

At the Edge of Knowability: Towards a Prehistory of Languages

Colin Renfrew

The issue of 'knowability' in relation to the origins and distribution of the language families of the world is addressed, and recent advances in historical linguistics and molecular genetics reviewed. While the much-debated problem of the validity of the concept of the language 'macrofamily' cannot yet be resolved, it is argued that a time depth for the origins of language families greater than the conventional received figure of c. 6000 years may in some cases be appropriate, allowing the possibility of a correlation between language dispersals and demographic processes following the end of the Pleistocene period. The effects of these processes may still be visible in the linguistic 'spread zones', here seen as often the result of farming dispersals, contrasting with the linguistic 'mosaic zones' whose early origins may sometimes go back to initial colonization episodes during the late Pleistocene period. If further work in historical linguistics as well as in archaeology and molecular genetics upholds these correlations a 'new synthesis', whose outlines may already be discerned, is likely to emerge. This would have important consequences for prehistoric archaeology, and would be of interest also to historical linguists and molecular geneticists. If, however, the proposed recognition of such patterning proves illusory the prospects for 'knowability' appear to be less favourable.

This article, based upon the eleventh McDonald Lecture,¹ is about linguistic diversity and the origins of the 6500 or so languages spoken in the world today. These origins may be traced back in some cases over the past 10,000 or so years. Controversial claims have been made that some features may be traced back even further. But the more general question of the origins of language itself, as a typically human capacity, will not be discussed, although it is one of the most fascinating issues in contemporary archaeology (Mellars 1998; Pinker 1998). There is a general consensus (Noble & Davidson 1996) that a fully modern language capacity is a feature of our species *Homo sapiens sapiens*, and that this is likely therefore to have been the case of our *sapiens* ancestors of more than 40,000 years ago, and I do not propose to consider greater time depths than that. The title of this article is drawn in part from the current interests of the Alfred P. Sloan Foundation of New York on the whole issue of *knowability*. The question is not so much about the extent of our current knowledge,

but about the limits to knowledge and the ways by which we can assess what is in principle knowable, given favourable conditions of research, and what is in principle unknowable, being beyond the reach of human investigation. Some of the research described here has in part been funded by a generous grant from the Sloan Foundation applied to the project 'The Prehistory of Languages' conducted at the McDonald Institute.

My own concern for this theme dates back to my first research in the Aegean and the realization (Renfrew 1964) that earlier inferences about the chronology of the prehistoric place names of the Aegean (Haley & Blegen 1928) were invalidated by the discovery of Neolithic settlements on the Cycladic Islands (Evans & Renfrew 1968). This line of reasoning, applied to the origins of the Greek language, led to a growing scepticism about the current consensus on the origins of the Indo-European languages of Europe (Childe 1926; Gimbutas 1973) that their advent was due to the activities of warlike nomad pastoralists

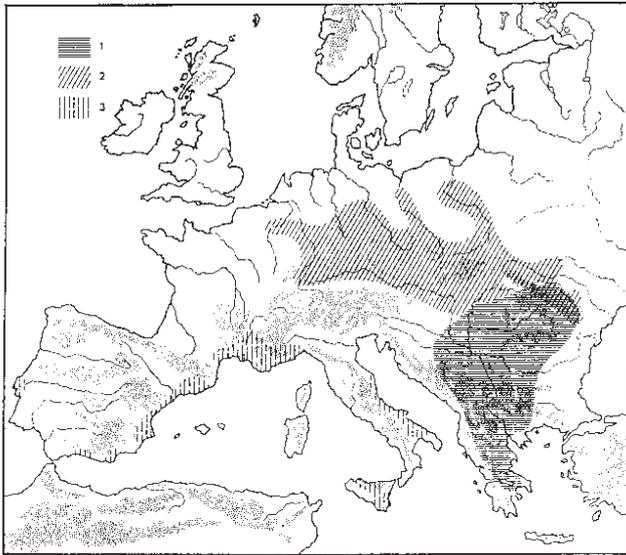


Figure 1. The early spread of agricultural communities in Europe, as summarized by Stuart Piggott in 1965. 1) Proto-Sesklo and Starčevo cultures; 2) Linear Pottery culture; 3) Impressed pottery culture. It is suggested below that the early spread of Proto-Indo-European speech accompanied the dispersal of farming in these areas. (From Piggott 1965, 57.)

from the Pontic steppes. It invited instead the formulation of an alternative theory.

The quest for some radical change in European prehistory with more convincing explanatory power than was offered by the alleged coming of the mounted warrior nomads was influenced by the clear picture on farming origins outlined for instance by Piggott (1965, 57, fig. 26) as seen in Figure 1. This was supported by the striking patterning for the chronology of the dispersal of farming in Europe (seen in Fig. 2) established on the basis of radiocarbon determinations by Grahame Clark (1965, 46). I therefore proposed (Renfrew 1973a) that the spread of farming to Europe from Anatolia was the principal agency responsible for the arrival and dispersal of Proto-Indo-European speech. If this proposal were accepted it would follow that the 'homeland' for the original Proto-Indo-European language would be situated somewhere in south-central Anatolia (where James Mellaart had been making striking Early Neolithic discoveries) and that the arrival of Proto-Indo-European speech into Europe from Anatolia could be dated using the available chronology for the coming of farming to around 7000 BC.

In developing the farming dispersal thesis for Indo-European into a more comprehensive statement in *Archaeology and Language* (Renfrew 1987), I uti-

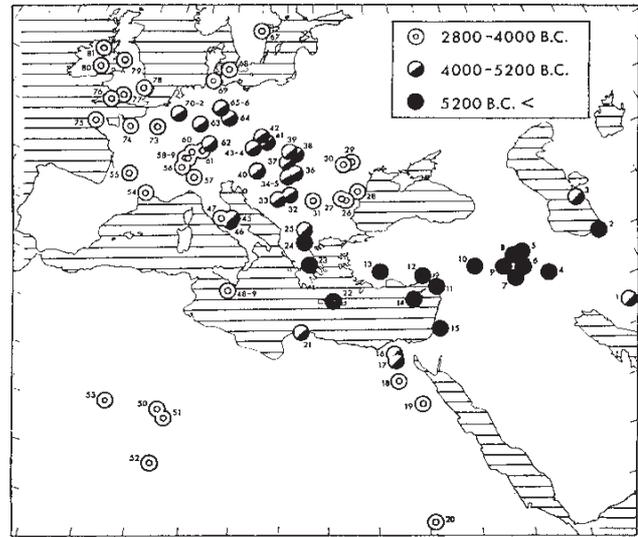


Figure 2. Radiocarbon chronology (uncalibrated) as proposed by Grahame Clark in 1965 for the spread of farming from Anatolia to Europe. Radiocarbon determinations are shown for the earliest sites of farming settlements as dated by 1965. The direction and source of the farming dispersal is very clear. (From Renfrew 1973a, 71.)

lized the elegant device proposed by Ammerman & Cavalli-Sforza (1973) of the 'wave of advance' which modelled the propagation of the farming economy by a process of 'demic diffusion'. Although this was soon criticized by Zvelebil & Zvelebil (1988; 1990) in its application to the European case, and with some justice, as an oversimplification, the basic notion of the arrival of Proto-Indo-European speech in Europe with the coming of farming has nonetheless been widely accepted (Zvelebil 1995). At the same time, the notion of a single Proto-Indo-European dispersal was seen by many linguists as too simple a hypothesis to account for the complexity of the relationships among the various sub-families of the Indo-European language family, well illustrated for instance by Raimo Anttila's diagram (Anttila 1989) of isoglosses. Both objections were valid. They are discussed further below.

In retrospect, then, *Archaeology and Language* suffered from two principal defects (among other deficiencies). In the first place, it laid too much store by the 'demic diffusion' model of Ammerman & Cavalli-Sforza and not enough on the phenomenon of contact-induced language change. Ironically perhaps it is the application of molecular genetics which has given new insights into the limited extent to which demic diffusion took place (see below). And

secondly it did not sufficiently deal with the complexity of the further developments of the Indo-European languages in the long time span after the initial Proto-Indo-European spread, with local convergence (or advergence) effects responsible for the formation of some of the sub-families (see now Renfrew 1999). But it did, perhaps for the first time in the context of recent discussions, establish two principles or processes which turn out to be applicable on a much wider canvas than that of prehistoric Europe. The first is the creation or at least the foundation of a language family not through some long-distance tribal migration (like that of the legendary mounted warrior-nomads) but through a spread phenomenon which may be seen as the result of an intelligible economic and demographic process (on the subsistence/demography model). In this case the spread of farming — a prime case of what Dixon (1997), following Stephen Jay Gould, was later to term a ‘punctuation’, leading to the formation of what Johanna Nichols (1992) was to name as a linguistic ‘spread zone’.

The second principle is a matter of time depth. Hitherto, historical linguists have in general come to employ what the archaeologist would term a ‘short chronology’, with a time depth for language families of often just five or six thousand years before the present. But many of the decisive demographic processes in world history are climate-related, either the direct consequence of such phenomena as the end of the Late Glacial Maximum or the end of the Pleistocene period, or the indirect consequence, dependent for instance upon the origins of farming, a process which is now viewed in most parts of the world as initiated by those climatic events but slower to develop in some areas. By establishing a date as early as 7000 BC for the first spread of a Proto-Indo-European language (or at least by seeking to do so) it became possible to view the origins of language families within a time-frame linked to the end of the Pleistocene period which allowed such global changes to be regarded as relevant to the issue. This is a principle which, for instance, Nettle (1999a) has used to very good effect.

It was the publication of Merritt Ruhlen’s very clear survey *A Guide to the World’s Languages* (Ruhlen 1991) which opened the way for the employment of the same principles, using explicit models (Renfrew 1989a) for language change. The wider application of the language/farming dispersal model, first in Eurasia (Renfrew 1991) and then much more widely (Renfrew 1992b; 1996), suggested a global solution to the problem of the spatial distribution of the world’s languages, although one which at first

presents a number of problems, particularly linguistic ones. It is noteworthy that over the same period Peter Bellwood, working in the Pacific, first with the Polynesian and then with the wider Austronesian language families, reached similar conclusions on the relationship between the dispersals of languages and of agriculture (Bellwood 1989; 1991) in the genesis of language family distributions. He has subsequently generalized these to a world level (Bellwood 1996; 1997).

The third strand in this interdisciplinary scene is the application of molecular genetics, in which Luca Cavalli-Sforza (Cavalli-Sforza *et al.* 1988; 1994) has been a pioneer, and where the pace of new research is extraordinarily rapid. I concluded a review of these matters in 1991 by predicting the emergence of a ‘new synthesis’ between historical linguistics, prehistoric archaeology and molecular genetics (Renfrew 1991, 20). These circumstances form the background to the present enquiry. Matters are made more interesting by a number of developments in historical linguistics offering greater insights into processes of language change. There is also the vexed question of so-called linguistic phyla or macrofamilies (such as ‘Nostratic’ and ‘Sino-Caucasian’) which, if their validity could reliably be established, would have considerable significance for our understanding of prehistory. For instance, the current distributions of the languages comprising the constituent language families of the hypothetical Nostratic macrofamily (including the Indo-European languages) could plausibly be explained by the wider application of the farming/language dispersal model (Renfrew 1991). But, as we shall see, the current status of such macrofamilies is problematic. There is also the potential application of the techniques of molecular genetics to linguistic problems through the mediating concept of population demography.

It was in the light of this very complex situation that it seemed appropriate to use the resources of the McDonald Institute (whose founder Dr D.M. McDonald was first led to visit the Department of Archaeology in Cambridge in 1988 through his interest in the issues raised in *Archaeology and Language*), and to enlist the support of the Sloan Foundation to examine some of these problems through the ‘Prehistory of Languages’ project.

On knowability

At first sight the suggestion of investigating the prehistory of languages may seem like following a will o’ the wisp. For if past languages are recorded only

though the medium of writing, first developed in western Asia and in Egypt in the fourth millennium BC and considerably later elsewhere, how could one hope to consider linguistic events prior to that date? Fortunately the well-established discipline of historical linguistics can supply at least a partial answer to that question. It does so through the concept of the language family, established already in 1786 by Sir William Jones (1807). He indicated, with reference to what is now termed the Indo-European language family, that there were a number of related languages (Greek, Latin, Sanskrit) and sub-families (Celtic, Germanic) which must be 'sprung from some common source', thereby implying both the notion of an earlier proto-language (such as Proto-Indo-European) and an original homeland. Such concepts are now basic to the discipline of historical linguistics and any history of the discipline will indicate how the Neogrammarians of the nineteenth and early twentieth centuries, notably Brugmann, established the principles of linguistic reconstruction (e.g. Durie & Ross 1996, ch. 1) with an understanding of regularities in sound change (phonology) and systematic modifications in word forms and syntax (morphology), accompanied by word loss and innovation in vocabulary (lexical change). The reconstruction of the hypothetical proto-language can thus be undertaken, and confidence is gained in the method, for instance when other languages are discovered which may be recognized as belonging to the same family and conforming with comparable principles in their inferred descent from the reconstructed proto-language.

It is therefore not a vain question to enquire, using the techniques of archaeology, what material traces may have been left by the speakers of these various languages both in the regions where they are recorded as having been spoken, and in the inferred homeland area. There is a long history of such researches, not least in the field of Indo-European studies (Mallory 1973). The matter is of considerable interest to archaeologists as well as to linguists, since many archaeological conclusions are based on broad statements about such issues as the 'coming of the Celts' or the arrival of the first Indo-European speakers (Schrader 1890; Childe 1926). Comparable issues emerge in every part of the world in relation to the relevant languages (e.g. McConvell & Evans 1997; Blench & Spriggs 1997; 1998; 1999a,b).

The relationship between archaeological and linguistic data has, however, always been a difficult one. The old equation, established by Childe (1929), between an archaeological culture, a people and a language is no longer directly acceptable (Renfrew

1989a). Moreover, while material cultural remains can be dated with considerable precision by such techniques as radiocarbon dating, linguistic events are more difficult to date (Renfrew *et al.* in press). And when the data from molecular genetics are brought into play, the situation becomes even more complicated (Cavalli-Sforza *et al.* 1994; Renfrew 1992a).

The matter has become more acute recently with the postulation of such macrofamilies as 'Nostratic' (Dolgopolsky 1998) and 'Amerind' (Greenberg 1987) or indeed 'Sino-Caucasian' (Starostin 1989). It is fair to say that these have been met with scepticism by the majority of historical linguists. But this has brought more directly to the fore the question as to how one should evaluate such proposals. How can one know?

Of course there is nothing inherently unfamiliar in the notion of making verifiable statements about events in the past which can themselves obviously no longer directly be observed. Historical linguistics and prehistoric archaeology are not alone in dealing with processes and events which are now not directly observable owing to the passage of time. Such is the case, of course, with palaeontology and with cosmology. The relevant deciding test, it is widely felt, is that which applies to all statements in the domain of the sciences, namely the Popperian criterion of testability or falsifiability (Bell 1994). But to say this certainly does not solve the problem, when it is not clear by what means a particular proposition may be subjected to testing.

In this field it seems at first sight difficult to formulate generally applicable criteria (McMahon & McMahon 1995; Sims-Williams 1998). Certainly there are cases where a probabilistic approach may be employed, and where it may validly be asked whether claims of significant patterning used to support a hypothesis in fact amount to anything more than configurations which might equally have arisen by chance (e.g. Ringe 1996; 1999). But this approach is more difficult to apply than might at first appear to be the case. For experience shows that in the development of any discipline, new configurations begin to emerge which are not at first clear, and which are sometimes at first contradicted by counter-examples. Often it is not easy to formulate the precise proposition which it is desired to test. One wonders, for example, whether the statistical tests proposed by Ringe (1996) would admit Albanian, Armenian, Tocharian and Sinhalese to the same language family, while few linguists today would doubt that all belong to the Indo-European family.

Sometimes theoretical criteria do not well supply the answer, and it may transpire that the proof of the pudding is in the eating. The criteria may then need subtly to be modified. The task of the Prehistory of Languages project has therefore been to address a number of current issues, seeking the support of a body of scholars with a wide range of opinions, and to see what emerges. The wider aim, however, must be to establish a number of general principles relating language change to demographic, social and economic change. If such principles could be established and supported by a number of well-understood cases, they might be of great value in elucidating other cases which are not yet clear. It is therefore not simply the falsification (or validation) procedure which counts: it is also the undertaking of proposing regularities and hypotheses which are open to further investigation. Recent developments in historical linguistics offer hope that this process is now under way.

Recent developments in historical linguistics

The enhanced level of self-awareness in the field of historical linguistics is well documented by the spate of recent overviews or textbooks to complement such earlier works as those of Hock (1986) and Anttila (1989), including such useful texts as those of McMahon (1994), Fox (1995), Lass (1997) and Campbell (1998) and the valuable collected work edited by Durie & Ross (1996).

More controversial, but of particular interest to the archaeologist because of the very great time depths which she contemplates, is *Linguistic Diversity in Space and Time* by Johanna Nichols (1992). Here she does not dispute the widely-held view among historical linguists that genetic (i.e. familial) relationships cannot, employing the standard comparative method, be observed between languages which became separated more than about 8000 years ago. As she puts it (Nichols 1992, 313):

Standard comparative method, then, requires accidentence of grammar and such things as shared arbitrary lexical categorization as evidence for an initial assumption of relatedness . . . it reconstructs not only a sound system and a lexicon but also a grammatical system which is an individual and not a type. The diagnostic kinds of accidentence apparently persist in genetic groups only for a few millennia, and dissipate entirely after about 8000 years.

Instead she advocates a population typology approach using what she terms stable structural fea-

tures (such as the presence of head/dependent marking, dominant alignment, complexity, word order, voice, etc.) and evaluates the frequencies of such features in broad geographic areas, claiming to draw therefrom conclusions about language dispersals going back to a very much greater time depth, as far as 30,000 years ago. For non-linguists, such as myself, her work presents difficulties — for the non-initiate such concepts as ergativity (see Dixon 1994) are taxing — and it is perhaps for that reason that her work has not been more widely discussed. There is no doubt, however, that on occasion she makes sweeping assumptions which are certainly open to question (e.g. Nichols 1997). For this reason the sustained criticism by Nettle (1999b) of her views on the date of the first colonization of the Americas is of particular interest. However, while some of the historical assumptions may be dubious, and her historical conclusions perhaps should be treated with caution until there has been a fuller evaluation, there is no doubt that her pioneering approach of population typology is one of the most interesting developments of recent years, and one whose further (and perhaps more cautious) application is likely to bring very fruitful results.

From the standpoint of the archaeologist, however, the most promising development in historical linguistics in recent years is a new willingness among some linguists to relate linguistic change to changes taking place also in the material world (the archaeologist is almost tempted to say the 'real world'), including economic, social and demographic changes. For too long, in my view, there has been the assumption in some quarters that language changes by its own rhythms and according to its own rules, quite independently of demographic or social factors. Several recent works have decisively broken that traditional mould. Among them is Christopher Ehret's *An African Classical Age* (Ehret 1998) where the linguistic history of Africa is related to the social and economic transformations which have taken place there over the past ten thousand years or so (although it should be noted that his view of the origins of the Afroasiatic language family differs significantly from that recently advanced by Diakonoff (1998)).

Perhaps yet more significant, because framed in more general terms, are two refreshing works which will perhaps open a new chapter in historical linguistics, and in particular in the understanding of linguistic change in relation to demographic and social change, and hence potentially to the archaeological record. The first of these is Daniel Nettle's

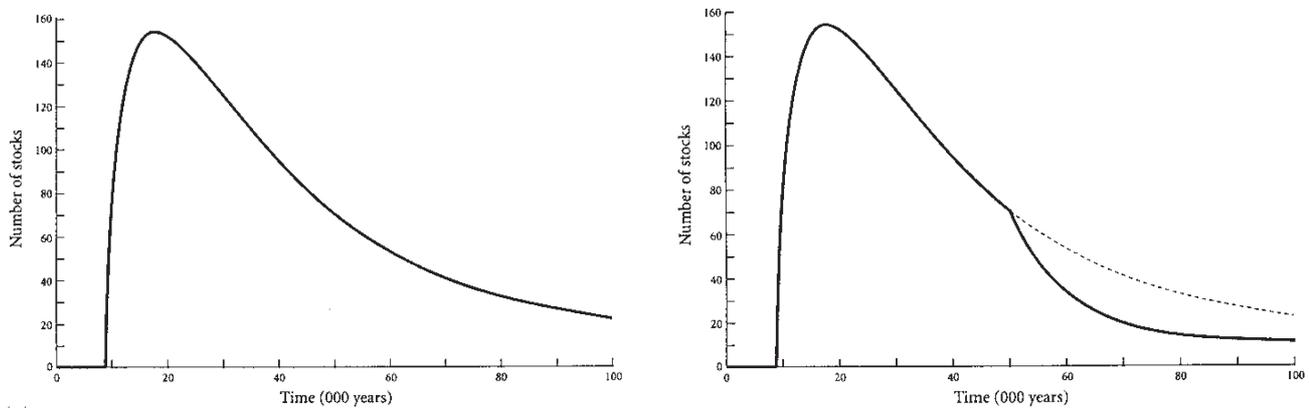


Figure 3. (Left) Diagram by Nettle (1999a, 124 & 125) predicting the number of language families (or 'stocks') over time for a notional continent using the assumptions described below. (Right) Predicted number of language families over time adding a Neolithic transition 50,000 years after first settlement.

Linguistic Diversity (Nettle 1999a), in keeping with recent archaeological work, especially in relation to hunter-gatherers (Binford 1999), he puts great weight upon demographic factors. In his treatment of patterns in space and his willingness to quantify he owes much to the earlier work of Nichols (1992), but his approach is anthropological as well as linguistic, situated very much in the 'real world' of the anthropologist as well as that of the historical linguist. It is self-evident to him that the dynamics of linguistic change are very different in the case of hunter-gatherers (typically with a population density of about one person per ten square kilometres) from that of subsistence farmers (often with a population density of ten persons per square kilometre — thus greater by up to two orders of magnitude). His treatment of phylogenetic diversity (Nettle 1999a, ch. 6) seems compelling, and an example of his thinking is seen in Figure 3.

Nettle argues (1999a, 121) that when humans first moved into a new area there would be a great variety of available geographical niches and the population would fission repeatedly and often. The newly split languages would go on changing until they were sufficiently different to be identified as different language families. On his diagram he allows about 10,000 years before divergence has proceeded so far as to produce different language families (no longer recognizable as genetically related) and about 20,000 years for the completion of this process. But he then follows Trubetzkoy and Dixon's convergence approach (see below) and predicts a *decrease* in the number of apparent language families. 'This could happen through extinction, as some peoples absorbed others or through large scale areal convergence on certain linguistic items, which

makes it difficult for subsequent linguists to identify the stock boundaries.' In effect, newer (and larger) family units would be produced though convergence effects, with a consequent reduction in the total number of language families in the area.

As noted earlier, Nettle's approach differs markedly from that of Nichols, who assumes a continuing linear increase with time in the number of language families or stocks within a given area. The controversy, then, is not simply about the date of first colonization of the Americas, currently much disputed, but rather one of a conflict of methodologies, both of which have much to offer to the more traditional brand of historical linguistics.

Another refreshing, if again controversial work, is R.M. Dixon's *The Rise and Fall of Languages* (Dixon 1997). Here Dixon in a sense takes a path already indicated by an earlier and influential study by Thomason & Kaufman (1988) and earlier by Trubetzkoy (1939), in that he lays emphasis upon convergence processes, whereas the traditional comparative method thinks mainly in terms of divergence and family trees, with the 'wave model' for linguistic change (in a sense a convergence model) very much in second place. He summarizes his position, inspired by the punctuated equilibrium model in biology, very clearly (Dixon 1997, 3):

Over most of human history there has been an equilibrium situation. In a given geographical area there would have been a number of political groups, of similar size and organisation, with no one group having undue prestige over the others. Each would have spoken its own language or dialect. They would have constituted a long-term linguistic area, with the languages existing in a state of equilibrium . . . Then the equilibrium would be punctuated and drastic changes would occur. The

punctuation may be due to natural causes such as drought or flooding; or to the invention of a new tool or weapon; or to the development of agriculture; or of boats, with movement into new territories; or to the development of secular or religious imperialism. These punctuations to the state of equilibrium are likely to trigger dramatic changes within languages and between languages. They give rise to expansion and split of people and of languages. It is during a period of punctuation — which will be brief in comparison with the eras of equilibrium that precede and follow — that the family tree model applies.

The book develops this thesis in greater detail, and emphasizes that what appear to be language families can emerge not only as a result of a punctuation and population dispersal (where the family tree model of the standard comparative method applies) but also as the consequence of long-term convergence resulting in equilibrium. Here there is no family tree, and no single proto-language, and the standard Neogrammarian methods of linguistic reconstruction do not apply. He applies this convergence model specifically to the case of the Pama-Nyungan language family of Australia. This has been a bone of contention in Australian linguistics for some years, and several attempts have been made, following the standard dispersal/family-tree model, to find a homeland for a postulated proto-Pama-Nyungan language and to elucidate the mechanism of its dispersal in material (and archaeological) terms (McConvell & Evans 1997). In Dixon's view such an effort is wasted on language families of his equilibrium type, being applicable only with language families (such as Indo-European) which can be shown to belong to the punctuation/family-tree type.

What is refreshing here is not only the greater prominence given to convergence models and the recognition that the family tree model of the standard comparative method may not always apply, but also the willingness to introduce historical processes, tangible material processes (floods, development of agriculture, imperialism), into the linguistic discussion. This, as we have seen, is what Nettle also does, but what many linguists, including Nichols, have generally failed to do.

These approaches, both the development of new patterns of linguistic analysis (by Nichols (1992), by Thomason & Kaufman (1988), by Ross (1996; 1997) and by Tryon (1999)), and the introduction of an interplay between linguistic and more material factors (demography, economy) by authors such as Nettle and Dixon, open up new avenues for the study of the prehistory of languages.

It certainly cannot be claimed that our Prehistory of Languages project has yet made any significant contribution to these developments in historical linguistics, which have been under way for at least a decade. But the attempt to bring together prehistoric archaeology and historical linguistics in a new way, and the emphasis upon the impact of agriculture in my *Archaeology and Language* and in the early work of Peter Bellwood, did perhaps lead to a new willingness among archaeologists to contemplate linguistic issues. Since the excesses of the German National Socialists before and during the Second World War, with their fantasies of Aryan supremacy, there had been a general and understandable reluctance to discuss such long-standing issues as the Indo-European question, with Marija Gimbutas (Gimbutas 1973; 1997) an exceptional and notable pioneer. Over the past decade, however, there has been a new willingness to discuss such matters, exemplified by the four-volume collective work also entitled *Archaeology and Language* published in the aftermath of the Third World Archaeological Congress held in New Delhi in December 1994, where a major session on that theme was organized by Roger Blench, Matthew Spriggs and myself (Blench & Spriggs 1997; 1998; 1999a,b). Durie & Ross (1996, 10) incorporate the language/farming model into their thinking, as indeed does Nettle, who with his reference to the 'Palaeolithic equilibrium', the 'Neolithic punctuation' and the 'Neolithic aftershock' (Nettle 1999a, 100–107) utilizes also the terminology of Dixon. Dixon himself, although his published acknowledgement of *Archaeology and Language* is scarcely fulsome ('Renfrew . . . has an agile mind but lacks an appropriate training in the methodology of historical linguistics for his work to constitute a linguistically significant contribution': Dixon 1997, 48n), was earlier kind enough to write after my visit to the Australian National University in Canberra and my Mulvaney Lecture delivered there in August 1993 saying that my reference in it to the convergence model of Trubetzkoy (1939) had been helpful in the development of his ideas (Dixon, *in lit.*). And indeed the model of the gradual convergence of initially independent languages to form a new language family as conceived by Trubetzkoy (see also Renfrew 1987, 108–9) does clearly anticipate aspects of Dixon's work, even though its application to the Indo-European case has always provoked the disapproval of historical linguists (e.g. Dolgopolsky, in Renfrew 1991, 9).

Moreover, while Dixon is correct in remarking that I am not a linguist, and in that observation he is

in good company (viz. 'Renfrew's erudition in linguistics is insufficient': Diakonoff 1988, 79), perhaps it is not unduly polemical in reply to point out that the two central points in the passages by Dixon quoted here (on language displacement 'punctuations' following either the subsistence / demography or elite dominance models, and on time depth) were both anticipated ten years earlier in *Archaeology and Language* (Renfrew 1987, 121–37, and 165–8 respectively).

There is much work which remains to be undertaken in the discussion of convergence processes, which are evidently in action all the time when languages are in contact, just as divergence processes continually operate when related languages have become isolated from each other. Strictly speaking, the notion of *convergence* generally applies to languages which are not themselves genetically related. But a very common phenomenon is the comparable process of mutual influence when two separate languages, which are in fact genetically related through descent from a common ancestor, occupy adjacent territories and continue to interact. In such a case it may be preferable to speak of *advergence*, and this process would appear to play a major role in the formation of the sub-families of which many language families are composed. In the Indo-European case, for instance, the Celtic sub-family may be less the product of a further punctuation episode from a notional proto-Celtic homeland than the result of advergence processes between different dialects of the early form of Proto-Indo-European which was spoken in north-western Europe after its initial spread. Such issues, which may underlie the internal structure of many language families, deserve closer attention.

The essential point, however, is that there is a new spirit abroad in the field of historical linguistics which indeed fulfils some of the earlier aspirations of sociolinguistics but on a wider scale, both spatially and temporally. New models are being developed, and new ways found of making linguistic data relate to those of prehistoric archaeology and even of molecular genetics.

Time depth

One of the principal obstacles to progress in seeking a *rapprochement* between the data of archaeology and those of historical linguistics has been that the two disciplines often use differing timescales. The timescales of prehistoric archaeology suffered acute dislocation a quarter of a century ago, when the impact of the tree-ring calibration of radiocarbon dating was acutely felt in some areas (Renfrew 1973b)

with a readjustment of ages by as much as 2000 years. Timescales in historical linguistics are naturally constructed with very different assumptions, and it is sometimes difficult to establish just what these are. Certainly most linguists reject the system of 'glottochronology' as originally applied by Swadesh (1960; 1972), and appear rather to extrapolate back from the well-dated development of the Romance languages from late Latin, or perhaps that of the Semitic languages from Old Akkadian. In some cases, however, an assumed chronology for the development of the Indo-European languages from Proto-Indo-European is used as a basis for further comparisons, when in fact the Indo-European chronology in question depends on a view of Indo-European origins which can scarcely be regarded as securely established (Mallory 1989).

Moreover it is common in historical linguistics to assume some specific chronological threshold beyond which the techniques of the comparative method cannot penetrate. We have seen above how Joanna Nichols felt able to pronounce a time limit of 8000 years. Similarly Ringe (1995, 72) has written:

If after ten millennia or twelve (or whatever the threshold is exactly) the similarities between diverging languages of common origin become indistinguishable from similarities which could have arisen by random chance, language relationships at that and greater time depths *simply cannot be posited by scientific linguists* . . .

but the scientific basis for this threshold has not been established. Kaufman & Golla (2000, 47) have similarly written: 'The possibility of establishing a genetic grouping requires . . . that the relationship is not older than 8000–10,000 years before the earliest date at which the languages are documented.' It should be noted, however, that for languages which are documented in writing at an early date, such as the earliest known Indo-European or Semitic languages, this ruling may permit a time depth of more than 14,000 years before the present.

It is not generally clear upon what principles such observations rest, although in the case of Kaufman & Golla they explicitly acknowledge the use of a form of glottochronological reasoning. The point has been made trenchantly by Dixon (1997, 47–9):

What has always filled me with wonder is the assurance with which many historical linguists assign a date to their reconstructed proto-language. (And these are, by and large, people who firmly reject the glib formulas of glottochronology.) We are told that proto-Indo-European was spoken about 6000 years ago. What is known with a fair degree of certainty is the time between proto-Indo-

Aryan and the modern Indo-Aryan languages — something of the order of 3000 years. But how can anyone tell that the development from proto-Indo-European to proto-Indo-Aryan took another 3000 years . . . Why couldn't proto-Indo-European have been spoken about 10,500 years ago? This would correlate with a major socio-economic development, the introduction of agriculture, which archaeologists date at about 10,500 bp for this part of the world . . . Surely the only really honest answer to questions about dating a proto-language is 'We don't know'.

The received opinion of a date of around 6000 bp for proto-Indo-European — with dates for other proto-languages being calibrated on this scale — is an ingrained one. I have found this a difficult matter to get specialists even to discuss. Yet it does seem to be a house of cards (reinforced, at one time, by the chimera of glottochronology). This is a question that demands careful re-examination with a full range of possibilities being discussed and compared.

Such considerations as these led to the Symposium held within the framework of the Prehistory of Languages project at the McDonald Institute in August 1999 on the topic 'Time Depth in Historical Linguistics', with the aim of reviewing these issues, drawing upon as wide a range of linguistic scholarship as possible. The papers, which were precirculated, are now being prepared for publication (Renfrew *et al.* in press). Although many different standpoints were expressed it was agreed that chronological evaluation is difficult for periods before language was recorded in writing. The lexicostatistical approach advocated by Starostin, which differs in a number of respects from the earlier glottochronological methods of Swadesh, attracted widespread interest (Starostin in press). One essential point is that some words in a language may be more stable (i.e. resistant to decay) than others, and that any valid method for evaluating age or date must take note of this factor.

Most speakers were in agreement that it is difficult to proclaim any precise time threshold, but also the point made in a number of papers by Ringe and others that there is a certain level of noise through the emergence of lexical resemblances by chance, was accepted. The other area of promise was the population typology approach of Nichols already mentioned above. But while the method was accepted as exceptionally promising, it was recognized that the adoption of dubious assumptions in utilizing it could inevitably lead to doubtful results.

The Symposium certainly succeeded in bring-

ing questionable assumptions to the fore, and in recognizing the considerable problems involved in establishing reliable chronologies on the basis of linguistic data alone. My own view is that the scepticism of Dixon is well warranted, but that the approaches of Starostin and of Nichols hold considerable promise for future progress. It would also appear that most of the obstacles which formerly seemed to prohibit a time depth of the order of 8000–10,000 years for such language families as Indo-European or Afroasiatic have now been removed. That does not, however, imply that greater time depths than those formerly accepted have been established: simply that they are no longer prohibited as a matter of principle.

Trouble with macrofamilies

The question of the validity or otherwise of linguistic macrofamilies has so far been one of the principal foci of the Prehistory of Languages project. The concept of the linguistic phylum or macrofamily is most familiar in the West from the work of Joseph Greenberg, and in the first place through his division of the languages of Africa (Greenberg 1963) into four major linguistic phyla: Afroasiatic, Niger-Kordofanian, Nilo-Saharan and Khoisan. This division met with quite widespread acceptance, although some commentators have subsequently questioned whether the Khoisan macrofamily represents a real cladistic grouping (i.e. one with common ancestry) or is merely an area grouping (of languages situated in proximity, without genetic relationship, but perhaps showing similarities due to convergence effects). The status of Nilo-Saharan has also been questioned.

In Russia, Vladislav Illič-Svityč was from 1967 publishing papers, and from 1971 to 1984 his dictionary in three volumes (see Illič-Svityč 1989; 1990) was devoted to the Nostratic macrofamily, a concept first adumbrated by the Danish linguist Holger Pedersen (1931). Aharon Dolgopolsky was, at much the same time, working along comparable lines (Dolgopolsky 1973). The Nostratic macrofamily, as formulated by Illič-Svityč and Dolgopolsky, is a broad grouping comprising the Indo-European, Afroasiatic, Kartvelian, Uralic, Altaic and Dravidian language families. Work along broadly similar lines has been undertaken by Greenberg, well summarized by Ruhlen in his 1991 postscript to *A Guide to the World's Languages* (Ruhlen 1991, 383–5) and the publication of Greenberg's *magnum opus* on this subject is awaited (Greenberg in press). Other proposed

macrofamilies are Sino-Caucasian (Starostin 1989) and Austric (Blust 1996), involving several language families of southeast Asia and the Pacific. By far the most controversial work of this kind, however, has been Greenberg's *Languages in the Americas* (Greenberg 1987) which has elicited a storm of protest from Americanist linguists.

It is not easy for a non-linguist to assess such matters, but their archaeological significance, if they were established as valid, would be considerable. Some years ago (Renfrew 1991), I pointed out how the distributions of some of the language families in the Nostratic macrofamily (Indo-European, Afroasiatic, Altaic and Dravidian) might all be regarded as the result of farming dispersal processes from the relevant areas of Western Asia. Comparable suggestions have been made suggesting a possible relationship between the distribution of the language families comprising the Austric macrofamily and the dissemination of rice agriculture (Higham 1996). These correlations would be of considerable interest to the archaeologist if the validity of the relevant classifications were accepted.

Although the principal conclusions of Joseph Greenberg's *Languages in the Americas* (Greenberg 1987) have not been accepted by the majority of linguists, his suggestion of three waves of colonization of the Americas, corresponding to the speakers of his three American macrofamilies (Amerind, Na-Dene and Eskimo-Aleut) has found some support from anthropologists (Greenberg *et al.* 1986) and from some molecular geneticists (e.g. Torroni *et al.* 1992; 1994).

The debate on these matters among linguists has been loud, even unseemly. Dixon is scornful both of the Russian Nostratic work and of Greenberg's 'Amerind' proposal, remarking tartly of this and Greenberg's earlier African work (Dixon 1997, 35): 'It is also rather like the clock that strikes thirteen — not only is it implausible in itself but it casts doubt on all that has gone before.' Lass (1997, 162 n. 76) carries such criticism beyond the usual confines of academic discourse in commenting upon an article in *Scientific American* by Greenberg & Ruhlen (1992): 'This article is now part of a long and rather disgusting debate, conducted in part in the lay press, and characterized by unedifying *ad hominem* attacks upon the anti-Greenberg camp; a particularly cheap and nasty specimen appeared in the prestigious *Atlantic Monthly*.' Since I had myself suggested to the editor of *Scientific American*, following publication there of an article of my own (Renfrew 1989b), that he might find profitable a contribution by Greenberg or by Ruhlen on this interesting macrofamily ques-

tion, I find myself slightly disconcerted by the vehemence of these remarks.

In order to form our own independent perspective upon these matters we invited Aharon Dolgopolsky, who is now resident in Israel, to deliver a seminar on the Nostratic hypothesis at the McDonald Institute. Following the success of that occasion we invited him to prepare a longer paper which was then published (Dolgopolsky 1998) in preparation for a more ambitious Symposium. This was duly held with precirculated papers in July 1998, and these have now been published under the title *Nostratic: Examining a Linguistic Macrofamily* (Renfrew & Nettle 1999) within the framework of the Prehistory of Languages project and with the support of the Sloan Foundation. As the reader may judge, the advocates of the Nostratic hypothesis adduced much detailed material. But the critics gently made the point that the accumulation of lexical examples may well not be conclusive until a quantitative analysis can be undertaken to exclude the possibility that the number of resemblances exceeds by a significant factor that which might have arisen through chance. What is still lacking is a more comprehensive morphological analysis, to show that the specific morphological features of the proto-language are inherited, albeit with modifications, in the daughter language families. My own assessment (for what it is worth) is that the case remains not proven, and I was impressed that a number of experienced linguists who are not in the Nostratic camp also feel that the matter is not yet resolved. The McDonald Institute is hoping to publish Aharon Dolgopolsky's forthcoming *Nostratic Dictionary* in the near future, and along with the forthcoming publication of Greenberg's Eurasiatic volume (Greenberg in press), this should offer scope for further assessment.

Sergei Starostin has agreed to produce a major introductory paper on the Sino-Caucasian hypothesis, of which he is the principal architect, and it is our hope to publish that as a preliminary to a further Symposium within the framework of our Project, which should in turn lead to a further publication. As the distinguished Russian linguist the late Igor Diakonoff (1990, 62) remarked: 'According to Starostin, the Caucasian languages are related to the Ket on the Yenisey River and to Sino-Tibetan. This seems to be rather a mad statement, but coming from a linguist of the stature of Starostin, cannot be just disregarded.' It is indeed difficult to imagine what the archaeological reality underlying a Sino-Caucasian language family might be, were the hypothesis to find acceptance, although Cavalli-Sforza

(Cavalli-Sforza & Cavalli-Sforza 1996, 184) has published a map, suggesting a pre-Nostratic linguistic unity over much of northern Eurasia, following the views of John Bengtson, who would seek to relate Basque and possibly Sumerian to the Sino-Caucasian macrofamily, to which the Na-Dene languages of North America would also be linked. But this is running before one can walk, and a proper evaluation of the Sino-Caucasian hypothesis must surely precede the inclusion of these other elements. So far, the hypothesis rests largely upon the work of Starostin, most of it published in Russian, and awaits a systematic and wider evaluation.

Turning once more to the Amerind controversy, there was opportunity to make some evaluations at first hand during a Symposium held in September 1998 at the McDonald Institute, again under the aegis of the Prehistory of Languages Project, when geneticists and linguists met to discuss the theme of molecular genetic variation at the local level ('population-specific polymorphism'). This phenomenon has been most intensively studied in Mesoamerica and South America, and the meeting inevitably discussed also the status of the proposed 'Amerind' macrofamily. The two linguistic papers offered, by Terrence Kaufman and Victor Golla, were critical of the concept. Merritt Ruhlen, in his role of Greenberg's principal advocate naturally supported it with lucid arguments, whereas Don Ringe was critical on carefully argued probabilistic grounds. I myself suggested that the molecular genetic evidence, while it might support Greenberg's insistence on the significance of the Na-Dene language family, was essentially neutral on the Amerind issue. It may not be in doubt that the speakers of languages classed by Greenberg as Amerind arrived earlier in North America than the speakers of Na-Dene or Eskimo-Aleut languages. But that does not in itself give any insight into the issue as to whether the Amerind languages are all genetically related, or whether they represent simply an area classification without genetic significance. The papers in question will be published later this year (Renfrew 2000). My own assessment (again, for what it is worth) is that there is less support for the 'Amerind' category among historical linguists than there is for instance for the Nostratic macrofamily, and I am impressed by the reluctance of scholars such as Starostin or Dolgopolsky (who would be classed as 'lumpers' rather than as 'splitters' by most observers) to endorse it.

My overall conclusion on this theme is that it would be imprudent at present for archaeologists to place much reliance upon the macrofamily concept,

or on the specific macrofamilies which have been coherently argued. To the outside observer, the evidence adduced by Greenberg and Ruhlen on the one hand, and by Illič-Svityč and Dolgopolsky on the other is cumulatively impressive. But the critics of the macrofamily concept are at their most effective when they are least shrill. It is the willingness of a number of internationally known historical linguists to sit down for a weekend and spell out their objections to the Nostratic hypothesis in a measured manner which I have found particularly impressive (rather than the polemic of more assertive critics), and it is not easy to set their scepticism aside. The arguments of Ringe (1995) that he was unable to verify the validity of the Nostratic hypothesis using a probabilistic approach do carry weight. Yet at the same time there are many cases in the history of science where pattern recognition is at first very difficult, and where uncertain and initially unverifiable hypotheses come later, through new developments, to be substantiated. There are moments when it appears that the 'multilateral comparison' method of Greenberg may sometimes elicit in outline patterns which are not yet open to documentation by the more rigorous procedures of the comparative method. However this may prove to be an erroneous impression. Moreover, even if it were to prove justified, it does not preclude the possibility that the Greenbergian method of multilateral comparison might give valid answers in some cases (perhaps the 'easy' ones, such as the Indo-European language family) and erroneous ones in others. These matters may become clearer over the next few years.

I look forward to learning more about the Sino-Caucasian and Austric hypotheses, and to hearing them debated by competent historical linguists. Meanwhile I believe that prehistoric archaeologists would be wise to regard all these macrofamilial claims as interesting and promising hypotheses, but for the present not as much more than that. In what follows I shall deliberately try to put the macrofamily hypothesis on 'hold', and to deal only with generally accepted language families. Thus if reference is made to 'Amerind' it will employ that term to mean Native American languages other than the Na-Dene and Eskimo-Aleut families, using it to designate an area group of languages without any implications of genetic relationship. The same observation holds for the designation 'Indo-Pacific' applied to the languages of New Guinea. The terms 'Nostratic', 'Eurasian' and 'Austric' will be avoided. It should also be borne in mind that the 'Altaic' language family is not everywhere regarded as a valid entity in

the genetic sense, so that it may be more cautious instead to refer to its Turkic, Mongolian and Tungus components. This should, however, be regarded as a benign scepticism: while keeping an open mind on these matters it is for the moment more prudent to avoid using these terms in the subsequent analysis.

The impact of molecular genetics

It is by no means easy to see how to establish equivalences between the data of molecular genetics and those of historical linguistics. Molecular genetics is informative about historical demography, and when language distributions may be regarded as the product of migrations or other demographic events and processes, then progress may be possible. Frequently, however, there are chronological obstacles. For, as we have seen above, linguistic features and events are notoriously difficult to date. Until recently this was precisely the problem for molecular genetics also, although we shall see below that the position is changing, and approximate datings can now be provided in favourable cases using internal molecular genetic evidence. Above all, the pace of progress in molecular genetic research is almost overwhelming, and it is to be predicted that over the next few years there will be further important developments.

In some senses, however, the applications of genetics in archaeology did not get off to an altogether favourable start. Out of three pioneering applications, using (it should be noted) classical genetic markers — blood group and gene frequencies, rather than specific nucleotide sequences as obtained from mitochondrial DNA analyses and the study of Y-chromosome polymorphisms — two have led to questionable claims. In one case the trouble was mainly chronological. In the second it arose from rather doubtful linguistic assumptions. These setbacks have unfortunately, if not surprisingly, resulted in a rather unfavourable reception of molecular genetic procedures among historical linguists (Bateman *et al.* 1990; McMahon & McMahon 1995; Sims-Williams 1998).

The first case was the production of synthetic maps of human gene frequencies in Europe (Menozzi *et al.* 1978; see also Cavalli-Sforza *et al.* 1994, 290) and more particularly their interpretation. Using 38 genes (later increased to 95) assessed at a large number of locations in Europe, they used standard statistical techniques first to calculate genetic distances between the locations and then to configure the data into six principal components. Geographic maps were then prepared for each principal component. Those for

principal components 1 and 2 are seen in Figure 4.

These maps are still generally regarded today as an accurate representation of the variability as indicated by the study of classical genetic markers. The difficulty comes with the interpretation. They proposed that the first principal component (representing 28.1 per cent of the variance) with its marked clines from southeast to northwest represented the genetic trace of the demic diffusion process which, as Ammerman & Cavalli-Sforza had suggested (1973), motivated the spread of farming from Anatolia to Europe. The second (representing 22.2 per cent of the variance), with its marked north–south axis, they felt might be temperature-dependent, while the third (representing 10.6 per cent of the variance), with its marked east–west axis, they suggested could be the result of the ‘Kurgan’ migrations seen by Gimbutas (1973) as the process which introduced Proto-Indo-European speech to Europe. The fourth principal component (representing 7.0 per cent of the variance), with its high values in west Anatolia, south Greece and south Italy (with Sicily) they associated with the Greek colonization process in the first millennium BC. The trouble with these ‘explanations’, it has always seemed to me, is that there is no framework of inference for any of them. They are plucked from the air, and used to substantiate whatever seems a convenient theory.

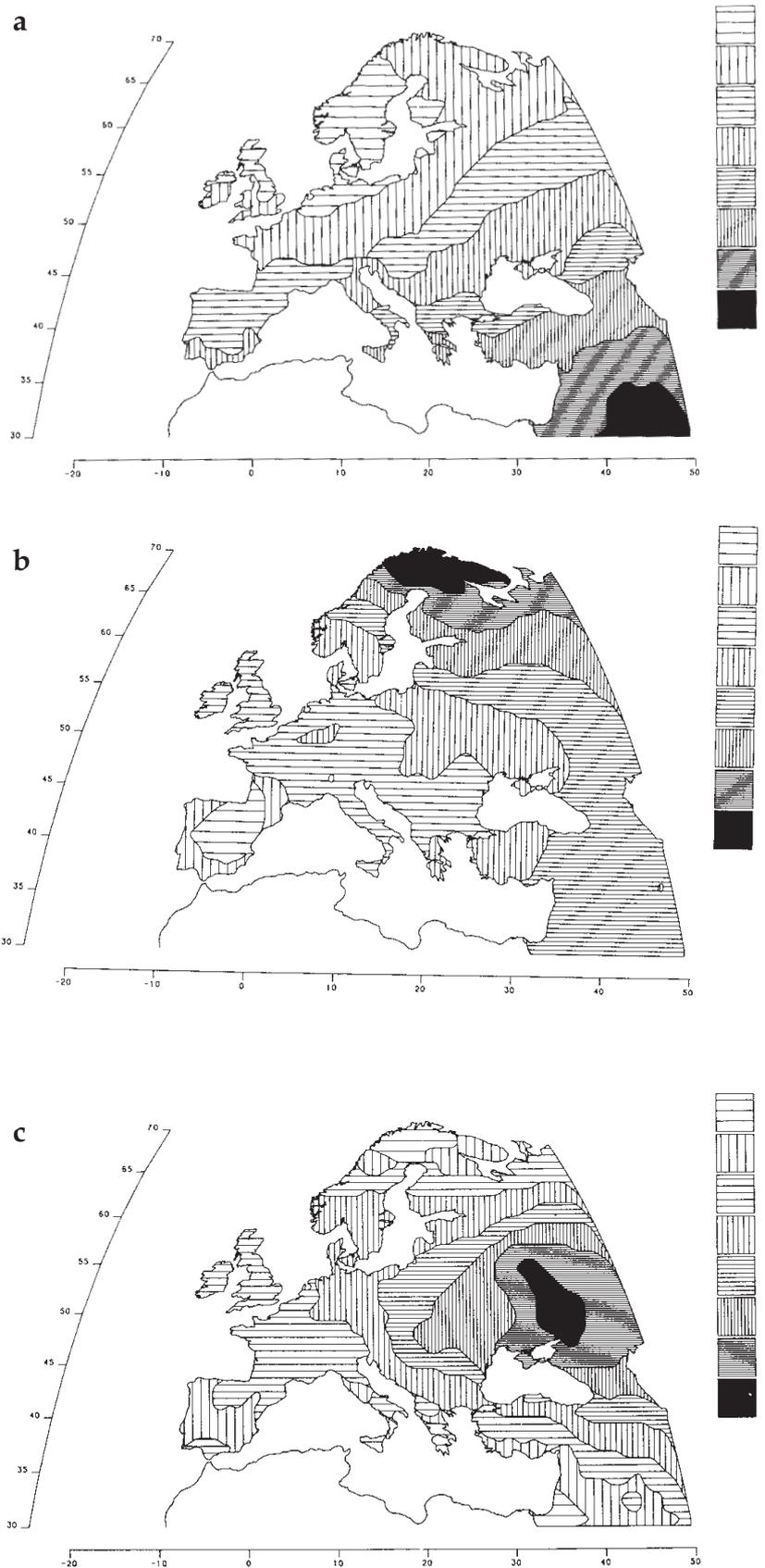
But the difficulty is that these maps, reflecting the current situation (since they are based on samples derived from living populations) represent a palimpsest of all the historical processes and events which have taken place since the first population of Europe by *Homo sapiens*. As we shall see below, the first principal component may now be attributed more to earlier colonization processes from Anatolia, taking place during the Late Pleistocene period, although the spread of farming may be involved at a subsidiary level. The second principal component, as noted below, may relate to the spread of population north from Iberia after the Late Glacial Maximum in the centuries following 14,000 years ago. The third may rather represent the spread of pastoralism from west to east (since these maps establish axes but not direction), as well as subsequent movements from east to west by Cimmerians, Scythians, Huns, and Mongols, as much as any notional exploits of ‘Kurgan’ warrior nomads. And I have always doubted whether the city-founding colonizations by the Greeks during the Classical period involved sufficiently large migrating populations to count for much in such an analysis — how about the effect of Roman colonization in the Mediterranean during the

Empire instead? So while the patterning is impressive, the explanations adduced fail to carry conviction.

The second piece of work, by Laurent Excoffier and his colleagues (1987, reproduced in Renfrew 1992a, fig. 5), used classical markers (gamma-globulin polymorphisms) sampled from a wide range of African populations, to establish a classificatory tree (dendrogram) for these African groups. These were initially classified, at the sampling stage, according to the language family of the language spoken by each group. The dendrogram resulting from the genetic data yielded a grouping which had the effect of classing together populations which happened to be speakers of the same language family. So a classification based on gammaglobulin analyses turned out to produce also a classification which was valid in linguistic terms. This seems to me the most effective application so far of genetics to historical linguistics. For the consequence is that the Afroasiatic speakers, the Niger-Kordofanian speakers and the Khoisan speakers are grouped together successfully purely on the basis of the frequencies of varieties of gammaglobulin in the blood. This impressive result has been overlooked by many of those who have criticized the application of genetics to historical linguistics.

The third early case is, however, more problematic. In 1988, Cavalli-

Figure 4. Synthetic maps of Europe and western Asia obtained by Menozzi et al. (1978) using (a) the first, (b) second, and (c) third principal components of the variations obtained by the analysis of the gene frequencies of 95 genes (using classical genetic markers). They interpreted the first component map (a) as reflecting the spread of agriculture from Anatolia to Europe. The pattern in map (b) was ascribed to climatic factors, and that in map (c) to the arrival of Indo-European speakers from the east, following the 'Kurgan' model of Gimbutas (1973).



Sforza and his colleagues (Cavalli-Sforza *et al.* 1988) published an analysis which purported to show a very general correlation between genetic and linguistic data pertaining to evolution. Indeed the article was widely interpreted as claiming the co-evolution of genes and of languages, which would be a remarkable result indeed, given the low rate of genetic mutations in humans (of the order of one per 10,000 years) and the high rate of lexical change (estimated by Swadesh in his glottochronological exercise, discussed above, at a word loss of 14 per cent of the total every thousand years). Their diagram is reproduced as Figure 5.

The genetic classification employed here might

now be regarded as controversial, and more sensitive approaches are now available using molecular genetic markers, notably mitochondrial DNA (see Cann *et al.* 1987) and Y-chromosome studies. But the linguistic classification embodies the macrophyletic speculations of Greenberg (Eurasian, Amerind) and of the Russian School (Nostratic) as well as the hypothetical Austric macrofamily. Moreover the sequence in the listing of populations in the genetic tree (or indeed of the linguistic tree) may legitimately be altered by rotating about any of the nodes in the tree (treating the diagram in effect like a three-dimensional mobile), so that the conjuncture, even accepting the linguistic classification, is in part illusory.

The risks involved in using macrofamily categories in this way should be apparent from the last section.

But even if such a conjuncture were established, it would not necessarily say much about genetic and linguistic co-evolution as long-term processes. For that to apply, the languages and the genes would have had to evolve in the same localities over long time periods. But it is clear (as will be further discussed in the next section) that many of the languages involved have had their distributions determined by relatively recent (i.e. over the past 10,000 years) dispersal processes. If these processes were accompanied by a significant degree of gene flow, then the correlations in the diagram would be the result of such dispersals and spatial relocations, without saying anything significant about change (evolution) in either genes or languages. To make these criticisms is not intended to show disrespect for the pioneering work involved in producing such a diagram. Speculation of this kind is the very basis of progress in any discipline. But that does not mean that the conclusions suggested have to be accepted. Nor does the rejection of those conclusions imply any

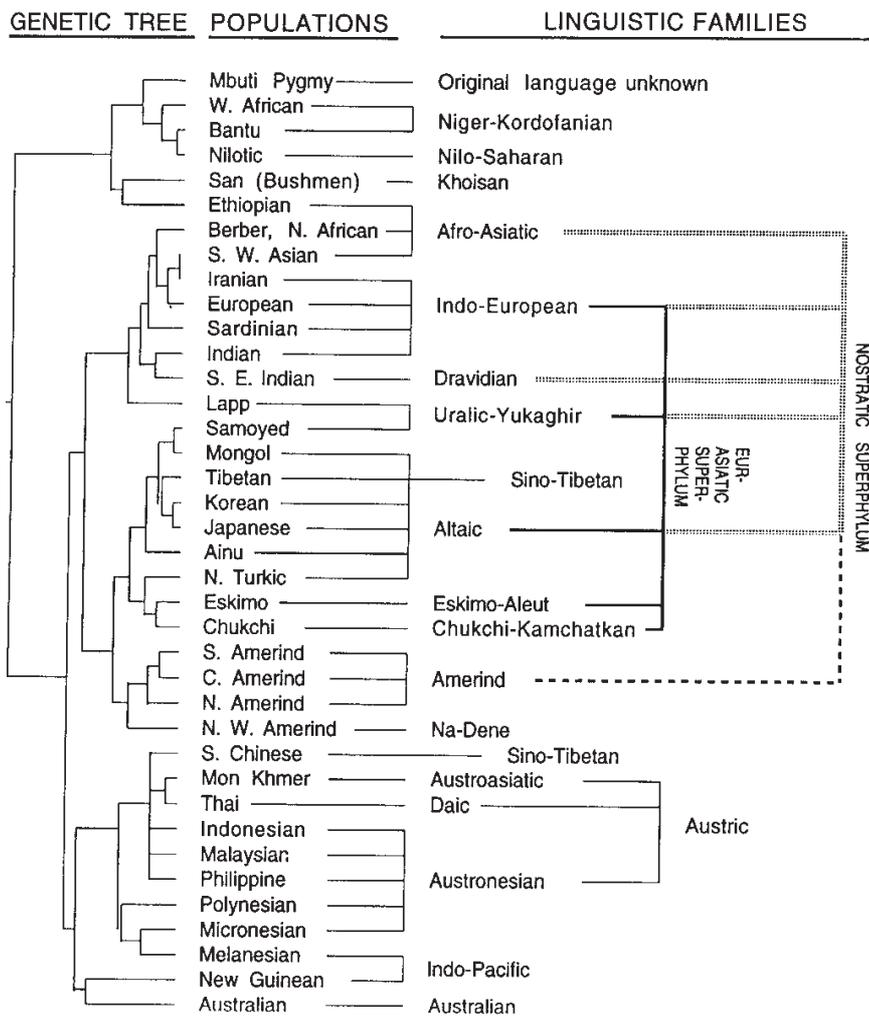


Figure 5. A comparison between genetic (left) and linguistic (right) family trees relating 42 populations of the world. The genetic classification is based upon the analysis of frequencies for 120 alleles for classical (non-DNA polymorphisms) analyzed for several dozen individuals at each locality. The linguistic classification follows that of Ruhlen (1991). (From Cavalli-Sforza *et al.* 1994.)

general criticism of the application of genetics to historical linguistics. It implies no more than the judgement that in this case the proposed correlation (a) does not work and (b) even if it did would not demonstrate co-evolution.

There is no doubt that there are meaningful correlations between linguistic and genetic data, as studies by Barbujani (1991; 1997; Barbujani & Sokal 1990; Barbujani *et al.* 1989; 1994) have shown, and as the work by Excoffier and colleagues, noted above, demonstrates. That is not surprising, nor need it necessarily be very informative. Yet the whole field of population-specific polymorphisms is a promising one, of relevance here when the population is defined in linguistic terms, as is indeed the case with many studies, for instance of African or South American tribes. At the Symposium held at the McDonald Institute in 1998 there was discussion of such issues, and the study of the linguistic relationships within the Yanomama tribe of the Lower Amazon (Merriwether *et al.* 2000) showed a very high degree of tribal-specific polymorphism.

It seems likely, however, that the main contribution to this field offered by molecular genetics, at least in the near future, is likely to be in the reconstruction of population histories. Such work is now being undertaken in every continent, and the position in the Americas is particularly well developed (Torrioni 2000; Forster *et al.* 1996; Merriwether 1999), although hampered there by the on-going archaeological controversy about the date of the first colonization of the Americas. However it is in Europe that some of the issues have been widely debated, and that chronological developments have been particularly informative.

The first modern genetic observations in Europe bearing upon linguistic interpretations were based upon maps of blood group frequencies (Mourant *et al.* 1976) and specifically upon that of the Rhesus negative gene (see also Renfrew 1992a, fig. 4). This has a particularly high frequency in the Basque area of Spain, and the proposal is an obvious one that either the 'coming' of the Basques might be reflected by this high frequency, perhaps indicative of an incoming population, or alternatively that this pattern is, like the Basque language itself, a residual feature of an earlier population which has in some respects been swamped by the arrival of a subsequent Indo-European speaking population (see Cavalli-Sforza 1988). Similar suggestions could be made with regard to the first principal component (PC1) of Cavalli-Sforza and his colleagues seen in Figure 4(a) above. Although these workers have been

hesitant in associating the farming 'demic diffusion' which they believe to be documented by PC1 with the spread of Proto-Indo-European, at first preferring to follow the Gimbutas theory and associate that spread instead with PC3, other commentators have suggested that the spread of Proto-Indo-European is indeed documented by the PC1 map interpreted as showing the effects of farming dispersal.

The application of mitochondrial DNA analysis has, however, caused a fundamental reappraisal of the situation. The network-joining method of H.-J. Bandelt, which allows the recognition of locally occurring haplotypes and a calculation of the amount of time subsequent to an initial dispersal necessary to allow for the relevant mutations to occur in the area under study (in this case Europe), allowed Richards and his colleagues (Richards *et al.* 1996) to make the controversial suggestion that the greater part of the mitochondrial DNA variation seen in Europe today derives from colonization episodes taking place during the Late Pleistocene period. They suggested that only a limited proportion, represented by haplogroup J, is to be assigned to a colonization event as recent as 8500 years ago (see Table 1).

More recent work (Sykes 1999) has refined the chronology (see Table 1). And although these findings have been criticized by Cavalli-Sforza & Minch (1997) and by Barbujani *et al.* (1998), their criticisms have been answered (Richards *et al.* 1997; Richards & Sykes 1998) and the findings are in general supported by the work of Torrioni and his colleagues (Torrioni *et al.* 1998) and through Y-chromosome work also (Malaspina *et al.* 1998). In addition, the recent paper by Torrioni and his colleagues (Torrioni *et al.* 1998) has announced an important new finding: the recognition of a mitochondrial DNA haplogroup, haplogroup V, which apparently originated in the Iberian peninsula some 15,000 years ago. Its subsequent wide dispersal along the Atlantic coast and up to northeastern Europe would be associated with the repopulation of the north European plain after the Late Glacial Maximum in the centuries after 15,000 years ago (Fig. 6), and indeed may underlie the pattern seen in the map of the second principal component of Cavalli-Sforza and his colleagues (Fig. 4b, above).

These findings, and there are others on a similar timescale from other parts of the world, suggest that many of the haplogroup frequencies seen in the world today were determined in large part already during the Late Pleistocene period. In addition, at the extreme north of Eurasia and of America subsequent climate-related population movements will

have had a major impact, for instance upon the Na-Dene and Eskimo-Aleut populations of North America, upon the populations of Siberia and upon the Uralic-speaking populations of northern Europe.

The careful investigation of these matters depends upon precise and well-dated climatic data, which are now becoming available. For northern Europe there is a recent article which sets out the radiocarbon evidence for the late glacial recolonization of northern Europe (Houseley *et al.* 1997) and for North America the picture is also becoming clearer (Rogers *et al.* 1990).

These very recent finds open an entire new vista for our understanding of the population history of the world. For it was not widely realized that so much of the demographic history of the Late Pleistocene period is still inscribed in the genetic

record of Europe and beyond, as reflected in the genes of present-day populations. In the years ahead it is likely that the archaeological record, controlled by radiocarbon dating, and the molecular genetic record, supported by the internal, network-derived chronology, will be brought into closer and closer adjustment to provide a well-documented demographic history, using Y-chromosome studies (see for instance Santos *et al.* 1999 for North America and Siberia) as well as mitochondrial DNA. Subsequent population bottlenecks will have determined the population composition for some areas, such as Finland. But in general it may turn out that most post-Pleistocene population events will appear as relatively minor modulations upon the pattern established already in the Late Pleistocene, as the frequency of only 20 per cent for the Neolithic dispersal

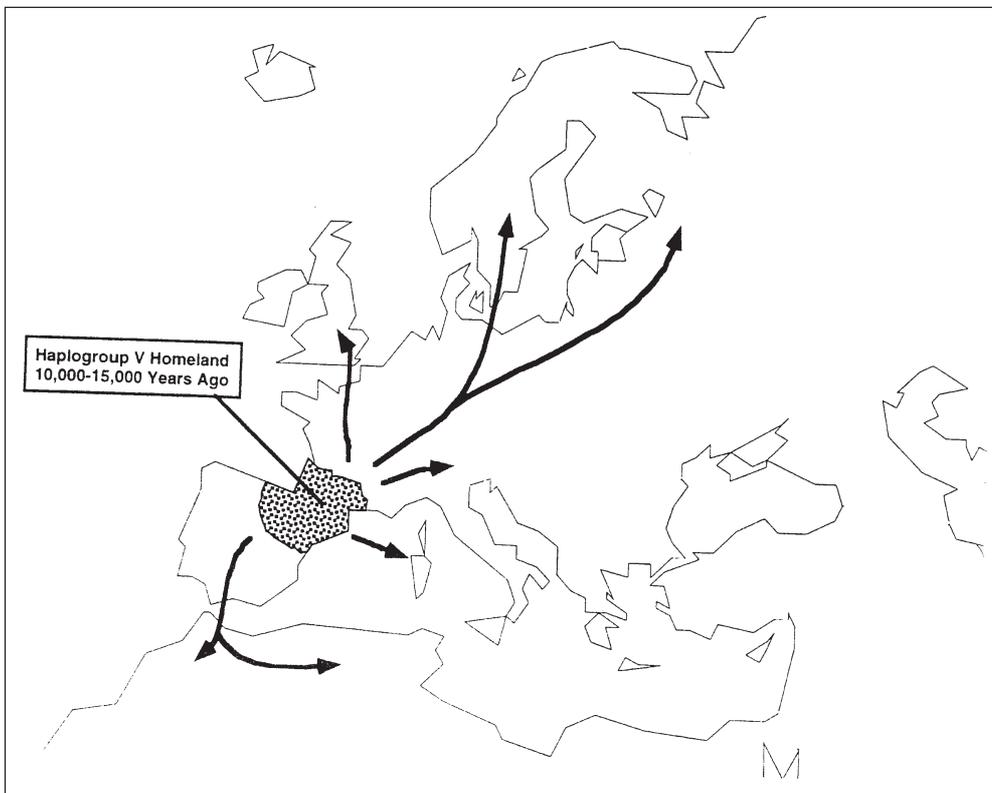


Figure 6. The suggested homeland for mitochondrial DNA haplogroup V, in the population refugium in Iberia during the Late Glacial Maximum cold spell, as proposed by Torrioni *et al.* (1998, 1148). It would be diffused from its homeland, as indicated, with the dispersal of population from the refugium over the period from 15,000 to 10,000 years ago.

Table 1. A summary of the three main waves of European colonization, as derived from mitochondrial DNA analysis. The letters refer to the classification of haplotypes. It should be noted that the Neolithic contribution is estimated at c. 20 per cent of the total. (After Sykes 1999, 137.)

Component	Dates (BP)	Main associated clusters	Contribution to modern gene pool
Neanderthal	300,000	unclassified	0%
Early Upper Palaeolithic	50,000	U5	10%
Late Upper Palaeolithic	11,000–14,000	H, V, I, W, T, K	70%
Neolithic	8500	J (+ more of H, T, K?)	20%

to Europe discussed above suggests.

During September 1999 a Symposium on 'Human Diversity in Europe and Beyond: Retrospect and Prospect' (The Third Biennial Euroconference of the European Human Genome Diversity Project) was held by the McDonald Institute and the resulting volume will be published with the aid of the Sloan Foundation (Renfrew & Boyle in prep.). Among the contributions were several bearing upon these issues (e.g. Macaulay 1999). It is only now that detailed sampling is being undertaken in Europe for mitochondrial DNA and Y-chromosome analysis. Although more data are needed, this gives clear support to the view of Zvelebil (1995) that the process of demic diffusion, while it may have been applicable to the Balkans and the *Linearbandkeramik* culture of the Danube, did not operate so significantly to the west or north of these limits. Beyond that it is more appropriate to think in terms of contact-induced language change with only very limited gene flow. This is one case where the molecular genetic evidence has a significant impact upon our understanding of the language prehistory. Another is the significant frequency in the Orkney Islands of Y chromosomes which may have reached the area from Scandinavia during the first millennium AD, against a background which, like that of Ireland, is otherwise typical for the western Celtic fringe of Europe (see Wilson *et al.* 1999). This can readily be associated with the Viking presence in Orkney and the complete replacement there of Pictish or Celtic speech by the Orkney Norn, of Scandinavian origin (Marwick 1929).

Reference has not been made here to the use of Ancient DNA which will, in the future, be increasingly informative. Already there have been applications relevant to the colonization of the Americas (Stone & Stoneking 1998; 1999). The most remarkable contribution, however, has been the recovery and analysis of mitochondrial DNA from Neanderthal remains some 40,000 years old (Krings *et al.* 1997). The findings seem to document clearly that the Neanderthals are not likely to have contributed genetically to the modern population of Europe, nor indeed of other parts of the world.

The big question

A broad picture is now beginning to emerge, from the comparison of the existing archaeological data with the new evidence now becoming available from molecular genetics, of the human population history of the world. For some years the outlines of the story based upon mitochondrial DNA have been becom-

ing clearer (Cann *et al.* 1987; Torroni *et al.* 1996), and the picture from Y-chromosome analysis can now at last be discerned (Underhill 1999). Moreover the processes and events can be dated, at least approximately, by the Bandelt network-joining approach, on internal criteria, without reliance on archaeological datings.

In Europe it is now possible to discern several successive phases in the population history:

- A. the first *sapiens* population episode in the Late Pleistocene *c.* 40,000 years ago;
- B. the retreat to southern refugia during the Late Glacial Maximum of *c.* 18,000 to 15,000 BC;
- C. the final Pleistocene from *c.* 15,000 BC and the retreat of the ice;
- D. the Holocene from *c.* 8000 BC prior to the advent of farming;
- E. the advent of farming during the Holocene;
- F. demographic processes during the Holocene subsequent to the advent of farming;
- (G). the Roman Empire;
- (H). post-Roman Europe from *c.* AD 400.

With the aid of mtDNA and Y-chromosome analysis, a comparable table can be established for any part of the world, although of course phases G and H (the Roman Empire and its aftermath) were an episode local to Europe, North Africa and western Asia. The advent of a farming economy was of course a late development in some areas such as Australia.

The *big question*, so far as the Prehistory of Languages is concerned, is the extent to which the linguistic history of the world can be related to world population history, as seen in these very broad terms.

For the Americas, one might argue that the first colonization came either during phase A or phase C: that is precisely the controversy currently under way both in the prehistoric archaeology of the Americas and in the debate between Nichols and Nettle in the historical linguistics. Certainly most observers would place the Na-Dene (or proto-Na-Dene) colonization in phase C, and the Eskimo-Aleut arrival in phase D (or possibly in phase C). The development of farming would then be responsible for the spread of certain language families such as the Maya and the Uto-Aztecan in phase E, and political events involving the rise of state society in Mesoamerica and in South America and then the emergence of the Aztec and Inca Empires would relate to phase F.

In Europe one could hypothesize that the demographic events of phases A and C (with D also) established the basis for the Basque language of northern Spain, seen as a relic of what was once a more widely distributed linguistic unit. Following the farming/language dispersal explanation the dis-

tribution of the Indo-European languages of Europe would have been effected during phase E, and those of Iran and India during some élite dominance episode of phase F. Obviously the distribution of the Romance languages would be a product of phase G, while the Germanic and Slavonic expansions (with the intrusion of the non-Indo-European Magyar tongue in Hungary) would be product of phase H.

But these equations in Europe would not all be accepted by the majority of linguists. As we saw earlier, issues of time depth are involved here. This is a decisive question which I first raised in 1987 with reference to Europe. It has been well identified by Dixon (1997) and firmly tackled by Nettle (1999a). It seems that we are close to the nub of the matter of knowability. For, on a pessimistic view, there are no adequate ways of establishing an equivalence between linguistic events and those of the kind under discussion here. Moreover, the generally accepted linguistic chronologies are much shorter than those envisaged here. If we follow these traditional views it is difficult to see how there can be further progress, and we have already reached the limits of the knowable. If, however, we take what I would regard as the optimistic view, and accept the proposals made earlier for greater time depths, we then reach the possibility for some correlations or equivalencies between linguistic events and climatic sequences and socio-economic processes. Indeed as we shall see again in the next section, there is at least one general model, of wide applicability, which offers a path towards a more general understanding of the patterns of linguistic diversity seen in the world. To imagine such a model, however, and even to show how its application might explain certain language distributions, is certainly not the same as establishing it. As the remarks on knowability in an earlier section made clear, there is no direct way of testing these matters. All that one can do is to frame the general proposition. If it can be shown to work well in a number of cases, and if it is not shown to be erroneous in others, and if it then goes on to forecast patterns not hitherto discerned, then it might prove acceptable as an explanation placing these matters in the realm of the knowable.

The model advanced in the next section is one such general model. It may not prove valid or acceptable, and other general models of a different kind may be developed in the future — indeed it could be argued that Johanna Nichols (1992) has already begun the task. But if no such general model can be developed and then found useful, then it

seems to me doubtful that much further progress could be made towards an understanding of the prehistory of languages.

The language/farming dispersal model and other 'punctuations'

Spread zone versus mosaic zone

Several linguists have remarked on the marked difference in the nature of the geographical distributions of language families, and Austerlitz (1980) has contrasted patterns of language-family density. In 1992 I suggested that there were contrasting geographical patterns (Renfrew 1992b, 59) when a region had undergone language displacement as a result of some dispersal process (Fig. 7, case A), or when, after initial colonization, there was a long period of stability resulting in a mosaic of small language units (Fig. 7, case B), but the term 'mosaic zone' or 'retention zone' may be preferable to 'residual zone'.

This distinction is very similar to the one drawn by Nichols (1992) between what she terms linguistic 'spread zones' and 'residual zones'. When a language family shows the 'spread zone' pattern, it displays what one may term a low genetic density (i.e. only a limited number of linguistic units unrelated by descent from a common language ancestor) over a sizeable area, and with a relatively shallow time depth. The 'mosaic zone' (or 'residual zone') pattern shows more language families, greater linguistic diversity within the language family, and greater antiquity of the linguistic stocks there. The north Caucasus is a good example of a 'residual zone', and is sometimes regarded as a linguistic refugium, and north Australia offers a further example.

It should be noted that pattern B, the mosaic zone configuration, is seen both with hunter-gatherers and with agriculturalists in those cases where the agricultural economy does not seem to be the result of an agricultural dispersal but may be regarded as indigenous, as in the case of New Guinea with its very early horticulture and perhaps in regions of South America. But the interesting point is that the spatial scale of the language unit differs in the two cases, with a hunter-gatherer linguistic unit occupying a much larger area than an agricultural or horticultural one. This is no doubt related to the much greater population density in the agricultural case, so that the size of the speech community in the two cases may be more similar than the territorial extent. This is a point addressed much more fully by Nettle (1999a).

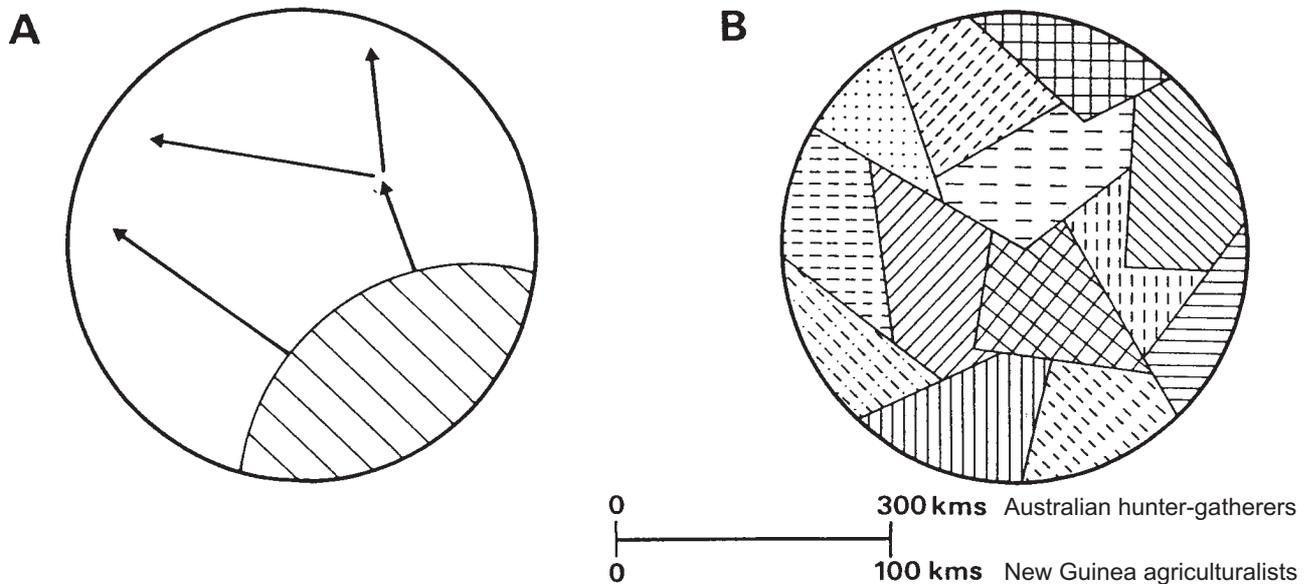


Figure 7. Spread zone (A) versus mosaic zone (B). Contrasting language distribution patterns seen when there has been a language replacement following a rapid dispersal process or punctuation (case A), or when, after initial colonization, there has been a long period of uninterrupted local divergence (case B) resulting in a mosaic of small language units. In B one may compare the different scales for the language map for hunter gatherers and for the agriculturalists whose agricultural economy is indigenous, resulting in both cases in a mosaic of small language units. (The language units are for convenience separated by lines: this is not intended to suggest that they are discrete or bounded entities nor to deny the existence of intermediate dialects.) (From Renfrew 1992b, 59.)

As I observed in 1992:

If we try to imagine a world whose language distribution was the product of initial colonisation, followed by convergence and divergence effects, but without the effects of agricultural dispersal or elite dominance, might we perhaps imagine the linguistic configuration of Australia generalised for those areas with a hunter-gatherer economy, and that of New Guinea for those with an indigenous agricultural economy? (Renfrew 1992b, 60).

In such a case as this the pattern does not in itself tell us whether the various languages involved are genetically related or not. For in the case of a mosaic zone, the time depth involved may be so great that through lapse of time most valid markers of relatedness, even if they existed initially, would have disappeared — we may be talking here of periods of stability enduring well over 10,000 years. And as Dixon has argued, the convergence effects operating over so long a period of time may produce a linguistic area or *Sprachbund* effect, with similarities developing between the languages which are not the product of genetic relatedness but simply of long-term interaction (on the model of Trubetzkoy (1939)).

On the other hand, the configuration seen in pattern A suggests a relatively recent dispersal, in

Dixon's recent terminology a punctuation. This may be due to a dispersal on the subsistence/demography model, or on the elite dominance model, or on the system collapse model (see Renfrew 1989a). Or, as I have now come to realize, it may be the product of contact-induced language change (see below). In most cases, as Peter Bellwood has also suggested (Bellwood 1996; 1997), it is likely to be the product of farming dispersal. Other applications of the model have been suggested for the consequences of rice cultivation in southeast Asia (Glover & Higham 1996; Higham 1996), for the Afroasiatic languages (Diakonoff 1998) and for some of the language families of Africa (Phillipson 1977; Ehret 1998).

The important implication of the evident distinction between 'spread zones' and mosaic zones' (or 'residual zones') is that they may in many cases be recognized by inspecting maps of language distributions, along with a basic classification of the languages in question into language families (or as isolates), following the work of linguists in the field. It is possible in this way to suggest that in many cases the language distribution in a spread zone is likely to have been the result of a dispersal phenomenon, very possibly a farming dispersal, although that needs to be evaluated by closer examination

including a consideration of the archaeological record. In cases where a mosaic zone is recognized, the pattern may well prove to be one of very early initial colonization, perhaps of the order of 15,000 years ago, followed by stability and local divergence. It is, of course, also possible that what may at first appear as a spread zone represents a point at the extreme right of Nettle's figure 3(b), where very long-term convergence processes have taken over in the manner proposed by Dixon, producing a language 'family' (but not in the genetic sense) through convergence. This, of course, is Dixon's proposal for the Pama-Nyungan language 'family' of Australia.

Mechanisms of farming dispersal

One of the central ideas of this article, as earlier of my book *Archaeology and Language*, is thus that in many cases spread zones of the kind discussed above are the product of farming dispersals. Table 2 (from Renfrew 1992a, 456) summarizes the argument.

In the first full elaboration of this model (Renfrew 1987) I emphasized the 'demic diffusion' mechanism of Ammerman & Cavalli-Sforza (1973; 1984) with its dependence upon demic diffusion — that is to say, the gradual outflow of farmers from areas where farming was practised into neighbouring areas, a process in part the product of the greater population density which developed in farming areas. But Zvelebil's criticisms of this model (Zvelebil & Zvelebil 1988; Zvelebil 1995) long ago persuaded me that the model was applicable, as noted above, only to southeast and central Europe, not to the north and

west. Their observations seem to be supported by the molecular genetic data for north and west Europe, which show very low frequencies of the haplogroups, notably haplogroup J, which may be identified as the product of the gene flow process which accompanied the coming of farming to Europe. This process, although discernible through the methods of molecular genetics (Sykes 1999) was on a smaller scale than Ammerman & Cavalli-Sforza (1973) and then Menozzi *et al.* (1978) had suggested.

It is appropriate to consider more carefully how the alternative approach, one of acculturation or 'contact-induced language change' would operate in practice. Zvelebil (1996, 325) has reformulated his 'availability model' (Zvelebil 1986, 12) to deal with such cases, and it is shown here with modifications of my own as Figure 8.

But we need to learn much more about the conditions, in social as well as in linguistic terms, under which one community comes to adopt the language of a neighbouring community. Ehret (1988) has interesting remarks to make about precisely the general case which concerns us most, the adoption by hunter-gatherers of the language of their farming neighbours: in some cases this occurs without a change in the economic basis of their own way of life. But, in the longer term, that perhaps rather surprising early adoption of the incoming language sometimes prior to the adoption of the new economy may simply be a prelude to that change also. The point here, of course, is that such change may be effected with only very limited gene flow, and hence

Table 2. *The language/farming dispersal model: note the two mechanisms proposed.*

A nuclear area is defined, supporting initially a specific range of wild plants (and sometimes animals) which later proved amenable to domestication. The farming 'package' of plants (and, where appropriate, animals), along with the appropriate exploitative techniques, becomes an expansive one dependent upon three factors:

- I *suitability for transplantation into new ecological niches* of the plants (and animals), when sustained with the appropriate exploitative technology by the accompanying human population, with propagation (i.e. seeding/planting or controlled breeding), protected growth (by weeding and manuring or controlled feeding, e.g. by transhumance) and organized harvesting (or culling);
- II *increased birth rate* and reduced rate of human infant mortality, and sometimes increased post-infantile life expectancy, associated with aspects of the new subsistence regime. These accompany the *sedentary* life which farming facilitated;
- III *greater intensity of production* as measured in terms of food (calories) per unit area, premitted by the new economy. Agricultural economies, even of a simple and non-intensive nature, are characteristically fifty times more productive in this sense than mobile hunter-gather economies, or have the capacity to be so.

In favourable cases the language or languages of the nuclear area are transmitted along with the plant and animal domesticate either through *demic diffusion* of the farming population (the 'wave of advance model'), or through adoption by local hunter-gatherer groups of the new language along with the new agricultural economy (*acculturation*: the 'availability model'). The genetic effects of the two mechanisms are significantly different.

leave little or no trace in the molecular genetic record as reflected in the gene frequency distributions of later generations. In many of the cases studied, the economy of the farmers carries with it, in the eyes of the hunter-gatherers, a higher degree of prestige than their own way of life. This may mean that it is difficult for male hunter-gatherers to seek wives from the farming communities, whereas it seems fairly common that the agriculturist communities should take wives from the adjacent hunter-gatherer groups. Such sexual or gender asymmetries are of course precisely what the matching techniques of mitochondrial (female lineage) and Y-chromosome (male lineage) analysis are equipped to analyze. We shall certainly learn additional information from the molecular genetic data about this issue in the next few years. The farming dispersal itself and its place of origin can usefully be investigated through the molecular genetics of domesticated plant (Heun *et al.* 1997) and animal (Bailey *et al.* 1996) species.

It is inevitable that language replacement without gene flow implies bilingualism, and this is a topic which Malcolm Ross (1997) has recently explored in terms of a social network model. But despite the insights from Ross, Ehret and Zvelebil, the circumstances in which hunter-gatherers accept the language and adopt the economy of their agricultural neighbours have yet to be more fully explored. The recent indication from molecular genetics that in some cases such linguistic replacements may have taken place without significant gene flow makes this an urgent topic for further study.

It seems possible, then, to suggest that the distribution of many of the world's spread-zone language families is the result of farming dispersal, and specifically those in Table 3. As noted above, the 'Altaic' family is not here assumed to be a valid genetic grouping, but the comments apply at least to its Turkic and Mongolian components, although in these cases the mechanism of 'farming dispersal' might be said to include also the development of

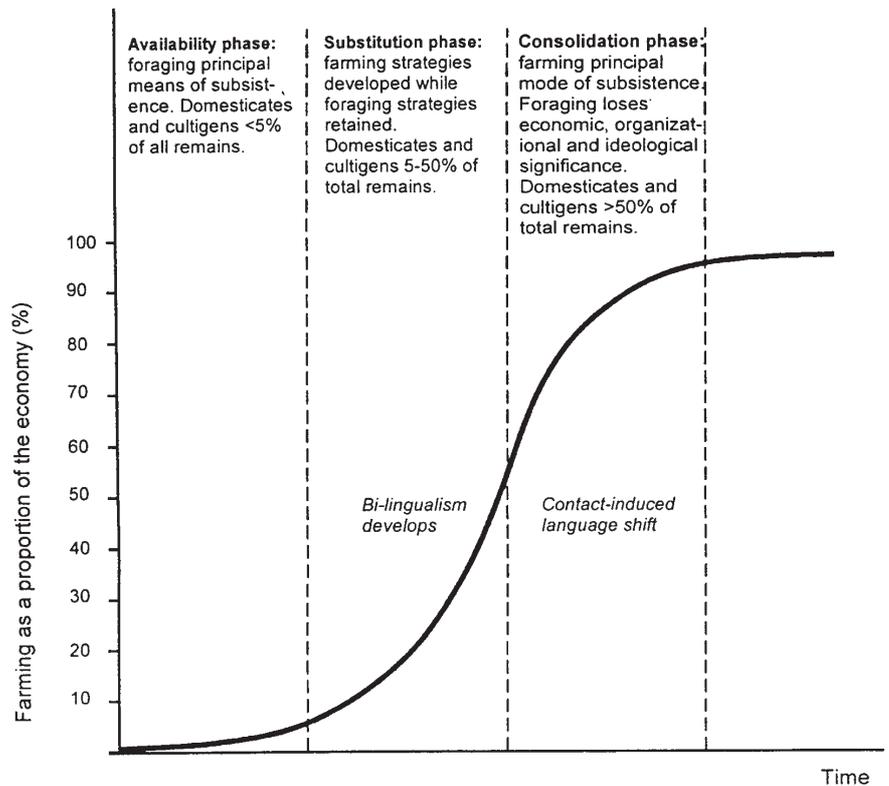


Figure 8. Linguistic adjacency acceptance. The adoption by hunter-gatherers of the language of neighbouring cultivators through contact-induced language shift. (Based on Zvelebil 1996, 325, with additions.)

nomad pastoralism. In the case of nomad pastoralists, however, the appropriate model may rather be one of élite dominance after the development of mounted warfare in the first millennium BC.

At the same time, as we have seen, it seems appropriate to suggest that many areas with mosaic-zone language distributions have not been subjected to a farming dispersal, but rather that the initial colonization took place during the Late Pleistocene period, and that there has been stability along with local divergence since that time.

The significance of the language/farming dispersal model may be gauged from the map seen in Figure 9, particularly when it is noted that the Uralic, Mongolian and Turkic families are not included on it. The Mongolian and Turkic languages are mentioned further in the next section.

These proposals are of course to be considered as tentative. It is hoped to explore them further at the McDonald Institute during the year 2001 in a projected Symposium on the language/farming dispersal hypothesis, to be organized jointly by Peter Bellwood and myself within the framework of the Prehistory of Languages Project.

Table 3. *Regions of early agricultural development and farming dispersals, with the language families whose spread-zone distributions may be ascribed to this factor. (From Renfrew 1998a, after Bellwood 1996.)*

Region of early agriculture	Associated language families
sub-Saharan Africa	Niger-Kordofanian
Southwest Asia	Elamo-Dravidian Indo-European Altaic Afro-Asiatic
China (north)	Sino-Tibetan
China (south)	Austroasiatic Austronesian Tai Hmong-Mien (Miao-Yao)
New Guinea	Many Papuan families
Mexico	Otomanguean Uto-Aztecan Mayan Mixe-Zoque
Andes/Upper Amazon	Chibchan/Paezan Qhechua/Aymara Arawakan Panoan

Mobility and pastoral nomadism

Farming dispersals offer the most obvious examples for the subsistence/demography model for language replacement. But it should not be forgotten that the development of pastoral nomadism, while it is often subsumed under the élite-dominance model (with the standard cliché image of the mounted warrior pastoralists carrying all before them) was an economic development before it developed further and came to represent also an innovation in warfare. It has recently been shown that the very high dates for the domestication of the horse and the development of pastoral nomadism have been set exaggeratedly early (Levine *et al.* 1999; Renfrew 1998b). Indeed it now seems that the sequence, in relation to the utilization of the horse in Eurasia, may be summarized as follows:

1. The *cart* is first seen in Europe *c.* 3000 BC, with solid wheels and drawn by oxen, in the Baden culture of the Middle Danube.
2. Horses are not seen in archaeological contexts in Central and western Europe at that time (although wild horses persisted since the Upper Palaeolithic). *Hippophagy*: in the Ukraine (Dereivka) horses were a major food resource from before

4000 BC: there is no secure evidence that they were domesticated at that time.

3. The *chariot horizon*, involving the horse-drawn chariot with spoked wheels, is seen in Europe and western Asia around 1600 BC (perhaps slightly later in China). The earliest known horse-drawn chariots are found in the Urals (e.g. Sintashta) *c.* 2000 BC. This is the first indication anywhere of the use of the horse in warfare.
4. The *warrior horseman* makes his appearance *c.* 1500 BC in the Eurasian steppes and *c.* 1200 BC in Europe and Western Asia. There are no earlier indications anywhere of fighting on horseback. Metal bits and harnesses are widely seen in Europe from *c.* 800 BC — the *horseman horizon*.
5. Fully *mobile nomad pastoralism* makes its first appearance on the Eurasian steppes not before the later second millennium BC and becomes established in the first millennium BC, about the time that horse-riding is first widely documented archaeologically.

It follows from this that the horse was not ridden for military purposes until towards the end of the second millennium BC. During the first millennium such use is documented historically for the first time with the case of the Scythians and Cimmerians described by the Greek historian Herodotus. But the horse was an important food resource in the steppe lands from a much earlier time, and then the basis for at least a partially nomadic existence for a couple of millennia well before its use in military conflict. It is possible to recognize here a further subsistence/demography model dependent less upon farming than on the development of mobile pastoralism. It should, however, be noted that the pastoral economy itself requires both animal domesticates derived ultimately from farming communities and plant foods often obtained through exchange with farmers (Khazanov 1984). Indeed, in discussing distributions of the Mongolian and Turkic language families (as well as the Indo-Iranian branch of Indo-European) it may be pertinent to consider also such developments rather than to rely exclusively upon the warlike character of warrior nomads and episodes of élite dominance to effect language replacement. A Symposium was held at the McDonald Institute in January 2000 on the theme of the exploitation of the Eurasian steppes in later prehistoric times, and the papers presented are expected to be published in a forthcoming volume (Renfrew *et al.* in prep.). The significance for historical linguistics of the development of Eurasian pastoral nomadism has often been misunderstood. The Symposium gave useful insight both into the

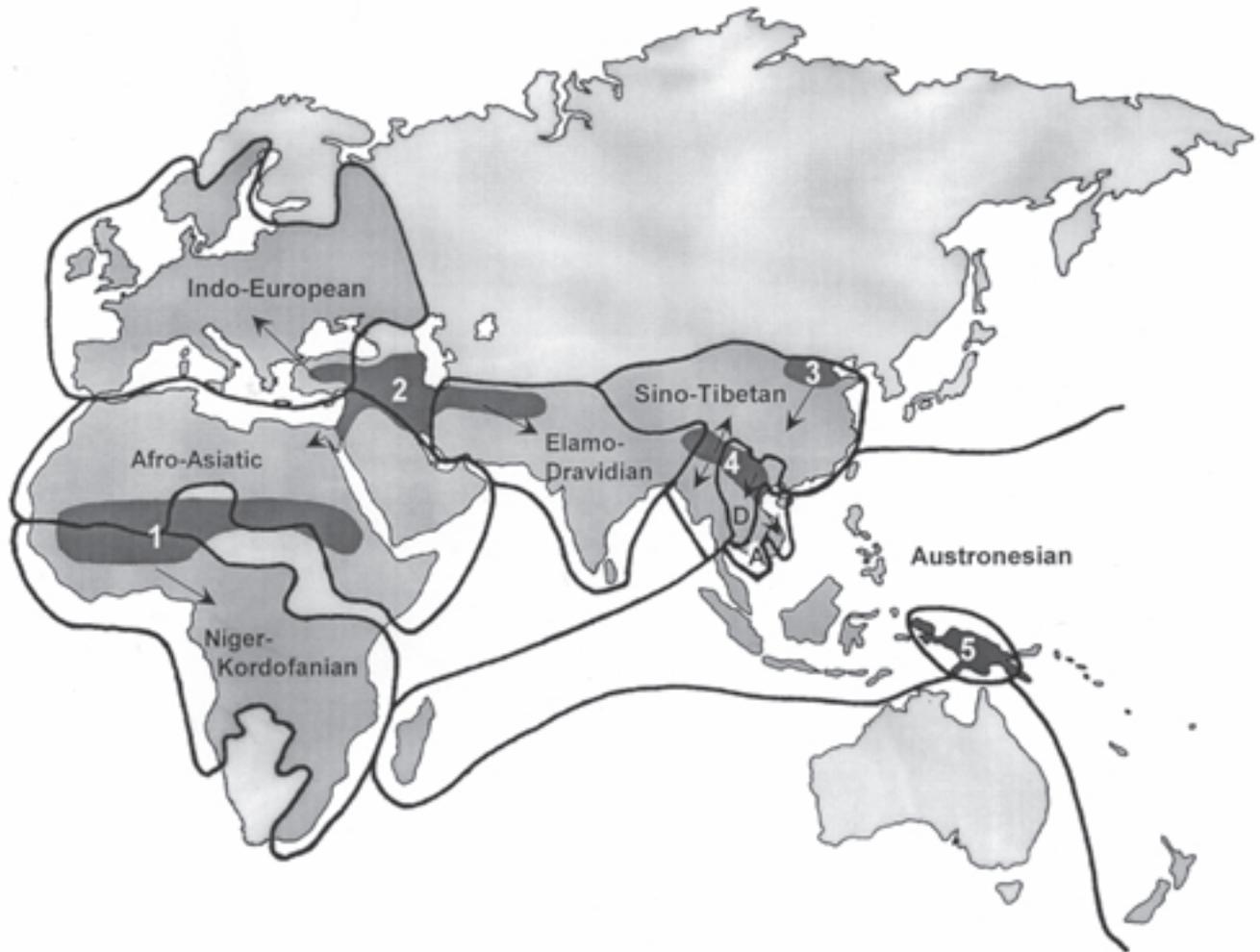


Figure 9. Map for areas of primary domestication of selected principal food plants and distributions of selected language families whose spread-zone distributions are here ascribed to agricultural dispersal. The areas of primary crop domestication are numbered: 1) sorghum/millet; 2) wheat/barley; 3) millet; 4) Asian rice; 5) taro/sweet potato. Southeast Asian language families indicated by letters are D: Daic; A: Austroasiatic. (Note that the Indo-European and Elamo-Dravidian distributions reflect the hypothetical language dispersals and do not show the subsequent spread of the Indo-Iranian languages. The agricultural dispersal underlying the Austronesian family distribution is believed to have originated in southeast Asia but was based subsequently on yam, taro and tree fruits.) (From Renfrew 1998a.)

process and its consequences.

From the foregoing it may be concluded that, when a linguistic spread zone is observed, it will in many cases be the result of a farming dispersal process.

Of course it should not be forgotten that initial colonization will itself have generated a spread-zone effect, with local divergence generating a series of new languages, initially all recognizable as descended from the proto-language of the initial spread. But then, as both Nettle (1999a) and Nichols (1992) agree, after the passage of between 10,000 and 20,000 years (see Fig. 3) the divergence processes will have proceeded so far that the familial genetic relationships are no longer visible, and what at first were recog-

nizable as related languages will have turned into apparently unrelated languages and indeed language families (through further splitting and divergence). That is how the configuration shifts from that seen in Figure 7 (A) to that of Figure 7 (B).

A recent initial colonization will thus give a spread-zone effect. This can be seen in the north of the northern hemisphere where recent, climate-related colonizations, such as that of the Eskimo-Aleut languages, are represented. It could also be argued that the spread of the Polynesian languages was an initial colonization of the Pacific islands (which indeed it was), but it was of course, at the same time, a farming dispersal — perhaps the only case in world history

where a farming dispersal has taken place in a territory not previously occupied by hunter-gatherers.

A linguistic spread zone can also be created by an episode of élite dominance. As we have noted above, such is the explanation usually offered for the distribution of the Indo-Iranian languages of the Indo-European family, and for the distributions of the Turkic and Mongolian languages also. But none of these processes is yet well understood historically, and the peaceful spread of nomad pastoralism may also have played a role.

Finally, the spread of English, Spanish and other Indo-European languages during the past five or six centuries over much of the world may at first seem an example of élite dominance, through the processes of imperial colonization. But in reality these languages have persisted most successfully when they have been accompanied by those basic domesticates — wheat, barley, cattle and sheep — which accompanied the initial Proto-Indo-European spread from Anatolia to Europe (Diamond 1997).

Conclusions

From the foregoing discussions it may be permissible to formulate a number of general conclusions, most of them of a provisional nature:

1. *Historical linguistics* is now developing methodologies for addressing linguistic behaviour in the 'real world' over long time-periods — in effect the development of sociolinguistics far beyond the life-span of the single observer. This is likely to be the single most important path for further insights into the origins of linguistic diversity.
2. Precise methods for evaluating *time depth* are not yet available in historical linguistics, but it seems that no absolute time barrier can be established, and time depths in excess of 10,000 years for some language families (e.g. Afroasiatic) have been plausibly argued.
3. The broad question of *macrofamilies* is not yet resolved. It is possible that the 'Nostratic' or 'Eurasian' and the 'Sino-Caucasian' macrofamilies will yet find broader agreement, but this will depend upon a more coherent analysis and presentation of the morphological correspondences between constituent families within the alleged macrofamily, and the further development of probabilistic statistical criteria for evaluating the lexical evidence.
4. The contributions of *prehistoric archaeology* have not yet been fully exploited: the mechanisms of farming dispersal remain controversial, episodes of alleged élite dominance are not yet well documented (e.g. for the Indo-Iranians in India and Iran) and the late emergence of full nomad pastoralism and the use of the horse in warfare (during the later second and early first millennia BC in Eurasia) is only now being appreciated.
5. *Molecular genetics* in its contribution to human population history is at an early stage. The Upper Palaeolithic picture is only now being elucidated. However the pace of progress is rapid in this fast-developing field.
6. The *spatial parameters* of language families including those suggestive of spread episodes have not yet been systematically studied in a comparative framework: the language/farming dispersal hypothesis remains a strong contender for the explanation of many spread episodes.
7. The issue of *knowability* is less likely to be decided on *a priori* criteria than by the success or failure of endeavour in the above areas. The matter is likely to be much clearer within a decade.

Colin Renfrew

McDonald Institute for Archaeological Research

Downing Street

Cambridge

CB2 3ER

Note

1. This article is based upon the Eleventh McDonald Lecture, delivered in Cambridge on 17th November 1999. Grateful acknowledgement is made to the Alfred P. Sloan Foundation of New York for generous funding to the Prehistory of Languages Project.

Acknowledgements

For permission to reproduce illustrations thanks are due to Luca Cavalli-Sforza and Princeton University Press (Figs. 4 & 5), Daniel Nettle and Oxford University Press (Fig. 3), Antonio Torroni and University of Chicago Press (Fig. 6), Marek Zvelebil (Fig. 8), and Edinburgh University Press (Fig. 1).

References

- Ammerman, A.J. & L.L. Cavalli-Sforza, 1973. A population model for the diffusion of early farmers in Europe, in *The Explanation of Culture Change: Models in Prehistory*, ed. C. Renfrew. London: Duckworth, 343–58.
- Ammerman, A.J. & L.L. Cavalli-Sforza, 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton (NJ): Princeton University Press.

- Anttila, R., 1989. *An Introduction to Historical and Comparative Linguistics*. 2nd edition. Amsterdam: John Benjamin.
- Austerlitz, R., 1980. Language family density in North America and Eurasia. *Ural-Altäische Jahrbücher* 52, 1–10.
- Bailey, J.F., M.B. Richards, V.A. Macaulay, I.B. Colson, I.T. James, D.G. Bradley, R.E.M. Hedges & B.C. Sykes, 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Philosophical Transactions of the Royal Society Series B* 263, 1467–73.
- Barbujani, G., 1991. What do languages tell us about human microevolution? *Trends in Ecology and Evolution* 6, 151–6.
- Barbujani, G., 1997. DNA variation and language affinities. *American Journal of Human Genetics* 61, 1011–14.
- Barbujani, G. & R.R. Sokal, 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the USA* 87, 1816–19.
- Barbujani, G., N.L. Oden & R.R. Sokal, 1989. Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology* 38, 376–89.
- Barbujani, G., A. Pilastro, S. de Domenico & C. Renfrew, 1994. Genetic variation in North Africa and Eurasia: Neolithic demic diffusion v. Palaeolithic colonisation. *American Journal of Physical Anthropology* 95, 137–54.
- Barbujani, G., G. Beretorelle & L. Chikhi, 1998. Evidence for Palaeolithic and Neolithic gene flow in Europe. *American Journal of Human Genetics* 62, 488–91.
- Bateman R., I. Goddard, R. O'Grady, V.A. Funk, R. Mooi, W.J. Kress & P. Cannell, 1990. Speaking with forked tongues: the feasibility of reconciling human phylogeny and the history of language. *Current Anthropology* 31, 1–24.
- Bell, J.A., 1994. *Reconstructing Prehistory: Scientific Method in Archaeology*. Philadelphia (PA): Temple University Press.
- Bellwood, P., 1989. The colonisation of the Pacific: some current hypotheses, in *The Colonisation of the Pacific: a Genetic Trail*, eds. A.V.S. Hill & W. Serjeantson. Oxford: Oxford University Press, 1–59.
- Bellwood, P., 1991. The Austronesian dispersal and the origin of languages. *Scientific American* 265(1), 88–93.
- Bellwood, P., 1996. The origins and spread of agriculture in the Indo-Pacific region: gradualism and diffusion or revolution and colonization?, in Harris (ed.), 465–98.
- Bellwood, P., 1997. Prehistoric cultural explanations for the existence of widespread language families, in McConvell & Evans (eds.), 123–34.
- Binford, L.R., 1999, Time as a clue to cause? *Proceedings of the British Academy* 101, 1–36.
- Blench, R. & M. Spriggs (eds.), 1997. *Archaeology and Language*, vol. I: *Theoretical and Methodological Orientations*. London: Routledge.
- Blench, R. & M. Spriggs (eds.), 1998. *Archaeology and Language*, vol. II: *Archaeological Data and Linguistic Hypotheses*. London: Routledge.
- Blench, R. & M. Spriggs (eds.), 1999a. *Archaeology and Language*, vol. III: *Artefacts, Language and Texts*. London: Routledge.
- Blench, R. & M. Spriggs (eds.), 1999b. *Archaeology and Language*, vol. IV: *Language Change and Cultural Transformation*. London: Routledge.
- Blust, R., 1996. Beyond the Austronesian homeland: the Austric hypothesis and its implications for archaeology, in *Prehistoric Settlement of the Pacific*, ed. W.H. Goodenough. (Transactions of the American Philosophical Society 26(5).) Philadelphia (PA): American Philosophical Society, 117–40.
- Boyle, K. & P. Forster (eds.), 1999. *Human Diversity in Europe and Beyond — Retrospect and Prospect: Final Program and Abstracts*. (The Third Biennial Euroconference of the European Human Genome Diversity Project 9–13 September 1999.) Cambridge: McDonald Institute for Archaeological Research.
- Campbell, L., 1998. *Historical Linguistics: an Introduction*. Edinburgh: Edinburgh University Press.
- Cann, R.L.M., M. Stoneking & A.C. Wilson, 1987. Mitochondrial DNA and human evolution. *Nature* 325, 31–6.
- Cavalli-Sforza, L.L., 1988. The Basque population and ancient migrations in Europe. *Munibe Antrop. Arqueol.* (Supl.) 6 (San Sebastian), 129–37.
- Cavalli-Sforza L.L. & F. Cavalli-Sforza, 1996. *The Great Human Diasporas: the History of Diversity and Evolution*, trans. S. Thorne. Wokingham: Addison-Wesley.
- Cavalli-Sforza, L.L. & E. Minch, 1997. Palaeolithic and Neolithic lineages in the European gene pool. *American Journal of Human Genetics* 61, 247–51.
- Cavalli-Sforza, L.L., A. Piazza, P. Menozzi & J. Mountain, 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. *Proceedings of the National Academy of Sciences of the USA* 85, 6002–6.
- Cavalli-Sforza, L.L., P. Menozzi & A. Piazza, 1994. *The History and Geography of Human Genes*. Princeton (NJ): Princeton University Press.
- Childe, V.G., 1926. *The Aryans: a Study of Indo-European Origins*. London: Kegan Paul, Trench & Trubner.
- Childe, V.G., 1929. *The Danube in Prehistory*. Oxford: Clarendon Press.
- Clark, J.G.D., 1965. Radiocarbon dating and the spread of the farming economy. *Antiquity* 39, 45–8.
- Diakonoff, I., 1988. Review of C. Renfrew: *Archaeology and Language*. *Annual Review of Armenian Linguistics* 9, 79–87.
- Diakonoff, I., 1990. Language contacts in the Caucasus and the Near East, in Markey & Greppin (eds.), 53–65.
- Diakonoff, I., 1998. The earliest Semitic society. *Journal of Semitic Studies* 43, 209–17.
- Diamond, J., 1997. *Guns, Germs and Steel: the Fates of Human Societies*. New York (NY): Norton.
- Dixon, R.M.W., 1994. *Ergativity*. Cambridge: Cambridge University Press.

- Dixon, R.M.W., 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Dolgopolsky, A., 1973. Boreisch — Ursprache Urasiens? *Ideen des exacten Wissens, Wissenschaft und Technik in der Sowjetunion* 73(1), 19–30.
- Dolgopolsky A., 1998. *The Nostratic Macrofamily and Linguistic Palaeontology*. (Papers in the Prehistory of Languages.) Cambridge: McDonald Institute for Archaeological Research.
- Durie M. & M. Ross (eds.), 1996. *The Comparative Method Revisited: Regularity and Irregularity in Language Change*. Oxford: Oxford University Press.
- Ehret, C., 1988. Language change and the material correlates of languages and ethnic shift. *Antiquity* 62, 564–74.
- Ehret, C., 1998. *An African Classical Age: Eastern and Southern Africa in World History, 1000 BC to AD 400*. Oxford: J. Currey.
- Evans, J.D. & C. Renfrew, 1968. *Excavations at Saliagos near Antiparos*. London: British School of Archaeology at Athens/Thames & Hudson.
- Excoffier, L.B., B. Pellegrini, A. Sanchez-Masas, C. Simon & L. Langaney, 1987. Genetics and the history of sub-Saharan Africa. *Yearbook of Physical Anthropology* 30, 151–94.
- Forster, P., R. Harding, A. Torroni & H.-G. Bandelt, 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genetics* 59, 935–45.
- Fox, A., 1995. *Linguistic Reconstruction: an Introduction to Theory and Method*. Oxford: Oxford University Press.
- Gimbutas, M., 1973. The beginning of the Bronze Age in Europe and the Indo-Europeans 3500–2500 BC. *Journal of Indo-European Studies* 1, 163–214.
- Gimbutas, M., 1997. *The Kurgan Culture and the Indo-Europeanization of Europe*. Washington (DC): Institute for the Study of Man.
- Glover, I.C. & C.F.W. Higham, 1996. New evidence for early rice cultivation in South, Southeast and East Asia, in Harris (ed.), 413–42.
- Greenberg, J.H., 1963. *The Languages of Africa*. Stanford (CA): Stanford University Press.
- Greenberg, J.H., 1987. *Languages in the Americas*. Stanford (CA): Stanford University Press.
- Greenberg, J.H., in press. *Eurasiatic*. Stanford (CA): Stanford University Press.
- Greenberg, J.H. & M. Ruhlen, 1992. Linguistic origins of native Americans. *Scientific American* 267(5), 94–9.
- Greenberg, J.H., C.G. Turner & S.L. Zegura, 1986. The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Current Anthropology* 4, 477–97.
- Haley, J.B. & C.W. Blegen, 1928. The coming of the Greeks. *American Journal of Archaeology* 32, 141–54.
- Harris, D.R. (ed.), 1996. *The Origins and Spread of Pastoralism in Eurasia*. London: UCL Press.
- Heun, M., R. Schäfer-Pregl, D. Klawan, R. Castagna, M. Accerbi, B. Borghi & F. Salamini, 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* 278, 1312–14.
- Higham, C.F.W., 1996. Archaeology and linguistics in Southeast Asia: implications of the Austric hypothesis. *Bulletin of the Indo-Pacific Prehistory Association* 14, 110–18.
- Hill, E., 1999. Y chromosomes and Irish origins, in Boyle & Forster (eds.), 13–14.
- Hock, H.H., 1986. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Houseley, R.A., C.S. Gamble, M. Street & P. Pettitt, 1997. Radiocarbon evidence for the late glacial human recolonisation of northern Europe. *Proceedings of the Prehistoric Society* 63, 25–54.
- Illič-Svityč, V.M., 1989. The relationship of the Nostratic family languages: a probabilistic evaluation of the similarities in question, in Shevoroshkin (ed.), 111–13.
- Illič-Svityč, V.M., 1990. Nostratic reconstructions (translated by M. Kaiser), in *Proto-Languages and Proto-Cultures*, ed. V. Shevoroshkin. Bochum: Brockmeyer, 138–67.
- Jablonski, N.G. & L.C. Aiello (eds.), 1998. *The Origins and Diversification of Language*. San Francisco (CA): California Academy of Sciences.
- Jones, Sir W., 1807. Third anniversary discourse: 'On the Hindus', reprinted in *The Collected Works of Sir William Jones II*. London: John Stockdale, 23–46.
- Kaufman, T. & V. Golla, 2000. Language groupings in the New World: their reliability and usability in cross-disciplinary studies, in Renfrew (ed.), 47–57.
- Khazanov, A.M., 1984. *Nomads and the Outside World*. Cambridge: Cambridge University Press.
- Krings, M., A. Stone, R.-W. Schmitz, H. Krainitzki, M. Stoneking & S. Pääbo, 1997. Neanderthal DNA sequences and the origin of modern humans. *Cell* 90, 19–30.
- Lass, R., 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- Levine, M., Y. Rassamakin, A. Kislenco & N. Tatarintseva, 1999. *Late Prehistoric Exploitation of the Eurasian Steppe*. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.
- Macaulay, V., 1999. The emerging tree of west Eurasian mtDNAs, in Boyle & Forster (eds.), 19–20.
- McConvell, P. & N. Evans (eds.), 1997. *Archaeology and Linguistics: Aboriginal Australia in Global Perspective*. Melbourne: Oxford University Press.
- McMahon, A.M.S., 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
- McMahon, A.M.S. & R. McMahon, 1995. Linguistics, genetics and archaeology: internal and external evidence in the Amerind controversy. *Transactions of the Philological Society* 93, 125–225.
- Malaspina, P., F. Cruciani, B.M. Ciminelli, L. Terrenato et al., 1998. Network analyses of Y-chromosomal types in Europe, northern Africa and western Asia reveal specific patterns of geographic distribution. *American Journal of Human Genetics* 63(3), 847–60.
- Mallory, J., 1973. A short history of the Indo-European

- problem. *Journal of Indo-European Studies* 1, 21–65.
- Mallory, J.P., 1989. In *Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames & Hudson.
- Markey, T.L. & J.A.C. Greppin (eds.), 1990. *When Worlds Collide: the Indo-Europeans and Pre-Indo-Europeans*. Ann Arbor (MI): Karoma Publishers.
- Marwick, H., 1929. *The Orkney Norn*. Oxford: Oxford University Press.
- Mellars, P.A., 1998. Neanderthals, modern humans and the archaeological evidence for language, in Jablonski & Aiello (eds.), 89–116.
- Menozzi, P., A. Piazza & L.L. Cavalli-Sforza, 1978. Synthetic map of human gene frequencies in Europe. *Science* 210, 786–92.
- Merriwether, D.A., 1999. Freezer anthropology: new uses for old blood. *Philosophical Transactions of the Royal Society, Biological Sciences* 354, 121–9.
- Merriwether, D.A., B.M. Kemp, D. Crews & J.V. Neel, 2000. Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA, in Renfrew (ed.), 89–124.
- Mourant, A.E., A.C. Kopec & K. Domaniewska-Sobczak, 1976. *The Distribution of the Human Blood Groups*. Oxford: Oxford University Press.
- Nettle, D., 1999a. *Linguistic Diversity*. Oxford: Oxford University Press.
- Nettle, D., 1999b. Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proceedings of the National Academy of Sciences of the USA* 96, 3325–9.
- Nichols, J., 1992. *Linguistic Diversity in Space and Time*. Chicago (IL): University of Chicago Press.
- Nichols, J., 1997. The epicentre of the Indo-European linguistic spread, in Blench & Spriggs (eds.), 122–48.
- Noble, W. & I. Davidson, 1996. *Human Evolution, Language and Mind*. Cambridge: Cambridge University Press.
- Pedersen, H., 1931. *The Discovery of Language: Linguistic Science in the Nineteenth Century*. (1972 reprint.) Bloomington (IN): Indiana University Press.
- Phillipson, D.W., 1977. The spread of the Bantu languages. *Scientific American* 236(4), 106–14.
- Piggott, S., 1965. *Ancient Europe*. Edinburgh: Edinburgh University Press.
- Pinker, S., 1998. The evolution of the human language faculty, in Jablonski & Aiello (eds.), 117–26.
- Renfrew, C., 1964. Crete and the Cyclades before Rhamanthus. *Kretika Chronika* 18, 107–41.
- Renfrew, C., 1973a. Problems in the general correlation of archaeological and linguistic strata in prehistoric Greece: the model of autochthonous origin, in *Bronze Age Migrations in the Aegean*, eds. R.A. Crossland & A. Birchall. London: Duckworth, 263–76.
- Renfrew, C., 1973b. *Before Civilisation: the Radiocarbon Revolution and Prehistoric Europe*. London: Jonathan Cape.
- Renfrew, C., 1987. *Archaeology and Language: the Puzzle of Indo-European Origins*. London: Jonathan Cape.
- Renfrew, C., 1989a. Models of change in language and archaeology. *Transactions of the Philological Society* 87, 103–55.
- Renfrew, C., 1989b. The origins of Indo-European languages. *Scientific American* 260(10), 106–14.
- Renfrew, C., 1991. Before Babel, speculations on the origins of linguistic diversity. *Cambridge Archaeological Journal* 1(1), 3–23.
- Renfrew, C., 1992a. Archaeology, genetics and linguistic diversity. *Man* 27, 445–78.
- Renfrew, C., 1992b. World languages and human dispersals, a minimalist view, in *Transition to Modernity*, eds. J.H. Hall & I.C. Jarvie. Cambridge: Cambridge University Press, 11–68.
- Renfrew, C., 1996. Language families and the spread of farming, in Harris (ed.), 70–92.
- Renfrew, C., 1998a. The origins of world linguistic diversity: an archaeological perspective, in Jablonski & Aiello (eds.), 171–92.
- Renfrew, C., 1998b. All the King's Horses: assessing cognitive maps in later prehistoric Europe, in *Creativity in Human Evolution and Prehistory*, ed. S. Mithen. London: Routledge, 260–84.
- Renfrew, C., 1999. Time depth, convergence theory and innovation in Proto-Indo-European: 'Old Europe' as a PIE linguistic area. *Journal of Indo-European Studies* 27(3&4), 257–93.
- Renfrew, C. (ed.), 2000. *America Past, America Present: Genes and Languages in the Americas and Beyond*. (Papers in the Prehistory of Languages.) Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, C. & K. Boyle (eds.), in prep. *Molecular Genetics and Early Europe: Papers in Population Prehistory*. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, C. & D. Nettle, 1999. *Nostratic: Examining a Linguistic Macrofamily*. (Papers in the Prehistory of Languages.) Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, C., A. McMahon & L. Trask (eds.), in press. *Time Depth in Historical Linguistics*. (Papers in the Prehistory of Languages.) Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, C., M. Levine & K. Boyle (eds.), in prep. *Towards Pastoralism: Prehistoric Communities of the Eurasian Steppe*. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.
- Richards, M. & B. Sykes, 1998. Reply to Barbujani *et al.* *American Journal of Human Genetics* 62, 491–2.
- Richards, M.R., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H-J. Bandelt & B. Sykes, 1996. Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* 59, 185–203.
- Richards, M.R., V. Macaulay, B. Sykes, P. Pettit, R. Hedges, P. Forster & H-J. Bandelt, 1997. Reply to Cavalli-Sforza and Minch. *American Journal of Human Genetics* 61, 251–4.
- Ringe, D., 1995. 'Nostratic' and the mathematics of chance. *Diachronica* 12(1), 55–74.

- Ringe, D., 1996. The mathematics of 'Amerind'. *Diachronica* 13, 135–54.
- Ringe, D., 1999. Language classification: scientific and unscientific methods, in *The Human Inheritance*, ed. B. Sykes. Oxford: Oxford University Press, 45–74.
- Rogers, R.A., L.D. Martin & T.D. Nicklas, 1990. Ice-age geography and the distribution of native North American languages. *Journal of Biogeography* 17, 131–43.
- Ross, M., 1996. Contact-induced change and the comparative method: cases from Papua New Guinea, in Durie & Ross (eds.), 180–217.
- Ross, M., 1997. Social networks and kinds of speech-community event, in Blench & Spriggs (eds.), 209–61.
- Ruhlen, M., 1991. *A Guide to the World's Languages*, vol. I. Stanford (CA): Stanford University Press (with postscript).
- Santos, F.R., A. Pandya, C. Tyler-Smith, S.D.J. Pena, M. Schanfield, W.R. Leonard, L. Osipova, M.H. Crawford & R.J. Mitchell, 1999. The Central Siberian origin for native American Y chromosomes. *American Journal of Human Genetics* 64, 619–28.
- Schrader, O., 1890. *Prehistoric Antiquities of the Aryan Peoples*. New York (NY): Scribner & Welford.
- Shevoroshkin, V. (ed.), 1989. *Explorations in Language Macrofamilies*. Bochum: Brockmeyer.
- Sims-Williams, P., 1998. Genetics, linguistics and prehistory: thinking big and thinking straight. *Antiquity* 72, 505–27.
- Starostin, S., 1989. Nostratic and Sino-Caucasian, in Shevoroshkin (ed.), 42–66.
- Starostin, S., in press. Comparative-historical linguistics and lexicostatistics, in Renfrew *et al.* (eds.).
- Stone, A.C. & M. Stoneking, 1998. MtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *American Journal of Human Genetics* 62, 1153–70.
- Stone, A.C. & M. Stoneking, 1999. Analysis of ancient DNA from a prehistoric Amerindian cemetery. *Philosophical Transactions of the Royal Society, Biological Sciences* 354, 153–9.
- Swadesh, M., 1960. Unas correlaciones de arqueología y lingüística, Appendix in *El Problema Indoeuropeo*, by P. Bosch-Gimpera. Mexico: Dirección General de Publicaciones, 345–52.
- Swadesh, M., 1972. *The Origin and Diversification of Language* (edited by J. Sherzer). London: Routledge.
- Sykes, B., 1999. The molecular genetics of European ancestry. *Philosophical Transactions of the Royal Society, Biological Sciences* 354, 131–40.
- Thomason, S.G. & T. Kaufman, 1988. *Language Contact, Creolization and Linguistics*. Berkeley (CA): University of California Press.
- Torrioni, A., 2000. Mitochondrial DNA and the origin of Native Americans, in Renfrew (ed.), 77–87.
- Torrioni, A., T.G. Schurr, C.C. Yang, E.J.E. Szathmary, R.C. Williams, M.S. Schanfield, G.A. Troup, W.C. Knowler, D.N. Lawrence, K.M. Weiss & D.C. Wallace, 1992. Native American mitochondrial DNA analysis indicates that the Amerind and Nadene populations were founded by two independent migrations. *Genetics* 130, 153–62.
- Torrioni, A., Y.-S. Chen, O. Semino, S. Santachiara-Benecere, C.R. Scott, M.T. Lott, M. Winter & D.C. Wallace, 1994. Mitochondrial DNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *American Journal of Human Genetics* 54, 303–18.
- Torrioni, A., K. Huoponen, P. Francalacci, M. Petrozzi, L. Morelli, R. Scozzari, D. Obinu, M.L. Savontaus & D.C. Wallace, 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 44, 1835–50.
- Torrioni, A., H.-J. Bandelt, L. D'Urbano, P. Lahermo, P. Moral, D. Sellitto, C. Rengo, P. Forster, M.L. Savontaus, B. Bonne-Tamir & R. Scozzari, 1998. MtDNA analysis reveals a major Palaeolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics* 62, 1137–52.
- Trubetzkoy, N.S., 1939. Gedanken über das Indogermanenproblem. *Acta Linguistica* 1, 81–9. [Reprinted in Scherer, A. (ed.), 1968. *Die Urheimat der Indogermanen*. Darmstadt: Wissenschaftliche Buchgesellschaft, 214–23.]
- Tryon, D., 1999. Language, culture and archaeology in Vanuatu, in Blench & Spriggs (eds.) 1999a, 109–26.
- Underhill, P.A., 1999. Y-chromosome biallelic haplotype diversity: global and European perspectives, in Boyle & Forster (eds.), 35–6.
- Ward, R. & C. Stringer, 1997. A molecular handle on the Neanderthals. *Nature* 388, 225–6.
- Wilson, J.F., D.A. Weiss, M.G. Thomas & D.B. Goldstein, 1999. Y-chromosomal variation in Orkney: identification of Viking signature haplotypes in the British Isles, in Boyle & Forster (eds.), 35.
- Zvelebil, M., 1986. Mesolithic prelude and Neolithic revolution, in *Hunters in Transition*, ed. M. Zvelebil. Cambridge: Cambridge University Press, 5–15.
- Zvelebil, M., 1995. Indo-European origins and the agricultural transition in Europe. *Journal of European Archaeology* 3(1), 33–70.
- Zvelebil, M., 1996. The agricultural frontier and the transition to farming in the circum-Baltic region, in Harris (ed.), 323–45.
- Zvelebil, M. & K.V. Zvelebil, 1988. Agricultural transition and Indo-European dispersals. *Antiquity* 62, 574–8.
- Zvelebil, M. & K.V. Zvelebil, 1990. Agricultural transition, 'Indo-European origins' and the spread of farming, in Markey & Greppin (eds.), 237–66.