences between singular and plural concord classes. The system is shared widely among the daughter branches and is identifiable as a system even when individual elements are greatly changed or lost. This kind of gender system is quite specific and quite rare worldwide and thus useful as a genetic marker. Yet systematic sound correspondences and regular lexical reconstructability are absent from Niger-Kordofanian, the internal structure of the genetic tree is still in doubt (cf overview chapters in Bendor-

Samuel 1989), and whole subgroups lack the gender system. The usual interpretation is that at least the gender-using branches are sisters descendant of a lineage so ancient that little detectable shared material survives apart from the distinctive gender system.

An atypically stock-like quasi-stock is Afroasiatic, another African group established by Greenberg (1963). It consists of the families Chadic and Berber, the isolate Egyptian, the Semitic stock, a likely isolate Beja, a stock or pair of stocks Cushitic, and possibly the family Omotic (see e.g. Bender 1989, Newman 1980). The Afroasiatic quasi-stock has a distinctive grammatical signature that includes several morphological features at least two of which independently suffice statistically to show genetic relatedness beyond any reasonable doubt (the entire set is listed in Newman 1980; for statistical significance, see Newman 1980 and Nichols 1996). Hence it is routinely accepted as a genetic grouping, though uncontroversial regular correspondences cannot be found and a received reconstruction may never be possible [recent serious attempts are Orel & Stolbova (1995) and Ehret (1995); both are controversial]. Its age is quite uncertain, though clearly older than that of classic stocks like Indo-European because some of the component branches are themselves stocks of Indo-European-like age. An estimate of 10,000 years is sometimes cited (e.g. Newman 1980).

Quasi-stocks can also be detected by lexical tests. Among the several cases of clear significance for known sisters and chance-level correspondences for random pairs of languages, the pairing of Indo-European and Uralic stands out as near-significant (Ringe 1997), though there are no genetic markers in the grammatical structure.

STRUCTURAL POOL This term, which I use ad hoc in this chapter, labels any group of stocks exhibiting some property or set of properties that is unusual or infrequent worldwide, though not so unusual or of such low probability as to be a genetic marker. The click languages of southern Africa have click consonants root-initially in most of their major-class words (nouns, verbs, adjectives) and nowhere else. This defines a structural canon type, not a genetic grouping (the click languages consist of one small family, two probable isolates, and possibly a third isolate; see below), but it is found nowhere else on earth. In fact, clicks themselves are found nowhere on earth except in these click languages and, sporadically and with a different distribution, in a few of their neighbors from the Bantu family and one Cushitic neighbor. (For clicks, see Ladefoged & Maddieson 1996.) This distribution is clearly nonrandom and testifies to some kind of historical interaction or connection among the click languages, but not to genetic relatedness.

Another structural pool is the set of stocks hugging the Pacific Rim of the Americas that exhibit personal pronoun systems with *n* in first person forms and/or *m* in the second person. Personal pronouns are normally inherited and almost never borrowed, but the *n:m* system is too small and too fraught with potential phonosymbolism and inflated probability of occurrence due to allomorphy and the cross-linguistic high frequency of nasals to be a good genetic marker. Dozens of languages and one or two dozen stocks with such systems, however, are found in the American Pacific Rim, the geographical skewing is highly significant, and other low-frequency features cluster there as well. Although not a demonstrable clade, the cluster is nonrandom and has some kind of historical identity (Nichols & Peterson 1996).

There are approximately 300 separate stocks on earth, which further comparative work may reduce to as few as 200 quasi-stocks, some of which will surely prove to be true stocks. The tests described above offer the prospect of being able to extend the fade-out point regularly to the time depth represented by the age of Afroasiatic or Indo-Uralic; reducing the world's stocks to 200 will require reaching such a time depth. (Estimates of these time depths are usually in the vicinity of 10,000 years.) Given present knowledge of language change and probability, however, descent and reconstruction will never be traceable beyond approximately 10,000 years. Methods now being developed reach back much earlier but do not trace descent. Among other things, this means that linguistics will never be able to apply phylogenetic analysis to the question of when language arose and whether all the world's languages are descended from a single ancestor.

Diversification

The rate at which languages diverge into dialects and then into daughter languages, and so on, is not constant, nor is the number of dialects or daughter languages, though whether they are relatively many or few depends on ascertainable cultural and historical factors, described below. Rates of change are accelerated by contact with other languages: The more profound the influence, the more rapid the change. There is no generally useful unit in terms of which grammar change might be measured. Lexical change is commonly measured by loss of items from a standard 100- or 200-word list of basic glosses. (Loss occurs when a word either changes meaning or drops out of use entirely.) A proposed constant rate of loss (for recent textbook summaries, see Crowley 1992; Trask 1996) makes it possible to estimate ages of families. This metric is called *glottochronology.* As the rate of vocabulary loss is not absolutely regular, accuracy is improved by computing retention percentages for larger numbers of daughter languages and by adjusting the constant for known areal effects or isolation. Embleton (1986, 1991) has proposed a different formula based on the number of adjacent languages, replacement rate in the basic word lists of the languages at issue, number of borrowings between the languages, and a measure of overall similarity between them; this method gives excellent accuracy for the language families Embleton has tested.

Glottochronology presupposes establishment of genetic relatedness and enough comparative work to distinguish cognate from noncognate vocabulary. Hence it is not valid in principle beyond the level of the stock, and even at that level it is not valid unless prior comparative work has been done.

For well-reconstructed families and under the right combination of circumstances, dating by comparison to dated archeological evidence can be quite accurate. The reconstructed Proto-Indo-European lexicon contains a sizable technical terminology for wheeled transport, terms for the major domesticated animals including the horse, but no clear terms for metals, all of which points to a date around 3500 BC (Anthony 1991, 1995; Mallory 1989). This in turn is consistent with glottochronological dates (Tischler 1973) and with the degree of differentiation exhibited by the earliest attested languages in the second and early first millennia BC.

Extinction

There are various causes of language death: the speech community is killed off, e.g. by genocide or natural disaster; the speech community is scattered; or the speech community abandons the language and shifts to another, as when Gauls shifted to Latin or Coptic speakers shifted to Arabic (for language shift, see Thomason & Kaufman 1988). The result for family trees is pruning, which may remove individual daughter languages (as when the Celtic branch of Indo-European lost Cornish through language shift to English) or entire branches (Indo-European has lost its Anatolian and Tocharian branches and others; the Afroasiatic quasi-stock has lost the Egyptian/Coptic branch). The death of a language isolate causes an entire lineage to die out, as has happened in historical times with Sumerian, Elamite, Etruscan, Iberian, northern Pictish, and no doubt countless others. Language spreads of all kinds cause extinction of languages previously in the area, usually through language shift.

Language shift usually involves an intermediate stage of society-wide bilingualism. It is favored when the target language is economically useful or functions as a vehicle of interethnic communication, or when the shifting community is accepting of linguistic variation (Hill 1996).

Language extinction by natural disaster and language shift are natural processes that have always gone on. Consequently, the branching rates in family trees drawn for surviving languages are not diversification rates but survival rates. Survival rates are less than diversification rates because there is always a nonzero probability of extinction, and this entails that older genetic groupings and more ancient nodes in family trees have, on average, fewer initial branches than younger ones. The only cross-linguistic survey of branching done so far (Nichols 1990) finds that stocks have, on average, about 1.5 initial branches; that is, many stocks are stock-level families or stock-level isolates.

Contact and Convergence

Languages whose speech communities are in contact acquire words, sounds, and even elements of grammatical structure from one another in what is known as *contact-induced change* (Thomason & Kaufman 1988). For the most part this has no impact on family tree structure or determination of descent, but in the occasional extreme case there is language mixture, in which one language can be seen as descended from two ancestral languages (examples include Mitchif, descended from both Cree and French, in southern Canada and the northern United States).

Languages long in contact can retain their discrete identities but come to resemble each other in sound structure, lexicon, and/or grammar. The resultant structural approximation is called *convergence,* and—especially when there is extensive and stable bilingualism or multilingualism—a set of languages showing convergence is called a *linguistic area* or *Sprachbund.* Well-known examples are the Balkan peninsula, the Caucasus, the Pacific Northwest of North America, Arnhem Land in Australia, and Mesoamerica (for major case studies, see Campbell et al 1986, Emeneau 1956, Heath 1978, Masica 1976, Ross 1996). A linguistic area is a population and even in some sense a speech community. Population formation through contact is not modeled with trees; it is sometimes described as requiring a *wave model* because the diffusion and adaptation involved in convergence are propagated in geographical space (see Trask 1996, 183ff).

Structural pools are likely to be dissipated former convergence sets and/or to contain ancient sisters whose inherited commonalities have faded away beyond the threshold of proof. At great time depths, it is impossible to distinguish between the two.

Within the geographical ranges of languages and even shallow language families, centers of political and economic importance are normally also centers of dialect or language prestige and epicenters from which linguistic innovations spread outwards. The spread involves adoption of the innovation and abandonment of the previous locution by speakers and local speech communities progressively farther from the center of innovation. Archaisms and archaic dialects then survive at the periphery of the area. (For these and other principles of linguistic geography, see Andersen 1988.) For instance, the spread of Inca empire from Cuzco also entailed the spread of a Quechuan dialect over a wider pre-Inca Quechua range. The modern result is a family tree in which one branch originated in the center of the Quechua range and the original range of the other extends both north and south (Mannheim 1985, summarizing earlier work). The same kind of geographical configuration usually accompanies the formation of a literary standard dialect. The same principle of innovating center and archaic periphery accounts for the essential geographical dynamics in all kinds of language spread and can be applied to the description of relations between languages, between families, and even between nonsisters or groupings of unknown phylogenetic status.

Genetic Density in Linguistics

Whether it is stocks, families, or languages that are counted, linguistic clades are not evenly distributed across the earth; their density shows highly significant skewings (Austerlitz 1980; Nichols 1990, 1992). The ratio of stocks to millions of square kilometers ranges from one to three for Africa and Eurasia, 10–20 for Australia and Central and South America, and over 100 for New Guinea (with its 60–80 stocks on less than a million square kilometers). The universal determinants of these differences are geographical and political-economical. Densities are higher in coastal regions, at lower latitudes, and in wetter and less seasonal climates; they are lower in continental interiors, at high latitudes, and in dryer and seasonal climates. Thus the densities of stocks or families are low, and the range of each is large, in such places as the central Eurasian steppe, the Eurasian and American arctic and subarctic, and the arid interior of Australia; they are high in (moist, tropical, mostly coastal) New Guinea, higher in California and Oregon than in Washington and British Columbia, and high in the Amazon basin. Densities are lower in complex societies, agricultural societies, and especially areas with a long history of empire; and they are higher in smaller and simpler societies. Hence they are higher in California than in Mesoamerica, and higher in Mesoamerica and the ancient Near East (both with young traditions of statehood and empire) than in the modern Near East or modern Europe. Accordingly, densities are lower where population density is higher and vice versa. Essentially, linguistic density is highest in areas where small societies can be more or less autonomous on small territories (Austerlitz 1980; Mace & Pagel 1995; Nichols 1990, 1992).

These principles apply equally well in areas of earliest human inhabitation (Africa), continents colonized early (Australia), continents colonized recently (the Americas), and formerly glaciated areas colonized recently (northern Europe, North America); to continents with no evident recent colonization (Australia) and those with evidence of multiple colonization continuing until recent millennia (New Guinea, the Americas). Thus, neither time settled nor number of colonizations has any appreciable effect on genetic density, which is determined entirely by geography, population density, and economy.

Stock density, especially as viewed over time, is the most visible attribute distinguishing two different kinds of language areas: *spread zones* and *accretion zones* (Nichols 1992, 1997c, 1998). An accretion zone (termed *residual zone* in previous works, but *residual* has an unrelated technical sense) is an area where genetic and structural diversity of languages are high and increase over time through immigration. Examples are the Caucasus, the Himalayas, the Ethiopian highlands and the northern Rift Valley, California, the Pacific Northwest of North America, Amazonia, northern Australia, and of course New Guinea. Languages appear to move into these areas more often than they move out of them. Kaufman (1990, p. 35) describes parts of South America as *sumps* or *invasion zones,* areas most of whose languages have originated elsewhere. This is a more precise notion than accretion zone: There are several different invasion zones within Amazonia, while the entirety of Amazonia can probably be described as a single accretion zone. Accretion zones generally contain representatives of major stocks in the vicinity as well as some languages with no outside kin. In the Caucasus, for instance, are found three indigenous stocks and representatives of two branches of Turkic and two branches of Indo-European.

A spread zone is an area of low density where a single language or family occupies a large range, and where diversity does not build up with immigration but is reduced by language shift and language spreading. A conspicuous spread zone is the grasslands of central Eurasia, in which, at roughly 2000-year intervals, four different spreads have carried different language families across the entire steppe and desert as well as into central Europe and Anatolia: Proto-Indo-European, Iranian, Turkic, Mongolian (discussed below). Another spread zone is central and southern Australia, in which the Pama-Nyungan quasi-stock has undergone several spreads to cover most of the continent (e.g. Evans & Jones 1997, McConvell 1996a,b). Another is northern Africa. Another is the Great Basin of the western United States, where the Numic branch of Uto-Aztecan spread from the Sierra slopes in the southwest of the range within the past two millennia (Bettinger & Baumhoff 1982, Lamb 1958, Madsen & Rhode 1994).

The dynamic of a spread zone is much like that of a dialect area. There is a locus or "center," rarely at the literal center of the range and usually at an edge, from which the language or family spreads. The locus of the various spreads on the Eurasian steppe is at the eastern edge of the range, in Central Asia; that of Pama-Nyungan is its northeastern corner; that of the Numic spread is in its southwestern corner. (These spreads are discussed below.) In a spread zone there is a general trajectory of spread: east and north in the Great Basin, west in the Eurasian steppe, south and west in Australia. At the periphery of a spread zone, remnant languages may survive from previous spreads. At the western periphery of the steppe, the Iranian language Ossetic survives from before the spread of Turkic. Possible remnants surviving from before the Numic spread are the languages of the pueblos to the east and the Maiduan and Washo languages of the Sierras to the north (Aikens 1994).

The fact that spreads are accompanied by extinction and that remnants survive, if at all, as isolates at the edges of spread zones, together with the general paucity of high-level branching structure in language families in accretion zones, explains why there are so few good candidates for quasi-stocks and so few genetic groupings of great age. Regular processes of extinction have turned the majority of stocks into isolates.

Hill (1996) proposes an anthropological dialectology that explains the different linguistic distributions of accretion and spread zones, and bunched vs extended language patterns more generally. The movement of linguistic variables of all kinds—phonemes, words, whole languages—across human populations depends on the relative dominance among speakers of two stances toward that variation: *localist* and *distributed.* The localist stance hinders the spread of variables, while the distributed stance favors them. "People with secure primary claims on essential resources are more likely to favor localist stances, while people who lack adequate primary claims and draw instead on a diverse range of secondary or indirect claims are more likely to favor distributed stances." Hence laissez-faire attitudes toward dialectal variation and weak language loyalty are favored by precarious economic circumstances.

LANGUAGE MOVEMENT AND SPREAD

Family Tree Structure and Language History

Certain family tree structures result from particular historical situations. Minimal branching—as in an isolate or stock-level shallow family—points to extinction of the rest of the family. Isolates and near-isolates are most common at edges of spreads where they represent remnants: in mountain highlands, where they are islands surrounded by languages of lowland families (e.g. Burushaski, an isolate of the Himalayas whose lower neighbors are Iranian); at coasts and continental peripheries (Basque; isolates Nivkh and Ainu in eastern Asia); between two spreads, where remnants can be trapped (Ket, the sole survivor of the Yeniseian family, trapped between Uralic to the west and Tungusic to the east); and, occasionally, as islands within spreads [Yukagir, a remnant of a former spread over much of eastern Siberia, now whittled down to small islands in northeastern Siberia as a result of Evenki and Yakut spreads; for the seventeenth-century range of Yukagir, see Levin & Potapov (1964, p. 5)].

Multiple branching at or near the root of a tree points to abrupt dispersal of the protolanguage in a large spread. The Indo-European family tree had up to a dozen major branches, all of which separated within about the first millennium of this 6000-year-old family's history (for the order of separation, see Taylor et al 1995, Warnow et al 1996). The initial diversification of the Austronesian tree may have been into three distinct Formosan branches plus Malayo-Polynesian (Pawley & Ross 1993, p. 435), pointing to rapid dispersal.

When a language family has dispersed gradually and in a more or less constant direction, its family tree assumes a distinctive, consistently left- or right-branching shape. Right and left have no theoretical place in phylogenetic trees, but for some families it is possible to exploit right and left and draw a tree that neatly projects onto a map of the daughter languages in real space. The Uralic family tree, for example, has an initial bifurcation into Finno-Ugric vs Samoyedic, and the root can be positioned over the region east of the Urals where Finno-Ugric and Samoyedic meet. Finno-Ugric bifurcates into Finnic and Ugric, and this node is positioned over the Urals where the two branches meet. Finnic bifurcates into Permian and others west (i.e. left) of the Urals; one or more branches split off farther west in the vicinity of the Volga; and finally the westernmost (leftmost) Baltic Finnic branch diversifies—into Finnish, Estonian, and others—near the Baltic coast. This is a west-branching (or left-branching) tree whose nodes and branches coincide remarkably well with the distribution of the daughter languages in real space. The westward branching is the result of consistent westward movement from a homeland in the vicinity of the Urals across northern Europe. A right-branching tree is that of the eastern Malayo-Polynesian branch of Austronesian, which when positioned over a map puts down daughter branches from progressively lower nodes as the seafaring early Austronesians moved eastward from island Southeast Asia through Melanesia and along the Solomon Island chain to the open ocean and thence out to Polynesia. For both the Uralic and the Austronesian trees, the distinctive aspect is the skewing, with the root of the tree close to one edge of the range on the map. This skewing is the result of a long-standing and consistent directionality of spread.

The projection of the root of the phylogenetic tree onto the ground in a map is known as the *center of gravity* in linguistic geography. The protohomeland of a language family is assumed to lie in the vicinity of its center of gravity (Diebold 1960, Dyen 1956, Sapir 1949), a principle that puts the homeland near the eastern edge of the range for Uralic, near the western edge for Austronesian, but in the center of the range for Slavic, which spread radially from central Europe. In spread zones, this principle works reliably only for the most recent spread. As soon as the next spread overtakes the center of gravity of the former spread, the apparent center of gravity shifts toward the far end of the range. Consider the Turkic family, whose modern center of gravity is in the western part of its range, near the middle Volga where Chuvash, the sole survivor of the Bulgar branch, meets Tatar of the other branch. The homeland, however, is known from historical sources to have been at the eastern edge of the Turkic range, near Mongolia. The Iranian family, which spread before Turkic, has two modern centers of gravity: one in the mountains south of Central Asia and one south of the Caucasus, both representing peripheral pile-ups and not the actual homeland, which was in the eastern steppe. Indo-European had early centers of gravity in central Europe, the Balkans, and western Anatolia, but these too are pile-ups and not the homeland. The modern Indo-European center of diversity is in the Balkan peninsula, and descendants of the earliest surviving offshoot, Italo-Celtic (Taylor et al 1995), are now found at the farthest periphery of the Indo-European range. [These examples from the Eurasian steppe are discussed in Nichols (1997c).] Thus family trees of early dispersants in continuing spread zones show inverted centers of gravity.

There are three known mechanisms of language spread: language shift (see Thomason & Kaufman 1988), demographic expansion (e.g. Bettinger & Baumhoff 1982, Madsen & Rhode 1994), and migration (Anthony 1990). There are probably no pure cases: Language shift is normally in response to the presence of at least a few influential immigrants; demographic expansion involves some absorption of previous population rather than extermination; and migration leads to language shift (either to or from the immigrants' language). The terms *language shift, demographic expansion,* and *migration* refer to the predominant contributor with no claim that it is exclusive. Almost all literature on language spreads assumes, at least implicitly, either demographic expansion or migration as basic mechanism, but in fact language shift is the most conservative assumption and should be the default assumption. There is no reason to believe that the mechanism of spread has any impact on the linguistic geography of the spread, but it has major implications for whether we expect to find linguistic substratum effects (which should occur with language shift but not with demographic expansion) and what degree of (biological) genetic admixture we expect to find as a consequence of spreading.

Geographical Factors

In addition to the large spread zones associated with high latitudes and arid interiors, other general tendencies of language movement can be discerned for particular geographical environments.

MOUNTAINS Midlatitude mountains in settled agricultural areas, in historical and protohistorical times, exhibit a standard trajectory of language spread whereby lowland languages spread uphill, mountaintops are islands where remnant languages survive before being absorbed, and the mechanism of spread is "vertical bilingualism," whereby highland villagers know lowland languages but not vice versa (for the Caucasus, see Wixman 1980; for Central Asia, see Èdel'man 1980). The causal mechanism is climate: At least since the advent of the Little Ice Age in the late middle ages (Grove 1988), highland economies have been precarious, whereas the lowlands, with their longer growing seasons, are prosperous and offer markets and winter employment for the essentially transhumant male population of the highlands. Prior to the global cooling, lowlands were dry and uplands moist and warm enough for agricultural security. When highlands are economically secure, they are loci of linguistic spreads, upland dialects spread downhill, and highland clans or polities extend to form islands in lowland outposts (which can then grow into centers of dialect spread when economic fortunes shift). This generalization is based on clan origin traditions and language and dialect isoglosses in the central Caucasus (Nichols 1997b) and on evidence of highland spread of Quechua (Stark 1985) and verticality in the prehistoric and protohistorical Andes (Stanish 1992). Verticality of political economy is standardly assumed for the Andes (for more on the "vertical archipelago" of John Murra, see Masuda et al 1985), but Stanish also finds evidence for ethnic identity, and this implies language spread.

COASTLINE Coastally adapted cultures can spread structural pools, linguistic areas, and even families far along coastlines. The various Austronesian languages of New Guinea are mostly coastal even after several millennia (see maps in Wurm & Hattori 1983). The Eskimoan spread across arctic North America to Greenland was initially entirely coastal and replaced the previous, also coastal, Dorset culture and language (Dumond 1984, Woodbury 1984). Austronesians and speakers of Eskimoan languages are coastally adapted peoples, and accordingly they have spread along coasts rather than inland. What is notable is the great extent of their coastal spreads. Earlier than these spreads is the extended spread from Beringia to South America, almost entirely along the Pacific coast, of the structural pool of languages identified by personal pronouns with first person *n* and second person *m* in addition to several other features (discussed above).

Coastal languages and language families often have discontinuous distributions. Sometimes these are due to known overwater migrations, as is the case with the various Austronesian enclaves of New Guinea. Intrusions of other languages have probably split some of the discontinuous coastal families of New Guinea (shown in Foley 1986, pp. 230–31; Wurm & Hattori 1983). The Yolngu branch of Pama-Nyungan (Australia) is isolated in northern Arnhem Land at some distance from the rest of the quasi-stock, whose distribution is otherwise monolithic (Blake 1988, Wurm & Hattori 1983), and the cause of this separation remains mysterious. Finally, accretion zones in the vicinity of coasts, like those of California and Oregon, contain languages that are geographically isolated and at some distance from their sisters; examples include the coastal Athabaskan languages of northern California and southern Oregon, distant from their relatives in Canada and Alaska. Movement of languages within accretion zones must be erratic overall and occasionally long-distance.

FOREST The eastern forests of North America and much of pre-medieval Europe fostered sizable spreads and relatively low language diversity, comparable to that of the prairie and steppe. In seasonal and temperate climates, then, forests seem to harbor spread zones just as grasslands do. For tropical and subtropical forest the picture is less clear. High diversity is found in the Amazon basin, New Guinea, and central America, but much lower diversity is found in Africa, Southeast Asia, and eastern Australia. Perhaps the lower diversities were associated with more sedentary societies and higher population.

Economic and Political Factors

As noted above, complex societies, and especially states and empires, favor spreads and produce considerable linguistic extinction. There is a recurrent opinion in the literature that in small societies with small populations, diffusion and other influence occur easily, even to the extent that convergence can obscure family boundaries. Austerlitz (1991) notes—chiefly with reference to the Eurasian arctic and subarctic—that the combination of small population, short average lifespan, and frequent slavery and exogamy would entail a great deal of systematic multilingualism and allow for rapid vocabulary transfer. Hill (1996), citing Miller (1970), mentions that in small desert populations with shifting group membership, children would have lacked stable peer groups of age mates, the principal forum in which the individual speaker's dialect identity crystallizes. Sorensen (1985) notes the effect of linguistic and tribal exogamy in the upper Amazon. Heath (1978) and Ross (1996) present detailed case studies of the consequences of systematic bilingualism or multilingualism in Australia and New Guinea, respectively. In Australia, the consequence is a great deal of diffusion, which brings about a rough convergence piece by piece. In New Guinea, the sociolinguistic situation is different; an interethnic language influences an ethnic-specific language in vocabulary and grammar, and this gives rise to profound convergence. Hinton (1991), citing

Florence Shipek, describes systematic intertribal marriage in southern California with consequent substratal effect and language shift.

Does the initial spread of agriculture or herding result in the spread of the language of first farmers or pastoralists? Geneticists often maintain that it does and that the spread is demographic (Ammerman & Cavalli-Sforza 1984, Cavalli-Sforza et al 1994). In some cases, this may have occurred. The spread of Austronesian is widely associated with the spread of rice agriculture (Pawley & Ross 1993, pp. 442–43). However, in Mesopotamia, western Eurasia generally, and eastern North America, there is no evidence for associating any extant stock or quasi-stock with the spread of agriculture. In Mesoamerica, the likely candidate for original agriculturalists is the Mixe-Zoque stock, notable for its lack of spread (Hill 1996, Justeson & Kaufman 1993).

LINGUISTIC POPULATION HISTORIES FROM SELECTED CONTINENTS

This section is a brief survey of four continents to show how a consistent stock-based genetic classification affects statements of language distribution and language density, and how spread zones have affected language distributions. Major language spreads that have often been seen as demographic expansions or biological diffusions prove to be very ordinary cases of language shift in long-standing spread zones.

Africa

Languages of Africa are conventionally divided into four groupings, from south to north: click languages, Niger-Kordofanian, Nilo-Saharan, and Afro-asiatic [Bender (1989), Heine et al (1981), all with maps]. (In addition, on Madagascar there is Malagasy, a Western Malayo-Polynesian language with close kin in Borneo.) These four are not comparable and are not all genetic groups. The click languages include the Khoisan family in southern Africa and two or three isolates—Hadza and Sandawe in Kenya, possibly Kwadi in the southwest—and constitute a structural pool. Niger-Kordofanian is a quasi-stock, defined by the gender prefixes described above, and consists of the outlier Kordofanian family in the eastern Sahara and several west African stocks. The center of gravity for the entire quasi-stock is in west Africa, and that of the widespread Bantu family and its superordinates is in the northwest of its range (for Bantu, the vicinity of Cameroon). Three separate spreads have carried the Kordofanian outlier to the northeast, the Adamawa branch eastward in the north of the Congo basin, and the Bantu family east and south to cover most of the southern half of Africa. Thus, the distribution and prehistory of Niger-

Kordofanian suggest a long-standing epicenter of spread in west Africa, with spreads through the forest and well to the south that have obliterated any previous languages other than Khoisan in the far south. The Bantu spread is associated with the spread of the Iron Age in the east (Bouquiaux 1980), which may explain its far southern reach, but otherwise it is an unexceptional instantiation of the standing eastward language trajectory through spread-producing geography (interior forest and grassland). Its mechanism was neither migration nor expansion (Shaw et al 1993) but language shift, absorbing the results of previous spreads.

The eastern highlands of Ethiopia and Kenya and the central and northern Rift Valley are an accretion zone into which the eastern Bantu frontier has advanced and where isolate click languages, several Afroasiatic stocks, and several Nilo-Saharan stocks are found. Bender (1983) has found two other Ethiopian languages that may well be isolates. Nilo-Saharan is probably not a genetic grouping and not even a structural pool but is simply a residual grouping. A core set of its stocks identified by Bender (1983)—Nilotic, Nubian, Central Sudanic, Kunama-Ilit, Koman-Gumuz, and Kadugli—may constitute a structural pool or even a quasi-stock. The other Nilo-Saharan groups are probably unrelated families and isolates found across the Sahel and central to eastern Sahara.

The Afroasiatic quasi-stock (Diakonoff 1988, Greenberg 1963, Heine et al 1981, Newman 1980) stretches across north Africa and the Near East and into Ethiopia and Somalia. The center of gravity of Afroasiatic is in the vicinity of the Ethiopian highlands (see Blench 1993), but because the highlands are an accretion zone, the Sahara is a continuing spread zone, and Afroasiatic is ancient and an early dispersant, it can be assumed that Afroasiatic exhibits a center-of-gravity inversion with a peripheral pile-up in Ethiopia. An origin on the central to western Sahara is most likely on geographical grounds. The northwestern Sahara is where domestication of cattle and sheep is first attested (Clutton-Brock 1993; see also Phillipson 1993), and this is a logical Afroasiatic homeland.

In sum, Africa may have as many as 30 stocks and as few as 17 combined stocks and quasi-stocks. Spreads mostly from west to east have peripheralized most of the genetic and structural diversity in the eastern highlands and near the Rift Valley.

Western Eurasia

Most of this half-continent has been linguistically populated by successive rapid westward spreads originating in the steppe to the east and occurring at roughly 2000-year intervals: in reverse chronological order, Mongolian, Turkic, Iranian, Indo-European (Nichols 1992, 1997c, 1988; for Indo-European, see also Anthony 1991, 1995; Mallory 1989). The northern forests in the same time frame have undergone only one spread, that of Uralic (bringing Estonian, Finnish, Saami, and others to northeastern Europe). The spreads mostly involved language shift, and each spread on the steppe absorbed the previous steppe languages, leaving remnants variously in central Europe, the Caucasus, and the Central Asian mountains. The Caucasus is an accretion zone with three unrelated indigenous stocks, of which Northeast Caucasian and Kartvelian have at least some structural affinities with ancient Mesopotamia and southwestern Asia, while Northwest Caucasian is radically unlike any other Eurasian stock in most respects. In far western Europe, Basque is an isolate, though with such southwest Asian affinities of structure that it could plausibly be a barely pre-Indo-European immigrant from the steppe. It is possible that no language presently found in or near Europe continues a pre-Neolithic indigene or a language of first agriculturalists.

Australia–New Guinea

The entire Australia–New Guinea landmass including the large Melanesian islands had been settled by about 40,000 BP and perhaps much earlier (Roberts & Jones 1994, White 1996). Colonization was by coastally adapted seafarers from insular southeast Asia and is believed to have been intentional and repeated. Overseas colonization has continued, and the most recent episodes are various coastal settlements of Austronesian-speaking peoples which have occurred at various times in the past four millennia (Pawley & Ross 1993, Spriggs 1995).

The entire south and east of modern Australia is a vast spread zone covered recently by languages of the Pama-Nyungan quasi-stock (Dixon 1980). Pama-Nyungan originated in the northeast of its range and spread by a combination of language shift and migration perhaps within the past six millennia (Evans & Jones 1997, McConvell 1996a,b). Northeastern Australia (southern Cape York), the likely Pama-Nyungan homeland, is a long-standing center of technological innovation (Morwood & Hobbs 1995), an area of deep divergence within Pama-Nyungan, and close to the Tangkic family, which represents a likely first sister to Pama-Nyungan (Evans 1995). The languages spoken by the first settlers of southern Australia have probably not survived, as these areas have been overrun probably more than once by spreads like that of Pama-Nyungan. The spread of Pama-Nyungan from the northeast through the drier regions does not explain why tropical Cape York and the forested eastern coastal strip are entirely Pama-Nyungan-speaking. On geographical grounds, genetic diversity should thrive in these areas. The lack of diversity suggests a dense and sedentary population.

The wetter and coastal north and northwest of Australia form an accretion zone in which a dozen or more stocks are found. These languages form a structural pool with those of coastal and especially northern New Guinea, which suggests an origin in relatively recent colonizations (Nichols 1997a). Hence it cannot be assumed that this area contains descendants of the earliest colonizers, though it contains the earliest archeological sites. Pama-Nyungan forms a structural pool with the languages of southeastern and highland New Guinea (for the New Guinea side, see Ross 1995), and on geographical and other evidence this pool is older than the coastal one. The distinction between coastal and interior strata goes back, on geographical grounds, to before the end of glaciation (Nichols 1997a).

Diffusion, multilingualism, convergence, and continent-wide areality occur on an unprecedented scale in Australia (e.g. Dixon 1980, Heath 1978). The degree of convergence even suggests genetic relatedness between Pama-Nyungan and some or all of the northern groups, at least to some investigators (Blake 1988, Dixon 1980, Evans 1988). Dixon offers potentially valid genetic markers in noun and verb paradigms, but their evidentiary value will depend on their distribution as a set among the various stocks and families. Sorting out convergence and inheritance at great time depths will require bottom-up reconstruction within genetic groupings.

The non-Austronesian languages of New Guinea and Melanesia present the greatest genetic density on earth, falling into perhaps 80 families and isolates (Foley 1986; M Donohue, unpublished surveys), the majority of which are likely to be independent stocks. Most of this diversity falls in the coastal stratum that disproportionately reflects the contributions of recent colonizations, and the Austronesian languages at least of the New Guinea mainland are fairly typical representatives of this stratum. Bilingualism in small communities has produced some cases of extreme convergence (Ross 1996), and continent-wide prevalence of certain structural features such as SOV word order (Foley 1986) also suggests convergence and certainly indicates contact-based change.

Thus in Australia–New Guinea, coastal languages have mingled and generally stayed near the coast, whereas interior language families have been decimated by spreads.

The Americas

There are about 150 separate stocks in the New World (Campbell & Mithun 1979; Goddard 1997; Kaufman 1990; Suárez 1974, 1983). Two well-described probable quasi-stocks [Hokan (Kaufman 1988), a grouping of up to a dozen stocks of California and Mexico; and Penutian, half a dozen stocks of

California and the Oregon plateau, and perhaps others] and a few other possible quasi-stocks reduce this diversity only slightly. The North American diversity is disproportionately clustered near the west coast, with several large eastward spreads extending from the coastal or intermontane west (Jacobsen 1989). The South American diversity is concentrated in the Amazon basin.

The most recent and most clearly traceable of the North American interior spreads is that of the Numic branch of Uto-Aztecan, which has spread from the Sierra foothills in the southwestern part of the Numic range to cover most of the Great Basin within the past two millennia. The anthropological literature generally assumes the spread to have been demographic expansion (Bettinger & Baumhoff 1982, Madsen & Rhode 1994), but the ecological context of the foothills-desert interface and the presumable sociolinguistics of such an interface indicate language shift (Hill 1996).

For the other two large spreads, the epicenters are also in the west, but their latitudes are uncertain. The Na-Dene stock consists of coastal Eyak and Tlingit in southern Alaska and northern British Columbia, and the widespread but young Athabaskan language family which stretches across interior Alaska and northern Canada and has coastal outliers in Oregon and northern California as well as a sizable interior outlier in Navajo and Apache of the southwestern United States. The center of gravity is on the southern Alaska coast where Tlingit, Eyak, and Alaska Athabaskan meet. Leer (1991) and Kari (1996) proposed a more southern origin, and Leer has presented evidence that unknown, now extinct coastal languages to the north were absorbed by the later Na-Dene spread. Ever since Sapir (1949), Na-Dene—the northernmost stock save Eskimo-Aleut—has been interpreted as the next-to-last entrant to the Americas, but on the evidence of linguistic geography, Na-Dene is not the latest pre-Eskimo entrant but merely the most recent subarctic spread. In any event, a shallow family like Athabaskan with a wide range in a spread zone can be assumed to have replaced earlier languages.

A still earlier spread is that of the Algic stock, represented by Yurok and Wiyot of the northern California coast (which are either independent Algic branches or a very deep branch) and the far-flung Algonquian family stretching from the western plains (e.g. Blackfoot) to the Atlantic coast (Micmac, Delaware). The center of gravity of Algonquian is in its western range (Goddard 1994), and that of Algic farther west. The Oregon-Washington plateau is a possible locus of spread, or at least a point from which Yurok and Wiyot could have moved coastward. On geographical grounds, Algic is near the Numic periphery on three sides (east, north, west) and could be the remnant of a pre-Numic spread in the Great Basin.

Whistler (1977) shows that the Penutian language families of California entered individually from the vicinity of the Oregon plateau and spread southward. The California and Oregon coastal Athabaskan languages have spread well to the south of any plausible homeland. These reconstructable histories suggest that language movement along the coast is erratic but southward overall, while the interior spreads move east and north. There is no geographical stratification that can identify the latest surviving pre-Eskimoan entrant, though it is likely to be coastal. If Proto-Athabaskan, dispersing only 2000 years ago, has daughters as far south as northern California and the Mexican border, there is no reason to assume that the necessarily much earlier pre-Eskimo entrant has remained in the north. The Pacific Rim structural pool is the result of a recent colonization episode, and its continuity essentially all around the Pacific indicates rapid spread.

CONCLUSION

Languages can move rather rapidly over human populations, and in most cases for which we have evidence, modern languages are spoken at some distance from where their ancestors originated. The combination of extinction in spread zones and convergence in accretion zones means that simple phylogenetic descent is insufficient for tracing the origin and dispersal of the world's languages and peoples. A natural task for joint linguistic-archeological investigation is tracing ancient spread zones, trajectories of movement, and densities of languages and human populations, as these have shaped modern ethnic distributions.

ACKNOWLEDGMENTS

I thank Victor Golla, James Kari, and Richard Rhodes for some bibliographical and factual references.

Literature Cited

Aikens CM. 1994. Adaptive strategies and environmental change in the Great Basin and its peripheries as determinants in the migrations of Numic-speaking peoples. See Madsen & Rhode 1994, pp. 35–43

Akazawa T, Szathmáry EJE, eds. 1996. *Prehistoric Mongoloid Dispersals.* Oxford: Oxford Univ. Press

Allen J, O'Connell JF, eds. 1995. *Transitions: Pleistocene to Holocene in Australia and Papua New Guinea. Antiquity* 69:Special No. 26

Ammerman AJ, Cavalli-Sforza LL. 1984. *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton, NJ: Princeton Univ. Press

Andersen H. 1988. Center and periphery: adoption, diffusion, and spread. In *Historical Dialectology: Regional and Social,* ed. J Fisiak, pp. 39–83. Berlin: Mouton de Gruyter

Anthony DW. 1990. Migration in archeology: the baby and the bathwater. *Am. Anthropol.* 92:896–914

Anthony DW. 1991. The archeology of Indo-European origins. *J. Indo-Eur. Stud.* 19(3/4):193–222

Anthony DW. 1995. Horse, wagon, and chariot: Indo-European languages and archaeology. *Antiquity* 69:554–65

Austerlitz R. 1980. Language-family density in North America and Eurasia. *Ural-Altai. Jahrb.* 52:1–10

Austerlitz R. 1991. Alternatives in long-range comparison. See Lamb & Mitchell 1991, pp. 353–64

Bender ML. 1969. Chance CVC correspondences in unrelated languages. *Language* 45:519–36

Bender ML, ed. 1983. *Nilo-Saharan Language Studies.* East Lansing: Mich. State Univ. Press

Bender ML. 1989. Nilo-Saharan pronouns/demonstratives. In *Topics in Nilo-Saharan Linguistics,* ed. ML Bender, pp. 1–34. Hamburg: Buske

Bendor-Samuel J, ed. 1989. *The Niger-Congo Languages.* Lanham: Univ. Press Am.

Bettinger RL, Baumhoff MA. 1982. The Numic spread: Great Basin cultures in competition. *Am. Antiq.* 47(3):485–503

Blake BJ. 1988. Redefining Pama-Nyungan: toward the prehistory of Australian languages. *Aborig. Linguist.* 1:1–90

Blench RM. 1993. Recent developments in African language classification and their implications for prehistory. See Shaw et al 1993, pp. 126–38

Bouquiaux L, ed. 1980. *L'Expansion bantoue. Actes Colloq. Int. CNRS, Viviers, France, 4–16 Avril 1977.* Paris: Soc. Études Linguist. Anthropol. Fr.

Campbell L, Kaufman T, Smith-Stark TC. 1986. Mesoamerica as a linguistic area. *Language* 62:530–70

Campbell L, Mithun M, eds. 1979. *The Languages of Native America.* Austin: Univ. Tex. Press

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton Univ. Press

Clutton-Brock J. 1993. The spread of domestic animals in Africa. See Shaw et al 1993, pp. 61–70

Crowley T. 1992. *An Introduction to Historical Linguistics.* Auckland: Oxford Univ. Press. 2nd ed.

Diakonoff IM. 1988. *Afrasian Languages.* Moscow: Nauka

Diebold A. 1960. Determining the centers of dispersal of language groups. *Int. J. Am. Linguist.* 26:1–10

Dixon RMW. 1980. *The Languages of Australia.* Cambridge: Cambridge Univ. Press

Dumond DE. 1984. Prehistory: summary. See Sturtevant 1984, pp. 72–79

Durie M, Ross MD, eds. 1996. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change.* New York/Oxford: Oxford Univ. Press

Dyen I. 1956. Language distribution and migration theory. *Language* 32:611–26

Èdel'man DzhI. 1980. K substratnomu naslediju central'noaziatskogo jazykovogo sojuza. *Vopr. jazykozn.* 1980(5):21–32

Ehret C. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary.* Univ. Calif. Publ. Linguist. 126. Berkeley: Univ. Calif. Press

Embleton S. 1986. *Statistics in Historical Linguistics.* Bochum: Brockmeyer

Embleton S. 1991. Mathematical methods of genetic classification. See Lamb & Mitchell 1991, pp. 365–88

Emeneau MB. 1956. India as a linguistic area. *Language* 32:3–16. Reprinted in *Language and Linguistic Area,* ed. MB Emeneau, pp. 105–25. Stanford: Stanford Univ. Press

Evans ND. 1988. Arguments for Pama-Nyungan as a genetic subgroup, with particular reference to initial laminalization. *Aborig. Linguist.* 1:91–220

Evans ND. 1995. *A Grammar of Kayardild: With Historical-Comparative Notes on Tangkic.* Berlin: Mouton de Gruyter

Evans ND, Jones R. 1997. The cradle of the Pama-Nyungans. See McConvell & Evans 1997

Foley WA. 1986. *The Papuan Languages of New Guinea.* Cambridge: Cambridge Univ. Press

Goddard I. 1975. Algonquin, Wiyot, and Yurok: proving a distant genetic relationship. *Linguistics and Anthropology: In Honor of C. F. Voegelin,* ed. MD Kinkade, KL Hale, O Werner, pp. 249–62. Lisse: de Ridder

Goddard I. 1994. The west-to-east cline in Algonquian dialectology. *Actes Congr Algonq., 25th,* ed. W Cowan, pp. 187–211. Carleton, Ott: Carleton Univ. Press

Goddard I. 1997. *Handbook of North American Indians,* Vol. 17, *Language.* Washington, DC: Smithson. Inst.

Greenberg JH. 1960. An Afro-Asiatic pattern of gender and number agreement. *J. Am. Orient. Soc.* 80(4):317–21
Greenberg JH. 1963. *The Languages of Africa.* Bloomington: Indiana Univ. Press
Grove AT. 1993. Africa's climate in the Holocene. See Shaw et al 1993, pp. 32–42
Grove J. 1988. *The Little Ice Age.* London: Methuen
Haas MR. 1958. Algonkian-Ritwan: the end of a controversy. *Int. J. Am. Linguist.* 24: 159–73
Haas MR. 1969. *The Prehistory of Languages.* The Hague: Mouton
Heath J. 1978. *Linguistic Diffusion in Arnhem Land.* Canberra: Aust. Inst. Aborig. Stud.
Heine B, Schadeberg TC, Wolff E, eds. 1981. *Die Sprachen Afrikas.* Hamburg: Buske
Hill JH. 1996. Languages on the land: toward an anthropological dialectology. David Skomp Lect., Dep. Anthropol., Indiana Univ., Mar. 21. Bloomington
Hinton L. 1991. Takic and Yuman: a study in phonological convergence. *Int. J. Am. Linguist.* 57(2):133–57
Hock HH. 1991. *Principles of Historical Linguistics.* Berlin: Mouton de Gruyter
Hock HH, Joseph BD. 1996. *Language History, Language Change, and Language Relationship. An Introduction to Historical and Comparative Linguistics.* Berlin: Mouton de Gruyter
Hoenigswald H. 1960. *Language Change and Linguistic Reconstruction.* Chicago: Univ. Chicago Press
Jacobsen WH Jr. 1989. The Pacific orientation of western North American languages. Presented at Circum-Pac. Prehist. Conf., 1st, Seattle
Justeson JS, Kaufman T. 1993. A decipherment of Epi-Olmec hieroglyphic writing. *Science* 255:606–8
Kari J. 1996. A preliminary view of hydronymic districts in Northern Athabaskan prehistory. *Names* 44(4):253–71
Kaufman T. 1988. A research program for reconstructing Proto-Hokan: first gropings. *Pap. Hokan-Penutian Lang. Workshop, 1988,* ed. S Delancey, pp. 50–168. Eugene: Dep. Linguist., Univ. Oreg.
Kaufman T. 1990. Language history in South America: what we know and how to know more. In *Amazonian Linguistics: Studies in Lowland South American Languages,* ed. DL Payne, pp. 13–73. Austin: Univ. Tex. Press
Klein HEM, Stark LR, eds. 1985. *South American Indian Languages: Retrospect and Prospect.* Austin: Univ. Tex. Press
Ladefoged P, Maddieson I. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell
Lamb SM. 1958. Linguistic prehistory in the Great Basin. *Int. J. Am. Linguist.* 24(2): 95–100
Lamb SM. 1959. Some proposals for linguistic taxonomy. *Am. Linguist.* 1(2):33–49
Lamb SM, Mitchell ED, eds. 1991. *Sprung from Some Common Source: Investigations into the Prehistory of Languages.* Stanford: Stanford Univ. Press
Leer J. 1991. Evidence for a northern Northwest Coast language area: promiscuous number marking and periphrastic posessive constructions in Haida, Eyak, and Aleut. *Int. J. Am. Linguist.* 57(2):158– 93
Levin MG, Potapov LP, eds. 1964. *The Peoples of Siberia.* Chicago: Univ. Chicago Press
Mace R, Pagel M. 1995. A latitudinal gradient in the density of human languages in North America. *Proc. R. Soc. London Ser. B* 261: 117–21
Madsen D, Rhode D, eds. 1994. *Across the West: Human Population Movement and the Expansion of the Numa.* Salt Lake City: Univ. Utah Press
Mallory JP. 1989. *In Search of the Indo-Europeans: Language, Archaeology, and Myth.* London: Thames & Hudson
Mannheim B. 1985. Southern Peruvian Quechua. See Klein & Stark 1985, pp. 481–515
Masica CP. 1976. *Defining a Linguistic Area: South Asia.* Chicago: Univ. Chicago Press
Masuda S, Shimada I, Morris C, eds. 1985. *Andean Ecology and Civilization.* Tokyo: Univ. Tokyo Press
McConvell P. 1996a. Backtracking to Babel: the chronology of Pama-Nyungan expansion in Australia. *Archaeol. Ocean.* 31: 125–44
McConvell P. 1996b. *Pama-Nyungan expansion: popular movement or movement of people?* Presented at Annu. Meet. Am. Anthropol. Assoc., 95th, San Francisco
McConvell P, Evans ND, eds. 1997. *Archaeology and Linguistics: Global Perspectives on Ancient Australia.* Melbourne: Oxford Univ. Press. In press
McMahon AM. 1994. *Understanding Language Change.* Cambridge: Cambridge Univ. Press
Miller W. 1970. Western Shoshoni dialects. In *Languages and Cultures of Western North America,* ed. E Swanson, pp. 17–36. Pocatello: Idaho State Univ. Press
Morwood MJ, Hobbs DR. 1995. Themes in the prehistory of tropical Australia. See Allen & O'Connell 1995, pp. 747–68

Moseley C, Asher RE, eds. 1994. *Atlas of the World's Languages.* London: Routledge
Newman P. 1980. *The Classification of Chadic within Afroasiatic.* Leiden: Universitaire Pers
Nichols J. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66(3):475–521
Nichols J. 1992. *Linguistic Diversity in Space and Time.* Chicago: Univ. Chicago Press
Nichols J. 1995. Shaped by some common contingency: genetic and historical markers. Workshop Math. Methods Comp. Linguist., Inst. Res. Cogn. Sci., Univ. Pa.
Nichols J. 1996. The comparative method as heuristic. See Durie & Ross 1996, pp. 39–71
Nichols J. 1997a. Sprung from two common sources: Sahul as a linguistic area. See McConvell & Evans 1997
Nichols J. 1997b. Chechen phonology. In *Phonologies of Asia and Africa,* ed. AS Kaye, P Daniels, pp. 941–71. Bloomington, Ind: Eisenbrauns
Nichols J. 1997c. The epicenter of the Indo-European linguistic spread. In *Archaeology and Language 1: Theoretical and Methodological Orientations,* ed. RM Blench, M Spriggs, pp. 122–48. London: Routledge
Nichols J. 1998. The Eurasian spread zone and the Indo-European dispersal. In *Archaeology and Language 2,* ed. RM Blench, M Spriggs. London: Routledge. In press
Nichols J, Peterson DA. 1996. The Amerind personal pronouns. *Language* 72(2): 336–71
Orel V, Stolbova O. 1995. *Hamito-Semitic Etymological Dictionary.* Leiden: Brill
Oswalt RL. 1991. A method for assessing distant linguistic relationships. See Lamb & Mitchell 1991, pp. 389–404
Pawley A, Ross MD. 1993. Austronesian historical linguistics and culture history. *Annu. Rev. Anthropol.* 22:425–59
Phillipson DW. 1993. The antiquity of cultivation and herding in Ethiopia. See Shaw et al 1993, pp. 344–57
Ringe DA Jr. 1992. On calculating the factor of chance in language comparison. *Trans. Am. Philos. Soc.* 82:1
Ringe DA Jr. 1995a. 'Nostratic' and the factor of chance. *Diachronica* 12(1):55–74
Ringe DA Jr. 1995b. The "Mana" languages and the three-language problem. *Ocean. Linguist.* 34(1):99–122
Ringe DA Jr. 1996. The mathematics of "Amerind." *Diachronica* 13:1.135–54
Ringe DA Jr. 1997. A probabilistic evaluation of Indo-Uralic. In *Nostratic: Sifting the Evidence,* ed. B Joseph, J Salmons. Philadelphia: Benjamins. In press
Roberts RG, Jones R. 1994. Luminescence dating of sediments: new light on the human colonisation of Australia. *Austr. Aborig. Stud.* 1994(2):2–17
Ross MD. 1995. The great Papuan pronoun hunt: recalibrating our sights. In *Tales from a Concave World: Liber Amicorum Bert Voorhoeve,* ed. C Baak, M Bakker, D van der Meij, pp. 139–68. Dep. Lang. Cult. Southeast Asia Oceania, Leiden Univ.
Ross MD. 1996. Contact-induced change and the comparative method: cases from Papua new Guinea. See Durie & Ross 1996, pp. 180–217
Sapir E. 1913. Wiyot and Yurok, Algonkin languages of California. *Am. Anthropol.* 15:617–46
Sapir E. 1949. (1916). Time perspective in aboriginal American culture: a study in method. In *Selected Writings of Edward Sapir,* ed. DG Mandelbaum, pp. 389–467. Berkeley/Los Angeles: Univ. Calif. Press
Shaw T, Sinclair P, Andah B, Okpoko A, eds. 1993. *The Archaeology of Africa: Food, Metals, and Towns.* London: Routledge
Sorensen AP. 1985. An emerging Tukanoan linguistic regionality: policy pressures. See Klein & Stark 1985, pp. 140–56
Spriggs MJT. 1995. The Lapita culture and Austronesian prehistory in Oceania. In *The Austronesians: Historical and Comparative Perspectives,* ed. P Bellwood, J Fox, D Tryon, pp. 112–33. Dep. Anthropol., Res. Sch. Pac. Asian Stud., Aust. Natl. Univ.
Stanish C. 1992. *Ancient Andean Political Economy.* Austin: Univ. Tex. Press
Stark L. 1985. Ecuadorian highland Quechua: history and current status. See Klein & Stark 1985, pp. 443–80
Sturtevant WC, ed. 1984. *Handbook of North American Indians,* Vol. 5, *Arctic.* Washington, DC: Smithson. Inst.
Suárez JA. 1974. South American Indian languages. *Encycl. Br.*
Suárez JA. 1983. *The Mesoamerican Indian Languages.* Cambridge: Cambridge Univ. Press
Taylor A, Warnow T, Ringe DA Jr. 1995. Character-based reconstruction of a linguistic cladogram. *Proc. Int. Conf. Hist. Linguist.,* 12th, Manchester
Thomason S, Kaufman T. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley/Los Angeles: Univ. Calif. Press
Tischler J. 1973. *Glottochronologie und Lexikostatistik. Innsbrucker Beiträge zur*

Sprachwissenschaft, Vol. 11. Innsbruck: Kowatsch

Trask RL. 1996. *Historical Linguistics.* London: Arnold

Voegelin CF, Voegelin FM. 1964–1966. Languages of the world. *Anthropol. Linguist.* 6:3–9; 7:1–2,4,7–9; 8:2–3,6–7

Warnow T, Ringe DA Jr, Taylor A. 1996. Reconstructing the evolutionary history of natural languages. *Proc. ACM-SIAM Symp. Discrete Algorithms*

Watkins C. 1985. *The American Heritage Dictionary of Indo-European Roots.* Boston: Houghton Mifflin

Whistler KW. 1977. Wintun prehistory: an interpretation based on linguistic reconstruction of plant and animal nomenclature. *Proc. Annu. Meet. Berkeley Ling. Soc., 3rd,* ed. KW Whistler, RD Van Valin Jr, C Chiarello, JJ Jaeger, M Petruck, et al, pp. 157–74. Berkeley: Berkeley Linguist. Soc.

White JP. 1996. Paleolithic colonization in Sahul land. See Akazawa & Szathmáry 1996, pp. 303–8

Wixman R. 1980. *Language Aspects of Ethnic Patterns and Processes in the North Caucasus.* Univ. Chicago Dep. Geogr. Res. Pap. No. 191. Chicago: Univ. Chicago Press

Woodbury AC. 1984. Eskimo and Aleut languages. See Sturtevant 1984, pp. 49–63

Wurm SA, Hattori S, eds. 1983. *Language Atlas of the Pacific Area.* (Pac. Linguist. C-86). Canberra: Aust. Acad. Hum. Jpn. Acad.