

Genetics, linguistics, and prehistory: thinking big and thinking straight

PATRICK SIMS-WILLIAMS*

Many claims have been made linking ancient languages with genetically identified prehistoric and modern populations. There is much new 'evidence' and intense debate on the validity and appropriateness of such interdisciplinary work. Here Patrick Sims-Williams provides a timely comment on linguistics and the quest for ancient populations.

Key-words: linguistics, Indo-European, wave of advance model, genetics, philology

The 'new synthesis'

After the Second World War most prehistorians pursued their own discipline without speculating about the languages and genetic make-up of prehistoric populations. The methodological — and perhaps ideological — objections to synthesizing archaeology, philology, and population genetics seemed too great. Gradually, however, the age of the grand syntheses returned. In the 1970s Ammerman and Cavalli-Sforza correlated the spread of Neolithic farming communities northwestwards from Anatolia with the distribution pattern for 27% of 39 selected modern European genes, the so-called 'first principal component' (Ammerman & Cavalli-Sforza 1984: 105–6 — revised in Piazza *et al.* 1995 to cover 26% of 95 genes). In 1987 Renfrew, referring to the non-genetic aspects of their work as well as drawing on earlier work of his own (1973), correlated the spread of Neolithic agriculture with the hypothetical diffusion of the Proto-Indo-European (PIE) language from Anatolia (Renfrew 1987). Two important differences were that Renfrew did not bring in genetic evidence at all in 1987, still treating it with reserve (Renfrew 1989: 145–9), and that Ammerman & Cavalli-Sforza had preferred to equate the 'Indo-Europeans', not with the Neolithic farmers and the genetic 'first principal component', but with the spread of the Battle Axe culture of the 3rd millennium BC which they tentatively correlated with the distribu-

tion of their 'third principal component' (corresponding to 11% or 12% of 39 modern European genes — revised in Piazza *et al.* 1995 to cover 8.8% of 95 genes). This 'third principal component' peaks north of the Black Sea, in the area of Marija Gimbutas' putative Indo-European homeland (Ammerman & Cavalli-Sforza 1984: 107–8; Cavalli-Sforza *et al.* 1994: 299–300). Despite these major differences, Renfrew's line of reasoning could be seen as moving along the same lines as Cavalli-Sforza's ideas about the feasibility of a 'Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data' (Cavalli-Sforza *et al.* 1988), and it is therefore not surprising that Cavalli-Sforza and his colleagues hailed what they saw as Renfrew's 1987 conversion to the Ammerman/Cavalli-Sforza theory about the spread of Neolithic agrarian populations through a 'wave of advance' as an important support for their own genetic theories (1994: 108, 257, 264–65, 297; but *cf.* Renfrew 1973). Renfrew himself, in his 1991 Huxley Memorial Lecture on 'Archaeology, Genetics and Linguistic Diversity', came round to endorsing 'the view that the spread of farming to Europe from Anatolia had significant genetic effects', adding 'These may plausibly be correlated with significant linguistic effects, namely the initial spread of a proto-Indo-European language or languages through much of Europe' (Renfrew 1992: 463, 472; also 1997: 88).

* Department of Welsh, University of Wales, Old College, King Street, Aberystwyth, Dyfed SY23 2AX, Wales.

Received 30 January 1998, accepted 19 February 1998.

ANTIQUITY 72 (1998): 505–27

Renfrew (1997: 82) notes that 'for some time now there have been indications that a new synthesis is emerging between the disciplines of historical linguistics, prehistoric archaeology and molecular genetics'. This synthesis embraces not only accepted language families such as Indo-European, but also super-families which are dismissed by nearly all historical linguists (see Trask 1996; Dixon 1997; Carstairs-McCarthy 1997), such as 'Nostratic' (putative ancestor of PIE and several other families) and 'Amerind'. Thus Cavalli-Sforza's 1988 'Reconstruction of human evolution', mentioned above, links nearly all the languages of the world, as does the 'Worldwide analysis of genetic and linguistic relationships of human populations' of Chen, Sokal, and Ruhlen (1995), while Renfrew's most recent hypothesis about the dispersal of farming populations from the Near East (Renfrew 1997: 87) (given genetic support by Barbujani *et al.* 1994a)

would place the respective homelands of the Proto-Indo-European, Proto-Afroasiatic, Proto-Elamo-Dravidian and Proto-Altaic languages within a relatively circumscribed Near Eastern heartland. *If these languages are indeed related*, then the archaeological arguments would place Proto-Nostratic as a language spoken in this area, presumably a few millennia prior to the dispersal phase around 8000 to 6000 BC.

Renfrew's proviso, which I italicize, is crucial. According to Dixon (1997: 44 and n. 13), 'no reputable historical linguist ['anyone who teaches the subject at a leading university in the USA or in an EEC nation!'] . . . accepts the claims of Greenberg and the Nostraticists'. It follows that attempts to derive 'Proto-Nostratic', 'Proto-Amerind', 'Proto-Austronesian' and so on from a 'Proto-X' or 'Proto-World' (e.g. Hegedüs 1997: 71; *cf.* Trask 1996: 391–6) and to compare the latter with 'the origin and history of our species' (Chen *et al.* 1997: 595–6) are even more precarious linguistically. Hence Dixon's complaint (1997: 43):

Archaeologists, geneticists and anthropologists like to be given a clear-cut linguistic hypothesis about where and when a proto-language was spoken and exactly how it split and spread. They happily accept any family tree that is produced, without stopping to ask whether it is soundly based, and whether it is accepted by the majority of linguists. The excesses of Greenberg and the 'Nostraticists' have thus received acceptance outside linguistics itself.

Genetics and ethnohistory

Alongside these grand theories linking the genetic history of mankind to the history of human language groups, recent years have seen a remarkable confidence among geneticists such as Cavalli-Sforza and Sokal in the objective existence and knowable prehistory of smaller modern ethnic groups. This confidence has not been shared by archaeologists, social anthropologists and historians (*cf.* Ehret 1988: 571; Roosens 1989; Moore 1995; Jones 1997; Hall 1997). Particular surprise and incredulity was expressed (Pluciennik 1996: 13) at Sokal's claim that his laboratory had compiled a 'European ethnohistory data base', consisting of 6161 records, which 'documents the known locations and movements of 891 ethnic units over the past 4,000 years' (Sokal *et al.* 1993: 56–7). A revised version is now available on the world wide web (Sokal *et al.* 1996). The quality of the data is revealed by one of the few examples of ethnic movement published in print: 'Celts from quadrats 42–45, 54, 57, 58, 66–68, to quadrats 27, 28, 40, 41, 59, 60, 64, 65, 69–71, 81–83' (1993: 64). When compared with Sokal's grid, this example reveals that Sokal's laboratory gives equal weight to the historical — if unquantifiable — movements of Celts into Greece and Anatolia in the 3rd century BC and to their quite hypothetical contemporary movements into the British Isles. Another 'major Celtic expansion' is said to have taken place c. 1000 BC 'from Central and Eastern France and Switzerland (quadrats 55 and 56)', and since this even more hypothetical movement is said to have had 'the greatest effect' on the researchers' 'genetics-ethnohistory correlation' (1993: 68), one is bound to treat this entire correlation with reserve. Note, furthermore, that for prehistoric periods their ethnic unit consists merely of 'an archaeological horizon'. A more accessible example of this strand in population genetics, again involving the Celts, is provided by Sir Walter Bodmer's British Academy 1992 Rhys Lecture on 'The Genetics of Celtic Populations': Bodmer speaks of 'middle European genes . . . brought by the ancestors of the Celts to Wales' (Bodmer 1993: 37), and of 'Celtic invasions of the British Isles, around 600 or 700 BC' (1993: 55), but suggests that Celtic 'culture and language may have come in part' with the 'farmers who were at the tail end of the spread of agriculture from the Middle East that

came to the British Isles around five and a half thousand years ago: 'No doubt these migrating farmers, who may be the closest relatives of the Celts of Central Europe, fancied the blond, blue eyed women of the north' (1993: 56). In his Rhys Lecture Bodmer thus re-synthesizes the disciplines of physical anthropology, ethnology and philology which earlier Celticists had striven to separate — as described by Glyn Daniel in his 1954 Rhys Lecture (Daniel 1954). The confident tone of these modern scientific studies is particularly curious because recent anthropological and archaeological scepticism about ethnicity has focused on the long-suffering 'Celts' more than on any other Indo-European-speaking peoples (Chapman 1992; Waddell 1995; Brown 1996; Collis 1996; Collis 1997; Cunliffe 1997; Sims-Williams in press). The modern population geneticists' perspective seems closer to the 19th-century conception of the 'Celts' as a discrete and long-term ethnic unit, as memorably expressed in 1854 by Ernest Renan, the effective father of 19th-century Celticism: '*Jamais famille humaine n'a vécu plus isolée du monde et plus pure de tout mélange étranger*' ('Never has a human family lived more apart from the world, and been purer from all alien admixture': Psichari 1948: 255; Hutchison 1896: 4). Or, as Matthew Arnold put it in 1865, 'when I was young, I was taught to think of Celt as separated by an impassable gulf from Teuton' (Super 1962: 299–300).

Leading geneticists still tend to equate languages with so-called 'tribes', on the grounds that 'there tends to be a one-to-one correspondence of tribal names to language names'; they then assert that these 'tribes, when not too large numerically, are a reasonable approximation to a population unit for the purpose of genetic analysis' (Cavalli-Sforza *et al.* 1994: 23). Their understanding of ethnology has been condemned as cavalier by Moore (1995) — 'many populations have been lumped together that ethnologists would consider to be quite distinct' — and Moore also complains that their judgements about 'which populations are "admixed" and which not' imply a naïveté about the historical complexity of ethnogenesis. For Moore, the 'distinction between admixed and unadmixed populations is illusory and the product of our ignorance of the ethnohistory of most of the world'. In fact, genetic evidence itself now seems to indicate that admixture has been

extensive (e.g. Bowcock *et al.* 1991) and Moore's criticism is therefore taken seriously by some geneticists (A.R. Templeton and J. Bertranpetit *apud* Ward & Valencia 1996: 20–21). The language=tribe approximation is also problematic. While it may often be true of the modern world, where tribes and languages have been categorized by missionaries and colonial administrators for whom a tribe X would be expected to speak X or X-ese, just as the English spoke English and the French, French, it is 'far from universal' (Robb 1993: 748 and 750) and it is not clear that it applied generally in the ancient world either. Indeed, some anthropologists and historians, noting how minor a role language has in the works of the early Greek ethnographers, tend to minimize its importance in the emergence of 'ethnicity', that is, the categorization of ethnic groups, whether by themselves or by outsiders (*cf.* Jones 1997; Bellwood 1996; Hall 1997). While this tendency may well be excessive (Ehret 1988: 569–70), linguistic data certainly cannot be used as a surrogate for ethnological data about 'ethnic units', as is routinely done by geneticists (e.g. Guglielmino *et al.* 1995; *cf.* Moore 1994: 930, 935, 937).

The example of Proto-Indo-European and Celtic

It is worth emphasizing how limited is our chronological and geographical knowledge even about Proto-Indo-European and its dialects such as Celtic, since naturally 'Proto-Nostratic' and so on are even more insecure. The term 'Celtic', for example, and its variants or (near-)synonyms are first attested in Greek writers in the second half of the 1st millennium BC, referring to barbarian peoples in western Europe, specifically in parts of modern Spain and Portugal, France and Austria. These Greek writers had little knowledge of areas remote from the Mediterranean, and therefore say little about northern European ethnology. Neither they nor any subsequent writers before the 17th century describe the peoples of Britain and Ireland as Celts, although this terminological problem does not of course disprove that they were ethnically related to the known Celts of the Continent. Some sort of ethnic kinship between the Celts of Gaul and the Britons (and arguably the Irish) was apparent to Caesar and Tacitus, about the beginning of our era, but we do not know how far further back this can be projected. Archaeologists

used to attempt to date the 'arrival of the Celts' in Britain and Ireland by correlating them with innovations in material culture interpreted in terms of hypothetical invasions, but since the 1950s doubts about this method have grown, and many archaeologists would say that we simply do not know when Ireland and Britain became in any sense 'Celtic' (e.g. Waddell 1995; Raftery 1995; Collis 1996; Collis 1997). Sokal's implied 3rd-century date and Bodmer's 'around 600 or 700 BC' (both quoted above) are quite hypothetical.

The classical writers also use the term 'Celtic' in a linguistic as well as an ethnic sense, to refer to the language of the Celts. There is direct evidence for this language (identifiable by various distinctive Celtic traits) in inscriptions in Italy, Switzerland, France and Spain, from c. 600 BC onwards, and also from place-names, ethnonyms and personal names. The inscriptions are naturally confined to the vicinity of the Mediterranean, where alphabetic writing developed, and no conclusions can be drawn from their absence from northern Europe. In the same way, we have relatively little onomastic evidence from Britain and Ireland until about the time of Christ, by which time the available evidence shows that Celtic languages ancestral to Welsh, Cornish, Breton and Gaelic were already spoken here. The evidence does not exist to map Celtic speech much before this date. This explains the wide range of modern opinions, which depend more on theories about the spread of PIE than on solid evidence. Thus Nichols (1997: 134), who derives PIE from central Asia, proposes that 'by the end of the third millennium or the beginning of the second, Proto-Thracian, Proto-Italic, Proto-Venetic (unless that was part of Italic), Proto-Celtic, and perhaps Proto-Germanic were somewhere in the vicinity of central Europe', whereas according to Renfrew (1987: 249), who derives PIE from Anatolia, Proto-Celtic would already have reached Ireland by that date. Similar uncertainties recur with most of the other Indo-European dialects. This complicates the task for biologists seeking to compare their distribution with modern genetic evidence. A glaring example is the attempt by Weng & Sokal (1995: 579–81) to examine genetic data for 'seven major Indo-European language branches or families in Europe' on the basis of 'geographic localities . . . chosen to represent, as far as possible,

populations that nowadays inhabit areas where the language families may have originated': Copenhagen for Germanic, Brest for Celtic, Rome for Romance, Athens for Greek, Shkodra for Albanian, Vilnius for Baltic and Minsk for Slavic. Starting from such speculative data it is hardly surprising that 'the differentiation of the seven major Indo-European language branches in Europe seems unrelated to the times of the origin of agriculture' or indeed to anything else.

It cannot be emphasized too often that Indo-European comparative philology is primarily concerned with language-comparison and only very secondarily with the location in time and space of PIE and its dialects. This is already clear in the famous statement of Sir William Jones in 1786 (Franklin 1995: 361):

no philologer could examine [Sanskrit, Greek, and Latin] without believing them to have sprung from some common source, which, perhaps, no longer exists: there is a similar reason, though not quite so forcible, for supposing that both the *Gothick* and the *Celtick*, though blended with a very different idiom, had the same origin with the *Sanscrit*.

Jones here focuses on the ultimate identity of the five languages, and merely alludes to two matters which are highly relevant to attempts to combine linguistics with prehistory and genetics.

One is what we would now call the non-Indo-European 'substrate' elements in Germanic and Celtic (*cf.* Hamp 1990b; Polomé 1990), which is presumably what Jones meant by 'the Gothick and the Celtick' being 'blended with a very different idiom'. (This does not apply only to Germanic and Celtic, since other Indo-European languages such as Greek also have many 'substrate' elements: Mallory 1989: 67.) The problem here is that it is difficult to say anything convincing about contacts between Germanic or Celtic and other hypothetical indigenous or intrusive western languages when these languages are either not available for inspection or unintelligible, as in the case of the apparently non-Indo-European Pictish inscriptions in Scotland (Jackson 1980; *cf.* Forsyth 1997); it is inevitably a more intractable problem than, say, the relationship between Indo-Iranian and non-Indo-European Dravidian in India, where both languages are known and understood (Harris 1990: 69). More than once, features of Celtic which have been regarded as 'non-Indo-European', such as the initial muta-

tions or the syntax of the verb, have turned out to be amenable to explanation as regular developments from PIE to Celtic (Sims-Williams 1997: 350). This is a warning against applying the non-Indo-European label too readily whenever an Indo-European explanation is not apparent. Nevertheless, in principle it is likely enough that Celtic, like the other Indo-European daughter languages, is not purely Indo-European but incorporates non-Indo-European elements. The 'Pre-Aryan' element in Insular Celtic was a popular line of research in the early 20th century, and it is still seriously investigated by some biologists and linguists (Watkin 1986: 138; Schmidt 1990; Birkhan 1997: 79–81). A word of caution is needed. If there are identifiable substratum elements in Insular Celtic, we cannot assume that they entered Celtic in the British Isles. For example, the Insular Celtic element *carn-* ('cairn'), which has a good chance of being non-Indo-European, is also attested in ancient Continental Celtic inscriptions in France and north Italy (Hamp 1990c). The term *substratum* should not be taken to imply a teleological picture of PIE marching on inexorably into new territories and overlying indigenous languages in the process. Alongside that process, there would surely have been periods and places where non-Indo-European languages temporarily gained the upper hand over PIE, in the way that Turkish has done in the historical period (*cf.* Nichols 1997). For example, there is no reason to assume that Pictish was an indigenous pre-Celtic language; the Picts could be relatively new arrivals in the British Isles compared to the Britons and Irish, as was indeed believed in the early Middle Ages (Wright 1984: 92; Sims-Williams 1985: 117–18). Such uncertainties need to be borne in mind when reconciling language with biology and prehistory.

The second matter raised by Sir William Jones is the location of PIE in time and space. This is implicit in his phrase 'some common source, which, perhaps, no longer exists'. Jones' hint that the parent language might still exist somewhere in the world was soon abandoned, but we still cannot state when PIE began to split up (whatever we mean by that, for PIE is itself a problematic concept), nor where it spread from, unless we are prepared to make unprovable equations between languages and archaeological cultures. The fundamental obstacle is that even

in the case of the Indo-European dialects which are attested as early the 2nd millennium BC, we have no means of guessing at the absolute chronology of the sound-changes before 2000 BC which led to their emergence. There is no equivalent to Carbon-14 in phonology; languages do not evolve at a steady rate (the 'glottochronological' fallacy: *cf.* R. Coleman *apud* Anthony & Wailes 1988: 450; Zimmer 1990: 312–13; Trask 1996: 362–6; Mallory 1997: 99; Dixon 1997: 11, 36–7). This is obvious, for instance, if we compare Irish and Welsh over the last 1500 years, when Welsh has been much more stable phonologically. This means that textbook dates for the divergence of PIE such as *c.* 4000 BC rest more on what Mallory calls 'intuition' and 'informed speculation' (Mallory 1989: 276, n. 28; 1997: 100, 110) than on any reliable method of prediction from the distance between the attested dialects. Those linguists who objected to Renfrew's much earlier date (*c.* 7000 BC) are properly criticized by him (Renfrew 1989: 140–42; 1990: 19–21) for bringing back glottochronology through the back door. In fact Renfrew's comment that 'the argument for the measuring scale has never been made clear to me' (1997: 88) is very much in tune with the recent observations on PIE by the Australian linguist R.M.W. Dixon (1997: 47, 49): 'What has always filled me with wonder is the assurance with which many historical linguists assign a date to their reconstructed proto-language. . . 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)'.

The usual way round this problem is 'linguistic palaeontology', that is, reconstructing putative PIE words for items such as 'king', 'horse', and 'wheel' and tying these to presumed stages of mankind's material and social development. A well-known difficulty here is the possibility of parallel or independent semantic developments within the dialects — a word for 'headman' or 'protector' coming to mean 'king', and so on (and for further problems with PIE 'king' see Mallory 1989: 125). Another is the possibility of terms for innovations being borrowed from one dialect to another. Philologists are usually confident about distinguishing loanwords from cognates on phonological grounds, but this becomes increasingly problematic as one goes back towards the time when

the sound-systems of the dialects were closer and the likelihood of sound-substitution becomes greater. It also depends on the phonetics of the word in question; to give a simple example, if someone wanted to claim that Old English *tun* (English 'town') derived from Celtic **dūnom* (Irish *dún* 'fort') he would have to place the borrowing before the Germanic sound-shift that changed PIE /d/ to /t/, but this chronological constraint would not apply to an otherwise similar attempt to claim that Old English *run* 'secret' was borrowed from Celtic (Irish *rún*), or *vice versa* for that matter. Since we know nothing of the early absolute chronology, there seems no way of estimating how many 'invisible' loanwords may have passed between the emerging IE dialects in the millennia before 2000 BC, and in certain phonetically favourable conditions, much later (cf. Mallory 1996: 13–16). The rather small number of words relevant to linguistic palaeontology need to be considered in this light (cf. Mallory 1996: 17–18; 1997: 101). Comparative philologists are well aware of the problem, at least intermittently. For example, Szemerényi, in support of his contention that 'Greek *taúros* was in all probability borrowed from Semitic' after 2000 BC (cf. Nichols 1997: 143) and passed thence into 'most European languages' through 'successive borrowings', remarked (Szemerényi 1985: 41) that

we must not overlook . . . that early historical contacts must have produced loan relations which in some cases were [*sic*] hard to distinguish from prehistoric kinship relations.

It might be objected that it is one thing to explain away the relatively restricted European 'bull' word as a loan-word and another to explain away the phonetically more challenging and widespread word for horse, **ekʷos*, which is attested from Ireland to India (cf. R. Coleman *apud* Anthony & Wailes 1988: 450; Renfrew 1989: 151; Hamp 1990b; Mallory 1996: 9–10 and 1997: 101, 109–10). This does not really present a problem, however, unless linguists limit themselves by unprovable assumptions about the absolute chronology of the prehistoric IE sound-changes and also about the possible degree and extent of communication across the Indo-European world. Prehistoric intercommunication is a matter largely outside the remit of linguists; they can, however, compare large language continuums across modern low-

density populations (cf. Renfrew 1989: 116, 142–4; Hamp and Renfrew *apud* Markey & Greppin 1990: 397–9; Bateman *et al.* 1990: 9; Robb 1993: 750–51, 758; Trask 1996: 358; Mallory 1996: 6–7; Dixon 1997: 90–91), and they can note the apparent implication of the widespread uniformity of European river-names (Vennemann 1994; Kitson 1996; but cf. Beekes 1995: 32: 'the very idea of such a large linguistic unity at such an early period can hardly be imagined'). To sum up, then, there seems to be no reliable way of establishing the upper limit for the break-up of PIE. If Renfrew were able convince his fellow archaeologists that the first farmers were the only possible bearers of PIE, then philologists could probably explain away all the shared vocabulary that has seemed to imply later phases of civilization. Unfortunately, however, Renfrew's is not the only available theory for the diffusion of PIE (cf. Hines 1991; Robb 1993; Mallory 1997; Nichols 1997), and it may be, as Robb (1991) has suggested, that PIE spread as the result of a series of random events rather than a single process.

Problems with semantic shifts and loanwords also arise with attempts to locate the 'IE homeland' on the basis of shared vocabulary items such as the word for the 'bright tree', PIE **bherǵʰo-*, in Latin the 'ash' but generally the 'birch', a geographically restricted species (see Mallory 1989: 275; Trask 1997: 355). The principle of seeking the centre of dispersal at the 'centre of gravity' — that is, where the greatest variety of dialects survived — has also been shown to be invalid (Mallory 1997: 95–7; Nichols 1997: 129–30 and 133–4); moreover, it will tend to privilege regions such as Mediterranean Europe where written evidence appears early. Indeed, it is quite conceivable, although Mallory (1997: 105) is reluctant to contemplate this, that Indo-European failed to survive in the original PIE homeland (cf. Nichols 1997: 134). Another method is to locate PIE with reference to the location of languages which seem to have borrowed from it, such as Proto-Kartvelian (Nichols 1997; cf. Harris 1990), or loaned to it, such as Semitic (Szemerényi 1989: 162); the objection to this is that 'the homelands of the other language families putatively in contact with Proto-Indo-European are no more securely located than that of the Indo-Europeans' (Mallory 1997: 98; 1996: 3). A more widely employed method is to start from the earliest known location of

the surviving dialects and to associate them with material cultures in the same areas, for example, Celtic with La Tène, Germanic with the Jastorf culture. (Renfrew (1989: 116–17) is in principle sympathetic to this method, which fits in with the ideology dubbed ‘immobilism’ by its critics.) The forerunners of these cultures will then be traced on purely archaeological grounds, La Tène back to Hallstatt to Urnfield and so on, with the languages being assumed to follow suit, until we reach a common archaeological horizon for all the languages, such as the Corded Ware culture, which is then equated with the PIE homeland. The main objections to this method are the impossibility of verifying matches between prehistoric material cultures and languages and, secondly, the fact that there are too many uncontrollable variables on the linguistic side. For one, we cannot estimate how many lost IE and non-IE languages need to be allowed for in areas such as northern Europe (*cf.* Mallory 1989: 146–7; Renfrew 1989: 143; Zvelebil & Zvelebil 1990: 265, figure 1; Hamp in Markey & Greppin 1990: 397; Robb 1993: 750), since in areas where we have fairly early evidence, such as Italy and Greece, no consistent pattern can be observed — ancient Italy being more linguistically mixed than Greece (*cf.* Mallory 1997: 95–6). We do not know how many pieces there may be in the jigsaw: New Guinea has a new language every 1000 square kilometres (Dixon 1997: 17)! Furthermore it may be true, as Robb (1991: 289) guesses, that ‘Indo-European languages have probably replaced other Indo-European languages at least as often as they have replaced non-Indo-European languages’. Another uncontrollable variable, noted earlier, is that there is no reason to assume that PIE spread inexorably through some inherent superiority, rather than contracting periodically as well. We need not assume ‘an enterprising go-ahead race’ (Szemerényi 1989: 164) behind the spread of PIE.

The true historical situation was no doubt more messy than we can imagine or reconstruct. Certainly a simple wave of advance from a single homeland, such as Renfrew’s agriculturists fanning out from Anatolia, conflicts with the fact that while some currently adjacent dialects appear to be close, others, such as Hittite, Greek, and Italic, are very distinct. This objection was raised by Robert Coleman (*apud* Anthony & Wailes 1988: 453) and Anna Morpurgo Davies

(*apud* Renfrew 1989: 166) and was recognized by Renfrew himself to be a problem (Renfrew 1989: 176; *cf.* 1990: 22; Hines 1991: 52; Beekes 1995: 49; Mallory 1997: 110). A similar objection is raised by Weng & Sokal (1995: 589), who claim that, of their seven European language groups, only Baltic and Slavic are close both as languages [‘LAN’] and in the date of the origin of agriculture among their speakers [‘OOA’]. Germanic-Celtic, Romance-Greek, Romance-Albanian and Greek-Albanian have ‘low OOA distances paired with high LAN distances’, while in the cases of Germanic-Baltic, Germanic-Slavic, Romance-Baltic and Romance-Slavic ‘high OOA distances are paired with relatively low LAN distances’. The value of Weng & Sokal’s ‘substantial negative correlation between the two kinds of distances’ is, however, diminished by two problems. The first is that their ‘Germanic-Celtic’ really means ‘Copenhagen-Brest’ and so on (see above). The second is that it is not at all clear what the linguistic distance between some of these language-pairs really is. Deciding the distance between, say, Germanic, Celtic and Italic (Romance) depends on assigning a time depth to the two-dimensional and contradictory evidence of the isoglosses, and here philologists disagree. For example, does the statistically-evidenced closeness of Celtic, Germanic and Italic (Sims-Williams 1997: 351–2) derive from some early period of linguistic unity located in western Europe or even elsewhere, of the genetic type often supposed for Greek and Armenian (Hamp 1982: 349; *cf.* Clackson 1994; Warnow 1997), or is it merely due to contacts between these adjacent languages in the last one or two millennia BC, as once claimed by Hamp (1982: 349)? Is it legitimate, by ascribing its links with Italic and Germanic to late contact, to portray Celtic as an eastern Indo-European language (Schmidt 1996), or is such an argument merely circular (Schrijver 1997: 109)? Such questions are easier to ask than to answer in view of the difficulty of distinguishing early loans from cognates. The latest model, using statistical techniques borrowed from genetics, sees Italo-Celtic as a unity, the third language (after Hittite and Tocharian) to branch off from PIE, and sees Germanic as a much later offshoot which ‘moved away . . . into close contact with the “western” languages (Celtic and Italic) and borrowed much of its distinctive vocabulary from them at a period early

enough that these borrowings cannot be distinguished from true cognates' (Warnow *et al.* 1997: 6587). Evaluation of this interesting hypothesis awaits the publication of the data on which it is based (noted as 'in press' by Nichols 1997: 148). Translating it into historical geography would depend on one's chosen chronological and geographical starting point for PIE. We may doubt, however, that the loci of Celtic and Germanic would be Brest and Copenhagen.

In summary, then, there are many reasons why geneticists should be wary when combining their own data with inherited ideas about even such relatively 'well-known' languages and peoples as the Celtic and Indo-European ones. They need to be up to date and above all critical; the prehistory of such entities is still very uncertain. In many ways, speculation about the distribution of ancient Indo-European languages before *c.* 2000 BC (or *c.* 200 BC in the case of a medium-range dialect like Celtic) is rather like looking at the clouds in the sky and trying to guess how they might have looked 10 minutes before. Yet speculation about the pre-modern distribution of genes often looks even more precarious. For example, Ammerman & Cavalli-Sforza (1984: 108), explaining the distribution of their 'third principal component', were prepared to consider 'the so-called barbarian invasions in late Roman times and the period just after the collapse of the Roman Empire' as a possible alternative to Gimbutas' Indo-Europeans (the latter correlation is rejected by Barbuji *et al.* 1995: 126 and Weng & Sokal 1995: 592–3). This was possible because 'unfortunately synthetic genetic maps are inherently undated' (Piazza *et al.* 1995: 5839). Again, it is suggested that the north–south pattern of the 'second principal component' may reflect *either* the spread of the Uralic language *or* the climatic difference *or both or neither* (Piazza *et al.* 1995). What is clearly needed is the secure genetic time-depth that could come either from the comprehensive examination of human DNA from archaeological excavations or from chronologically reliable and agreed systems of 'triangulation' from modern genetic data. Either of these approaches could be as revolutionary as radiocarbon dating and dendrochronology, as rightly noted by Renfrew (1992: 448, 467–8). To date, however, most of the data employed by population geneticists has been as nebulous as that of the linguists.

Languages and genes

There are understandable reasons in the history of both disciplines why linguistics and biology have tended to look to each other for support. But is there any direct connection between them? Apparently not. While human speech depends on physiology, language differences do not seem to depend on physiological factors to any serious degree (Moore 1994: 930); as trivial exceptions which prove the rule we can note the view that features of the English of Glasgow and Liverpool depend on problems with the teeth and adenoids among the respective populations (Samuels 1972: 19)! Basically, language differentiation has nothing to do with genetics. A Chinese baby adopted by Americans will learn to speak perfect American English, and to mimic American physiognomy — the 'articulatory setting' (Samuels 1972: 20) — while doing so. Languages are thus transmitted independently of genes. Yet obviously the two may be, and often are, transmitted along parallel paths — to all appearances 'together' — from generation to generation. Hence the historically virtually useless tautology that 'Australian' or 'African' languages tend to co-occur with 'Australian' or 'African' genes. More interesting historically are such unusual cases as the correlation between blood-groups and linguistic evidence which seems to confirm that the Romani language and some inbred European gypsy populations derive from India some 1500 years ago, reaching Greece in the 14th century and Britain in the 15th (E.M. Williams 1986: 187, 199; for problems see Fraser 1992: 22–5). We have to remember, however, that the Romani language and the genes were transmitted 'together' in a weak sense only, and that indeed many European Romani-speakers lack distinctive 'gypsy' genes; in other words they are linguistically and culturally gypsies but not genetically gypsies *if* we define genetic gypsies on the basis of Central European gypsies (*cf.* Sunderland 1976: 260; E.M. Williams 1986: 206–8; Bodmer 1993: 42).

This example leads to the question of whether genetic similarities can be used to predict linguistic similarities and *vice versa*, even though languages are not transmitted genetically. 'Predict' is too strong a word. 'Raise the possibility' would be safer. For example, if biologists were to discover a genetic similarity between the populations of Scotland and the Pyrenees

(a possibility casually entertained by Falsetti & Sokal 1993: 226; *cf.* Collins 1986: 6), linguists would certainly want to re-investigate the apparent non-relationship between Pictish and Basque (Jackson 1980: 130, 152; Trask 1997: 391; Forsyth 1997: 8, n. 14); and, conversely, it would presumably be worth looking at biological evidence in the unlikely event that a link between Basque and Pictish were ever to be proved (note the general assumption of geneticists, e.g. Bowcock *et al.* 1991: 841, that Basque is a remnant of Palaeolithic pan-European; similarly Vennemann 1994, opposed by Kitson 1996 and Trask 1997: 364–8). Yet, nothing could be predicted. Furthermore, in such an investigation the disciplines of linguistics and biology would not be in any special relationship *vis-à-vis* archaeology, history and other human sciences which might hint at such a connection. Both linguistic and biological similarities between populations must testify to their contiguity or former contiguity either with each other or with other — perhaps unknown — populations with similar characteristics. Yet language is just one of various pointers to possible connections between populations. It is not in a different league from archaeological or ethnological or historical testimonies. The connections between populations suggested by linguistics deserve checking out by geneticists, but so do those revealed by archaeology and those asserted by ancient and modern ethnographers. (Particular care is needed, of course, in using origin legends; *cf.* Ó Corráin 1985; Sims-Williams 1985; Hall 1997.) Linguistic evidence is especially abundant, but it is not obvious that it should be regarded as exceptionally powerful.

We need therefore to avoid naïve interpretations of the shorthand equations of language and genetics in some of the biological literature. For instance, Harper states that ‘the physical and linguistic distinction of Brythonic from Goidelic Celts is reflected in the difference in prevalence of the phenylketonuria gene’ (Harper 1986: 283–5). His map (1986: 277) duly shows a high incidence of phenylketonuria per 10,000 births in southern Ireland (2.1), northern Ireland (1.3) and Scotland (1.2), *versus* a low incidence in Wales (gypsies apart: E.M. Williams 1986: 205) and the other ‘Brythonic’ areas (all of them, like England, having 0.7 or less). Yet it is typical of such maps that the Icelanders (with 1.3) have to be treated as honorary Goidels,

which is historically quite plausible (Pálsson 1986; Sigurðsson 1988), and, more alarmingly, that the Danes also have an awkwardly high incidence (1.1). The matter has been put in a new light by Zschocke and colleagues (1997), who show that the main Danish mutation is independent (they do not comment on Iceland), whereas R408W-H1, the mutation common in Ireland and Scotland, ‘originated locally, possibly in Ireland’. Although there is no real proof, they suggest that it may be older than 2500 BC. If so there is no point in attaching a ‘Goidelic’, or ‘Celtic’ (Scriver *et al.* 1996: 81), label to it. Another example of a dubious ethnic label is provided by Lucotte & Hazout (1995), in their study of G551D, the third most common cystic fibrosis mutation. ‘Regions with a high frequency’, they note, ‘correspond to areas with large past and actual populations of Celtic descent’ (1995: 566). The highest figures are from the British Isles, the Czech Republic (‘ethnic Czechs’ only), and Brest (‘Celtic populations’ only). It is difficult to agree that this is ‘consistent with a Celtic origin of the G551D mutation for populations where prevalent Celtic origins persisted until the sixth century AD’ (1995: 566). If the Czech Republic can be counted as Celtic (on archaeological evidence), so should France (i.e. Celtic Gaul), but the figures for France (from Paris, Lyon and Montpellier) are extremely low. Thus, even if we accept the judgement of Lucotte & Hazout that ‘earlier population expansions’ spread cystic fibrosis mutations, there seems no valid reason to envisage movements of ‘Celts’ in particular. In the same way, it is difficult to take seriously a recent suggestion that the Phoenicians carried G542X, the second most common mutation, round the Mediterranean in the 1st millennium BC (Loirat *et al.* 1997). One hopes that such ethnological speculations do not influence public health policy. It has been suggested, for instance, that such studies linking genetic disorders with past population movements ‘can assist mutation screening by predicting common mutations in other populations’ (Zschocke *et al.* 1997: 193).

One is bound to worry about the extent to which the many distribution maps produced by linguists and biologists are bound to overlap from time to time by mere coincidence. For instance, how ‘surprising’ is it that Welsh dialect distributions in Montgomeryshire ‘almost

exactly coincide with the . . . areas found by dermatoglyphics' (W.R. Williams 1986: 220)? The case of the gypsies differs interestingly from such examples because it involves very specific long-distance linguistic and biological (and historical) connections. Unfortunately, many linguistic/genetic correlations, such as that between 'native Australian' languages and 'native Australian' genes on the Australian continent, have nothing very interesting to say to historians, including historical linguists, because they can be understood in simple geophysical terms. This is also a problem with more complex cases. Suppose, for instance, that one day DNA data from excavations showed that the population of Ireland was genetically more homogeneous *c.* 800 AD than that of Britain: could we correlate this with the fact that Gaelic was essentially the only language of Ireland then, whereas Britain had Welsh, Cornish, Cumbric, Gaelic, Pictish and Anglo-Saxon? If there was an inherent link between biology and language perhaps we could. But since there is not, we would have to admit that merely geophysical factors could have allowed more diversity — both genetic and linguistic — to co-exist on one island than on the other. In other words, linguistics and biology are on all fours not only with the other human sciences, as noted above, but also with the whole empire of geographical sciences. An example is provided by the genetic and linguistic peculiarities of the Basques. These are often assumed to be inter-related (e.g. Ammerman & Cavalli-Sforza 1984: 86; *cf.* Cavalli-Sforza *et al.* 1994: 280 and 285–6; Bertranpetit *et al.* 1995; Côté-Real *et al.* 1996 — note also the correspondence claimed by Vennemann 1994: 277 n. 91 between the European distribution of the typically Basque Rhesus-negative gene and the area of 'Old European' hydronymy). Instead, one could envisage the valleys of the Pyrenees as rock-pools catching the debris from successive, rather than concurrent, genetic and linguistic tides. Here it is salutary to compare the linguistic minorities in the Italian Alps. At the level of mitochondrial DNA, Ladin speakers are even more isolated from other Europeans than the Basques are, and their language is also a very restricted one (Stenico *et al.* 1996); yet, since Ladin is derived from Latin we know that it cannot be the ancestral language of this isolated population. Just as the widespread language Latin replaced this Al-

pine population's earlier language (perhaps Rhaetic), so Proto-Basque, formerly widespread in Aquitaine (Trask 1997), may have replaced an earlier (unknown) minority language of the Pyrenean population that now speaks Basque.

The ambiguity of genetic maps

Historical geographers have an important role to play in evaluating the maps produced by biologists. The fundamental problem with most of these maps is that they are open to multiple interpretations. Consider, for instance, Coon's map of 'Pigmentation of Hair and Eyes' (Coon 1939: 270–71, reproduced with slightly less detail in Bodmer & Cavalli-Sforza 1976: 583, Bodmer 1993: 54 and Cavalli-Sforza *et al.* 1994: 267); the general trend is north–south, but there are interesting variations which might tempt us into historical explanations. For example, the obvious resemblance between the lightly-pigmented populations of eastern Britain and Scandinavia/northern Germany might be attributed to the Anglo-Saxon settlement in the 5th and 6th centuries AD. Equally well, however, it could be due to remoter prehistoric relations across the North Sea, or perhaps even to long-term environmental factors. Already in the 1st century AD the population of Britain appeared to be of mixed origin to Tacitus — yellowy-red haired (*rutilae*) Caledonians looking as if they came from Germania (compare Boudica: Cunliffe 1997: 110); dark curly-haired Silures (in south Wales) looking like descendants of Iberians (just the same Spain/south Wales correspondence as on Coon's modern map!); and southerners resembling Gauls — whether for genetic or environmental reasons Tacitus was unsure (*Agricola* §11). One concludes that the genetic similarities across the North Sea *may* pre-date the Anglo-Saxons and should not be associated with them too definitely. For similar reasons, genetic resemblances across the Irish Sea, such as the prevalence of the HLA-A1, B8 tissue haplotype in Wales, Scotland and Ireland *vis-à-vis* England cannot meaningfully be labelled 'Celtic' (Bodmer 1993: 48–9) until we can discover when they originated — and when these areas became linguistically 'Celtic'. A plausible way of looking at such a case was provided by Sokal and his colleagues. They noted that (speakers of) peripheral European language families (Greek, Basque, Finnic and Celtic) are frequently genetic outliers; for instance, 'a

northwest–southeast cline for HLA allele frequency *B8* is lowest for Greek speakers and highest for Celtic speakers' (Sokal *et al.* 1989: 500). Such data suggested the importance of 'clinal patterns, caused by either gene flow or adaptation, resulting in the differentiation of language families located at the extremes of the cline' (Sokal *et al.* 1989: 500). It would, however, have been more cautious to speak of 'differentiation of *populations* at the extremes of the cline'; the fact that these populations happen to speak Greek or Celtic may be irrelevant. It may be their peripheral, peninsular position that allowed the Greek and Celtic languages to survive in these areas.

Genetic patterns linking adjacent areas cannot point to unique historical explanations. This is why it would be difficult to find *provable* biological correlates for known migrations between adjacent areas, such as those in about the 5th century AD from Ireland to Gaelic Scotland and from Britain to Brittany. Very specific genetic correlations between *non-contiguous* areas are more likely to lend themselves to unique historical explanations, the correlation between gypsy genes, language and migrations being a fine example; another is the genetic evidence that confirms that the modern population of Tenerife descends from an admixture of specifically male Spaniards and specifically female native Guanchos (Pinto *et al.* 1996). To find anything as striking in the British Isles, one would have to undertake a systematic biological investigation of linguistic (and other) correlations between *non-contiguous* areas. Here non-biologists can contribute by setting a testable agenda. A good place to start would be the Celtic ethnic names in the *Geography* of Ptolemy (c. 150 AD) and other comparable sources (O'Rahilly 1946: 1–42, 453–66; Rivet & Smith 1979). Ptolemy has several examples of ethnonyms recurring at a distance, some of which quite likely refer to offshoots of the same peoples. The *Atrebat*es in northern France and southern England are still too close to be interesting (compare Tacitus' caution above). On the other hand, the connection between the *Parisii* of the Paris region and east Yorkshire, which is confirmed by archaeology (Wait 1995: 499–500), deserves comparison with the modern biological evidence for the populations in the same areas, on the working hypothesis of continuity of populations. The *Gangani* are another

interesting group. Ptolemy calls the Llŷn peninsula in north Wales the 'headland of the Gangani', implying a connection with the *Gangani* whom he places on the southwest coast of Ireland (in Kerry according to Mac an Bhaird 1991: 7; cf. Rivet & Smith 1979: 107). In this case, any genetic tie-up would have to be very specific to Kerry to be interesting, as the name of the Llŷn peninsula itself implies another, later settlement from Ireland; *Llŷn* comes from *Laighin* 'Leinstermen' (Lloyd-Jones 1928: 6). It would not be good enough, then to find vaguely 'Irish' rather than 'Kerry' genes on the 'headland of the Gangani'. Similar problems arise in the case of the southeastern Irish peoples whom Ptolemy calls *Manapii* and *Brigantes*. These names are also found in Flanders (the *Menapii*: Schrijver 1994: 82, 93) and in north Britain (the *Brigantes*), and phonetically they look like imports, unless sound-substitution by foreign informants has occurred (cf. Mac an Bhaird 1991: 14). *Manapii* shows the sound /p/ which did not exist in Irish at this stage, and *Brigantes* shows an /nt/ which has not undergone the Irish change to /d/, a change which had already occurred according to Pokorny, who cites Ptolemy's northern Irish river-name *Argita* [=arge:da: ?] 'silvery-river' with /d/</nt/, as in the Primitive Irish ogam inscriptions (Pokorny 1950; cf. Thurneysen 1946: 127 and Mac an Bhaird 1991: 16). It is sometimes argued that the other two southeastern tribes, the *Cauci* and *Coriondi*, have Continental connections as well (Warner 1991: 48; cf. O'Rahilly 1946: 24–7, 33–4, 39, 462–3; Mac an Bhaird 1991: 13–15). In a pioneering study of 'history and blood-groups in the British Isles' Potts (1976: 243) suggested a link between the *Manapii* and *Brigantes* and the relatively high frequency of the A blood-groups in southeast Ireland, as in Britain. Yet, as Potts himself notes, other candidates would be the Vikings of Wexford or the Anglo-Normans (as argued by Dawson 1964). Sokal and others relate the 'east–west cline of ABO allele frequencies in Ireland . . . to the penetration of the Celtic substrate population by Anglo-Saxon migrants from Britain' (Sokal *et al.* 1993: 56; cf. Relethford & Crawford 1995: 34–5). Here, then, is yet another good example of a geographical distribution open to multiple interpretation. As a Parthian shot, Potts states that 'genetically the Belfast man is at least as Irish as the Dublin man and rather more Irish

than the man from county Wexford' (1976: 245). That could be predicted merely from Ireland's geographical orientation.

Even a very precise linguistic/biological correlation may be open to doubt. If one looks at maps of blood-groups in Wales, there is a remarkable coincidence between the high frequency of the A gene in southern Pembrokeshire, which is known as 'Little England beyond Wales', and the evidence for the very long-standing strength of English in the same area, as revealed for example by non-Celtic place-names; the A gene and the non-Celtic place-names cluster in southern Pembrokeshire and the hinterland of Chester (Carter 1986: 35 and Watkin 1986: 122, whence Bodmer 1993: 43). The matter is complicated, however, by the historical fact that the de-Cambricization of Little England beyond Wales was originally due to its settlement by Flemings and Anglo-Normans in the 12th century (Thomas 1986: 54; Watkin 1986: 123–5; Toorians 1990; *cf.* Toorians 1996). Moreover, Morgan Watkin pointed out that the A gene frequency here is higher than in modern England, Normandy and Flanders, and is in fact best paralleled in Scandinavia (Watkin 1986: 125–7, 139). This led him to the conclusion that the area, like that round Chester (Davies 1990: 51–6), was heavily settled by Vikings, despite the lack of solid historical evidence for this (Loyn 1994: 37). Potts, on the other hand, has argued that the prevalence of A blood-groups in Little England beyond Wales is due to genetic drift in an isolated area (1976: 251–2). This genetic isolation is supported by the evidence of marriage registers (Woolley 1986: 243). To complicate matters still further, Sunderland has argued that when a wider range of genetic markers in Wales are considered, 'north Pembrokeshire is the atypical region and not south Pembrokeshire, as might have been expected on historical grounds' (1976: 255). Matters are not as straightforward as the precise modern linguistic/biological correlation in Little England beyond Wales might imply. This should make one hesitate before advancing historical explanations for Dawson's (1964) maps of blood types in Ireland, let alone for genetic maps of areas where the history is less well known. A further worrying possibility is that what appears to the layman to be a non-random distribution, may conceivably be random, as was demonstrated mathematically by Upton & Fingleton (1985:

161) in the case of a map of the A blood-groups in the 26 counties of the Republic of Ireland.

The maps correlating modern European genes and the spread of Neolithic farming and (hypothetically) of Proto-Indo-European can be reconsidered in the light of the above object lessons. It is agreed that agriculture spread northwest through Europe from Anatolia, although there is room for disagreement over the correctness of the 'wave of advance' model — Anthony (1990: 901–3) argues for 'leapfrogging', and others have argued that acculturation of Mesolithic populations may have played a part (*cf.* Renfrew 1992: 454–5; Zschocke *et al.* 1997: 191), in which case there would have been less demic diffusion and movement of genes than posited by the original model. As we have seen, maps of the so-called 'first principal component' of European genes show a similar northwest/southeast pattern, as first shown by Ammerman & Cavalli-Sforza (1984: 105–6; Cavalli-Sforza *et al.* 1994: 292; Piazza *et al.* 1995), and accepted by many other geneticists (e.g. Bodmer 1993: 52–4; Sokal *et al.* 1993: 55; Barbujani *et al.* 1994a; 1995). The overall resemblance between the selected modern genes and Neolithic agriculture is indeed very clear, even without the statistical underpinning provided by Sokal, Oden and Wilson (1991). The problem is that the broad patterns for agriculture and the 'first principal component' are both so very general that it is difficult to be sure that there has to be any connection between them (Mallory 1997: 111). The distribution of genetic markers from southeast to northwest makes excellent sense in terms of long-term gene-flow across Europe along this axis, perhaps in both directions, but does not point to the Neolithic farmers in any strikingly specific way. In Ammerman & Cavalli-Sforza's own words (1984: 106),

There is no necessary reason to link a particular principal component with a particular explanation. The fact that we are looking for precisely the pattern found in the first principal component could, of course, be a coincidence, and other explanations can never be entirely excluded.

In fact the authors of a recent study of the European mitochondrial gene pool (Richards *et al.* 1996: 197) argue that

it seems theoretically possible that the cline was established not by Neolithic immigration but by

earlier Upper Paleolithic colonizations that, our analysis suggests, brought with them the ancestors of most modern Europeans.

If the connection between the genes and agriculture is regarded as unproven therefore, the likelihood of connecting the genes with the spread of Proto-Indo-European becomes remote, since Renfrew's model for the diffusion of Proto-Indo-European is dependent on tying it to the spread of agriculture. As we have seen, Weng & Sokal (1995), in their admittedly imperfect study, found no genetic data to support Renfrew's 'out-of-Anatolia' theory of PIE *over and above* the genetic data commonly held to be correlated with the diffusion of agriculture (this is also true of Barbuani *et al.* 1995). And in an earlier study, Sokal, Oden and Thomson (1992) found no genetic evidence to support in detail either Gimbutas' theory of the spread of PIE or Renfrew's. The cautious conclusion must be that the spread of PIE has not yet been securely correlated with any modern genetic evidence.

Would we expect it to be? Surely not, despite the opinion of Cavalli-Sforza, Menozzi and Piazza, who have argued that 'languages offer a powerful ethnic guidebook, which is essentially complete' and which can be significantly correlated with genetic groupings (Cavalli-Sforza *et al.* 1994: 22–4).

Language: an ethnic guidebook?

Even if we allow that language plays an important part in ethnicity, the relationship between language and human biology remains problematic. Does linguistic similarity within an area promote endogamy or *vice versa*? Presumably a bit of both (as we see in Little England beyond Wales). One is usually dealing with a geographical common denominator rather than a direct one-to-one connection between language and 'tribe'.

In giving such prominence to language, Cavalli-Sforza, Menozzi and Piazza ignore the fact that bilingualism has been more the norm than the exception in world-wide terms. Between 70% and 80% of the world's population are said to be bilingual or multilingual, especially in areas *less* affected by European colonization (Trask 1996: 308). Yet from Ruhlen's figure of 4736 languages in the world Cavalli-Sforza and colleagues (1994: 22) calculate an average of a million speakers per language, apparently by a simple division. This is not

trivial criticism, because bilingualism is the prerequisite for language-shift, a phenomenon which has occurred on a massive scale not only in modern times, for example in Ireland, but also in antiquity, for example in France where Aquitanian, Gaulish and Frankish were all replaced by the form of Latin which became French. Cavalli-Sforza and colleagues underplay the importance of language-shift, and of language death which actually comes to almost the same thing: when languages die out, the populations who spoke them usually live on, but speaking other languages. What they say about language death and language shift is as follows (1994: 22–3):

Languages in the last category [i.e. the 'many [spoken] by only a few hundred or less than one hundred'] are likely to become extinct in a few generations. The same fate has already befallen a *number* of others, some of which were studied before their disappearance.

(1994: 157):

Language replacement is more likely to happen, *perhaps, in recent history*, and there are well-known examples of it.

These vague formulations (my italics) play down our ignorance of the scale of language death and language shift in the past (*cf.* Trask 1996: 324–5). In fact, it is not inconceivable, firstly, that more languages have died out in the last few millennia than at present exist (Robb 1993) and, secondly, that most languages are spoken by people to whom they are not ancestral in the long term; thus in Robb's computer-simulation of language diffusion and replacement (by a sort of random osmosis), only two out of his 64 'original' languages survive, and one of these ends up in sub-Saharan Africa, far from its 'homeland' in India (Robb 1991: 289). Dixon (1997: 29) has calculated that even if there were only a single human language 100,000 years ago, it would by now have spawned 10 million billion billion billion languages 'if IE-type splitting were the norm'. The obvious implication is surely not that the Indo-European rate of splitting (from 1 to 100 in 6000 years) must be unusual, as Dixon argues, but that language shift has been the rule rather than the exception.

The idea that languages are often non-ancestral obviously applies to modern world lan-

guages, such as English, but it may equally be true of minority languages. For example, we cannot — in the absence of evidence (perhaps from DNA?) — assume that the populations who currently speak Irish in the Gaeltacht have anything special in common, apart from language, with the Celtic populations who erected the first known Celtic-language ‘Leponitic’ inscriptions in north Italy in the middle of the 1st millennium BC, or with the *Keltoi* of the upper Danube in the Hallstatt period. (As non-evidence, note the Aran islanders illustrated by Coon 1939: plate 29, as ‘Hallstatt and Keltic Iron Age types’!) The last speakers of a language may be as distant biologically from its first speakers as it is possible to be. This is because languages (as opposed perhaps to Language itself) are cultural rather than biological phenomena. Cavalli-Sforza, Menozzi and Piazza are of course aware that language is an aspect of culture, but by stressing the ‘vertical’ transmission of culture, including language, from parent to child they attempt to explain why ‘congruence between genetic and linguistic evolution’ should be expected (1994: 380–81). This argument seems misleading and almost circular. While in the short run children do tend to share their parents’ language, religion, politics and style of dress and cooking, no evidence is produced to show that these are transmitted vertically over many generations, like genes, from remote paternal and maternal ancestors. Wherever we have historical and archaeological evidence, it is clear that culture in the widest sense, including language, travels horizontally as well as vertically. To assume that this has not always been so is to fall into the trap of Primitivism — the assumption that the past was simpler than the present (see Moore 1994). So convinced are many geneticists, however, that language is transmitted vertically, that a recent investigation of the extent to which ‘culture’ in sub-Saharan Africa has been transmitted by ‘demic diffusion’ — that is, vertically — rather than by cultural diffusion or environmental adaptation, uses ‘linguistic affiliation as an indirect measure of demic diffusion’ (Guglielmino *et al.* 1995: 7586).

The main reason why biologists have given linguistic evidence such precedence is, one suspects, the fact that ‘there exist phylogenetic classifications of languages, which in linguistics jargon are called “genetic”’ (Cavalli-Sforza

et al. 1994: 23). Indeed, Cavalli-Sforza and colleagues endorse Darwin’s belief that ‘a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world’ (1994: 381; *cf.* Zegura *et al.* 1990: 420; Renfrew 1992: 445–6). It has to be stressed that the resemblance between the family-tree model for languages (as currently understood) and modern genetic classifications of populations is very superficial. A language tree is like an old-fashioned family tree tracing an individual’s pedigree back through the male line *via* Charlemagne to Adam. Biologically, an individual inherits genes from two parents, four grandparents, and so on geometrically, so that the direct paternal line is of limited though not negligible genetic significance (relating to the Y-chromosome). By contrast, the paternalist linguistic family tree really means what it says, for according to classic philological theory, as laid down by the 19th-century Neogrammarians and still held by most linguists, there are no ‘mixed languages’ (*cf.* Hopper 1989; Trask 1996: 312; Dixon 1997: 11–13, 71). Thus English is defined as a descendant of the Germanic branch of Proto-Indo-European (on the basis of phonology, lexicon, and syntax in descending order), not as a hybrid of Germanic, Latin and French. This doctrine makes good sense in the context of comparative philology, at least in its primary aim of distinguishing ‘native’ from ‘loaned’ elements in languages in order to reconstruct ancestral languages such as Common Germanic and PIE. From this point of view, the non-Germanic elements in English are so much ‘noise’. For the social historian, however, and therefore for the population geneticist, they *are* significant: the English themselves are not necessarily basically ‘Germanic’ even though their language is defined as such. If the classic Indo-European rake-shaped family-tree is potentially misleading, so too, for the same reasons, are variants with many nodes such as those first by proposed by Schleicher (see Renfrew 1992: 450) and most recently by Warnow and her colleagues (1997; *cf.* Trask 1996: 368–70). None of these linguistic trees lend themselves to biological interpretations. (‘Wave-models’ and *Sprachbund* analyses are rather more helpful in suggesting some cross-language social interaction (*cf.* Trask 1996: 185–7, 315–17), as are the models advocated by Dixon 1997.) A ‘racial’ classification of languages, as mis-

takenly urged by Darwin, would never reveal that Sinhalese is closer to Irish than it is to Tamil; nor, conversely, can the linguistic contrast of Tamil *versus* Sinhalese and Irish be expected to have many genetic exponents.

It might be supposed that historians, geneticists, and others could adjust the linguistic data by giving an appropriate weighting to loan elements such as the French borrowings in English. Up to a point this would be valid; the Norman Conquest, as a demographic event, must have had genetic repercussions. There are two major problems, however. One is that we cannot make quantitative deductions from the linguistic facts (*pace* Ehret 1988: 573). For example, whereas there are innumerable medieval French loanwords in English, there are only a tiny handful of loanwords from British Celtic, even though many historians argue that ex-British Celtic-speakers were numerous in Anglo-Saxon England — far more than the number of French-speakers after 1066 — and perhaps even formed the great majority of the population of Anglo-Saxon England (Higham 1992). The general lesson to be derived from work in this field is that one cannot infer population sizes from linguistic outcomes (or *vice versa*, *pace* Sokal *et al.* 1993: 58–9), because other, irrecoverable sociolinguistic factors are involved. Thus the death of Frankish in France and the triumph of English in Britain may be due more to the relative prestige of Latin *versus* British Celtic than to the relative numbers of Franks and Anglo-Saxons involved (Charles-Edwards 1995: 731). The other problem is that we can rarely estimate the sizes of populations in the remoter past — even with the help of Domesday Book, this is a problem for England in 1086, and that is a unique resource. Hence modern estimates of the density of immigrations such as the Germanic ‘invasions’ that accompanied the fall of the Roman Empire (Whittaker 1994: 211–12) or the Viking settlement of England (Sawyer 1971: 154–76) vary widely. To take the late Roman Empire as an example, Drinkwater (1996: 23) notes that Whittaker’s (1994: 231) “‘bullish” estimate of the “huge number” of outsiders who crossed illegally into the Empire across the whole extent of its borders during the fourth century — 1,000,000 over 10,000 km. in the course of 100 years — works out at only one person per kilometer per year’. We can compare this with a population density in Gaul in

Julius Caesar’s time averaging just under 8 persons per square kilometre (my calculation from figures in Birkhan 1997: 190). If these figures are accurate — which may be doubted — the genetic impact of the 4th-century barbarian immigrations would have been slight (*cf.* Ammerman & Cavalli-Sforza 1984: 108). On the other hand, the migration of the Helvetii, observed by Caesar in 58 BC, involved over 300,000 people (Cunliffe 1997: 90; *cf.* Anthony 1990: 898–9).

Insofar as the geneticists have thought about these messy problems, they often seem to have assumed that such complications only became serious in the fairly recent past (post-Columbus? post-Prehistory?) and that ‘originally’ the ethnic situation was simpler — the fallacy of Primitivism. Thus Moore (1995: 531) has objected to the assumption of Cavalli-Sforza and his co-workers ‘that the year 1492 constitutes some kind of watershed in human migration and evolution before which the world distribution of genes, ethnicities, and languages was simpler’. Admittedly, some archaeological theorists have argued that large-scale language replacements such as we can document in Europe have only occurred since élites emerged in the last 5000 years (Cavalli-Sforza *et al.* 1988: 6005; Trask 1996: 325). This unverifiable view has not gone unchallenged (*cf.* Bateman *et al.* 1990: 9; Robb 1991; 1993) and, in any case, there is an unbridgeable gap between 3000 BC and much of the modern genetic and linguistic data deployed by Cavalli-Sforza and his colleagues.

Genetic and linguistic phylogeny

Some biologists maintain that there is an ‘intrinsic relation between genetics and language’ (e.g. Chen *et al.* 1995: 607). In the light of what has been said above, this seems a peculiar use of the word ‘intrinsic’. The reference is to the influential paper published in 1988 by Cavalli-Sforza and others, ‘Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological, and Linguistic Data’. This included a table correlating genetic and linguistic data on a world-wide scale by means of two trees, one for populations and the other for languages (also reproduced by Bateman *et al.* 1990: 3; Renfrew 1992: 466; Ruhlen 1994: 33; Moore 1994: 926; McMahon & McMahon 1995: 167; and Trask 1996: 401; revised in Cavalli-Sforza *et al.* 1994: 99). Some biologists and anthropologists (e.g.

Bateman *et al.* 1990; Moore 1994; Moore 1995; McMahon & McMahon 1995) have questioned both the criteria behind the division of mankind into 42 aboriginal 'populations' and the validity of the tree for these 'populations' as an evolutionary statement (cladogram) as opposed to a taxonomy (phenogram). Basically, the problems derive from the fact that humans cannot be divided into species, since they have always been able to interbreed. For the biological tree to work we have to assume a demographic background of continuous expansion, in which the various groups saw less and less rather than more and more of each other (Cavalli-Sforza *et al.* 1988: 6005). Yet, 'In view of the continuum of variation and the numerous interconnections resulting from gene flow', to quote Chen, Sokal and Ruhlen (1995: 610), '[t]here is a considerable question whether a dendrogram is an appropriate representation for the populations of the world'. Or, in A.R. Templeton's more trenchant words (*apud* Ward & Valencia 1996: 20):

The idea of a gene tree versus the population tree is a critical issue. Gene trees exist within a species, but there may not be a bifurcating population tree because of the exchange of genes by gene flow. . . by looking at mitochondrial DNA, we have found that there were recurrent genetic contacts between Africans, Europeans and Asians throughout their entire evolutionary history.

The 1988 linguistic tree is even more controversial, and most linguists would object to the inclusion of Greenberg's 'Amerind' and so on alongside proven families like Indo-European (Hopper 1989; Bateman *et al.* 1990: 4; Nichols 1990; McMahon & McMahon 1995; Trask 1996: 401–2, 406). Over and above this, we have the question of whether the congruence of the biological and linguistic trees is as 'remarkable' as is insisted by the authors (Cavalli-Sforza *et al.* 1988: 6002; 1994: 101, 157; and *apud* Bateman *et al.* 1990: 17–18), and by Penny and others (1993). It can be argued that the one-to-one fit between, say, the Nilo-Saharan 'population' (which is defined linguistically) and the 'Nilo-Saharan' language group is tautologous (even if 'Nilo-Saharan' were not a questionable language family: Trask 1996: 190) and that it is more significant that biological groups like 'Caucasoid' are split between several language families. Only the 'Amerind' language group corresponds neatly and exclusively to a bio-

logical grouping (S. Amerind+C. Amerind+N. Amerind), as noted by Bateman *et al.* (1990: 6), but the value of this correlation is damaged by the fact that 'Amerind', as a language family, is extremely dubious (McMahon & McMahon 1995; Trask 1996: 386–9; Nichols & Peterson 1996; Dixon 1997: 34–5) and presumably would not have been hypothesized in the first place were it not for the contiguity of the relevant languages — precisely the same geographical factor that presumably accounts for the genuine genetic similarity — however exactly it is to be explained — between the various Amerind populations. Moreover, 'with very few exceptions . . . genetic and linguistic differences [between native American populations] appear largely uncorrelated. . . , or negatively correlated. . . , which does not suggest coevolution of biological traits and languages' (Barbujani *et al.* 1994a: 150).

Thus the congruence the two trees, while mathematically undeniable (Penny *et al.* 1993), seems insignificant (*pace* Ruhlen 1994: viii, 32–4, 216–17, 273 and *passim*) when one bears in mind that geography has influenced not only the spread of genes and languages but also modern linguists' search for families like 'Nilo-Saharan' and 'Amerind'. Similar objections apply to the argument of Chen *et al.* (1995: 599, 607) that there is a significant correspondence between genes and languages for entities such as 'Afro-Asiatic', 'Australian' and so on.

The above criticism of Cavalli-Sforza and his colleagues' 1988 scheme is more far-reaching than Renfrew's mildly critical comment (1997: 82) that

the notional correspondence between linguistic change and linguistic phylogeny on the one hand, and the development of genetic diversity among human populations on the other (Cavalli-Sforza *et al.* 1988, Figure 6.1) is, I think, in part an illusory one [T]he correspondences arise more from a significant number of dispersal episodes in post-Pleistocene human history (in which influential demographic processes were involved, where linguistic and genetic replacement were indeed correlated) than from strictly comparable processes of linguistic and genetic evolution or equivalent rates of linguistic and genetic divergence.

While this criticism undermines the idea of an 'intrinsic relation between genetics and language', mentioned above, Cavalli-Sforza and his colleagues could find little to quarrel with in Renfrew's formulation, since they are all agreed

on the parallel 'vertical' transmission of language and genes rather than any illusory direct link between them. The real objection both to them and Renfrew is that the languages in question have not been proved to be related in ways remotely parallel to those alleged for the populations (Carstairs-McCarthy 1997). A more recent attempt to assess a speculative family tree for all the world's languages (one devised by Merritt Ruhlen) has found 'no genetic support for the proposed higher linguistic structure', and this despite the candid admission that 'Ruhlen might have been influenced by the existing genetic tree of human populations, especially at the highest level (i.e. the African/non-African separation)' (Chen *et al.* 1995: 609, 600). To date, then, attempts to graft genetic and linguistic world trees have not yet borne substantial fruit. Or, as Bodmer and Cavalli-Sforza put it back in 1976, 'Linguistic classifications are only moderately useful in defining human races' (Bodmer & Cavalli-Sforza 1976: 568).

Language and genes in Europe

In Europe linguistic history is better understood, which makes it a better testing-ground than the vast areas occupied by 'Nostratic', 'Proto-World' and so on. If we compare Cavalli-Sforza and his colleagues' 'genetic tree of 26 European populations based on 88 genes' (their 'first principle component') (1994: 268 and *apud* Bodmer 1993: 51) with the established linguistic family trees we see a poor correlation. For example, their 'Finnish' population (as defined and sampled) is no closer genetically to the 'Hungarian' than it is to the 'Irish' or 'Polish' populations, whereas linguistically Finnish and Hungarian are known to be related non-Indo-European languages. A way round this problem — which tends to undermine confidence in the predictive power of the whole exercise, however — is to argue that the Finnish-speaking population absorbed a larger Indo-European-speaking population (Piazza *et al.* 1995: 5838). Rather similarly, in a study of 14 linguistic groups in the Caucasus, the speakers of the only Indo-European languages, Armenian and Ossetic, turned out to be genetically very distant, forcing the researchers to the conclusion that genetic-linguistic correlations have been obscured by language replacement in hypothetical periods of 'elite dominance' (Barbujani *et al.* 1994b: 857). A further curiosity of this sort

is that in Gojobori's phylogenetic tree based on HLA-A and B genes (*apud* Bodmer 1993: 47), the 'British' are the closest relatives of the 'Hungarian' population, while the 'Cornish' are closest to the 'Swedish', completely contrary to the linguistic situation. Similarly, the above-mentioned 'genetic tree of 26 European populations' shows the 'French' population closer to neighbouring Germanic-speaking populations than it is to the 'Portuguese', 'Italian' and 'Spanish'. Admittedly, these last three Romance-speaking populations come out together, but the fact that they are contiguous countries effectively explains that away as a sort of pseudo-fit; their linguistic similarity is due to the spread of the Latin language among these populations, obliterating Venetic, Etruscan, Iberian, Tartessian, Celtiberian and so on, during the Roman Empire, the spread of which was itself subject to geophysical constraints. Contiguity is also relevant to the fact that the 'Scottish' and 'Irish' populations are genetically very close. This must surely be due to the geographical proximity which also led (very likely independently) to the spread of the Gaelic language to western Scotland in the 5th and 6th centuries AD, carried by an unknown number of settlers. It is difficult to be confident that the link between the 'Scottish' and 'Irish' in this tree indicates 'the distinctiveness derived from the Celtic ancestry of these populations' (Bodmer 1993: 52). Since the modern genetic data cited have no inherent time-depth, and since we are also ignorant about when Ireland and Scotland became in any sense 'Celtic', it is improper to attach any such ethnic or linguistic label to the biological correspondence, which could theoretically reflect intercourse over several millennia. Furthermore, it must be added that the present population of Scotland, late immigrants apart, descends from a mixture of British-Celtic-, Pictish-, Anglo-Saxon- and Gaelic-speakers; all such historical niceties are ignored by labelling them as 'Scottish' on the European genetic tree. When the data behind this 'genetic tree of 26 European populations' are mapped as the 'first principal component' (Cavalli-Sforza *et al.* *apud* Bodmer 1993: 53) we can see how the variation agrees with geographical distance and, in a very general way perhaps, with the diffusion of agriculture from Anatolia, rather than with what we actually *know* about European linguistic history. (Renfrew's theory about

the diffusion of Indo-European from Anatolia is of course a theory rather than a fact, and was originally suggested by Grahame Clark's map of the diffusion of farming (Renfrew 1973), as was the parallel explanation of the 'first principal component' by Ammerman and Cavalli-Sforza.)

Light was cast on the European situation by a study by Harding and Sokal (1988: 9372) demonstrating (on the basis of 3369 localities) that 'genetic distances between [speakers of] the European language families do not reflect their accepted linguistic relationships' and that 'affinities between modern European gene pools have been formed primarily by relatively short-range gene flow between geographically adjacent populations'. Examples of genetically related populations include Germanic and Celtic speakers, Slavic and Hungarian speakers and Turkic and Greek speakers. The exceptions were genuine linguistic and genetic outliers: Basque speakers; the Semitic speakers of Malta; and 'the Finnic language family . . . given its unique genetic profile by the inclusion of the Lapps'. Only in such exceptional cases does it seem reasonable to consider connecting modern linguistic peculiarities with ethnic prehistory. Even then considerable caution is needed, as emerges from a recent study of 'Genes and languages in Europe' by Sajantila and others (1995). After noting that the Basques, Sardinians, Icelanders, Finns and Saami (Lapps) are outliers when nuclear gene frequencies are analysed, they state (1995: 42):

The fact that the Basques, the Finns, and the Saami speak non-Indo-European languages . . . supports the notion that genes and languages coevolve.

This initial statement, which of course passes over the Indo-European languages in Iceland and Sardinia, ignores the fact that many other modern Indo-European-speakers who are not genetic outliers must be the descendants of non-Indo-European-speaking Etruscans, Iberians, Picts and others, so that the fit between eccentricity in genes and languages in the case of the Basques, Finns and Saami is more simply explained by their geographical isolation having inhibited the spread of mainstream Indo-European languages. Sajantila and colleagues are then puzzled by the problem that at the level of mitochondrial gene lineages, which they take to reflect an earlier stage of population history than nuclear gene frequencies, the Basques, the

Finno-Ugric-speaking Saami, Finns, Karelians, Estonians and Volga-Finnic peoples, and the Indo-European-speaking Icelanders, Sardinians, British and Swiss are all very similar with the sole exception of the Saami, whose 'unique position . . . in the genetic landscape of Europe could mean that they are an old population in Europe, which diverged from other European populations prior to subsequent linguistic and cultural diversification. Alternatively, they may have come to Europe from another, currently unknown region' (Sajantila *et al.* 1995: 49). The problem is therefore that (1995: 49)

no distinction according to language can be observed in the mitochondrial gene pool of Europe. . . . Rather, although the Saami languages are related most closely to Finnish and Estonian . . . , the mitochondrial lineages in the latter are related most closely to those in populations that speak languages belonging to other linguistic phyla than those in Saami. At first glance, this seems to contradict the coevolution of genes and languages demonstrated [by Cavalli-Sforza *et al.*] for nuclear markers.

The suggested solution is that all the European peoples under consideration, apart from the Saami, have the same comparatively recent ancestry, but that the Finns, Icelanders, Sardinians and Basques 'have undergone bottlenecks in population size in prehistoric times' leading to divergent nuclear gene frequencies through founder effects and genetic drift (Sajantila *et al.* 1995: 50). From the genetic point of view this is elegant (but note that the medieval 'colonization of Iceland from Scandinavia and the British Isles' was later than 'the Iron Age'). It does not, however, explain the linguistic situation. For this Sajantila and his colleagues simply suggest (1995: 50):

A tempting hypothesis is that populations that go through bottlenecks in size will tend to accelerated [*sic*] their rate of language evolution as a result of drift-like linguistic process. The result would be a congruence between trees based on language and gene frequency. In contrast, analyses of genetic lineages would tend to disagree with language trees. Such a disagreement has been observed not only in Europe, but also in the New World.

There is no evidence for such a linguistic tendency; for instance, Icelandic and Basque are generally regarded as very conservative languages, and in the case of the former this can

be measured relative to the other Germanic languages. Nor is it clear how it could explain the linguistic *status quo*. How could accelerated evolution make Basque differ from all other world languages but make Finnish become similar to Saami (and Hungarian)? Really, the linguistic hypothesis is only required by deference to the supposed demonstration of the congruence of linguistic and genetic trees in Cavalli-Sforza and his colleagues' paper 'Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological, and Linguistic Data' (1988), which we have already found wanting. At the deeper level of 'human evolution', of course, the above deductions from mitochondrial DNA undermine the thesis of that paper completely. (Note, however, the comment of W. Bodmer *apud* Ward & Valencia 1996: 24: 'People are concentrating far too much on mitochondria'.)

The interrelationship of genetics, language and geography in Europe was investigated in 1988 by Robert Sokal and colleagues on the basis of a very large sample of localities (mapped in Sokal 1988: 1723 = Sokal *et al.* 1988: 342). They found that although genetic distance correlated better with geographical distance than with linguistic difference, 'speakers of different language families in Europe differ genetically and that this difference remains even after geographic differentiation is allowed for' (Sokal 1988: 1722). In the absence of ancient DNA, it is not clear that we can decide how far this is due to long-standing biological differences between differing ethno-linguistic groups and how far to the effects of language on ethnic self-definition and choice of marriage partners. Sokal himself has often expressed this chicken-and-egg problem as follows: 'since language differences themselves are likely to be barriers to free gene flow, they will enhance genetic differentiation'; this 'may influence the results *in favor* of a language-gene frequency association' (1988: 1722; Sokal *et al.* 1988: 338, his italics; 1989: 490; Chen *et al.* 1995: 607). Note furthermore that some language boundaries are also political boundaries (Sokal *et al.* 1988: 353). The linguistic boundary which proved genetically strongest (chiefly owing to HLA-B differences) is Germanic-Romance, which runs from the Channel Islands along the English Channel and then southwards through Belgium to Italy (Sokal *et al.* 1988: 340, 345, 360). Yet this cannot be a very ancient linguistic boundary — Germanic only reached Britain in the 5th century AD — and it does not

replicate any known earlier boundary such as Germanic-Celtic either. It is therefore difficult to accept these researchers' extremely prehistoric explanation for modern 'rapid gene change across language boundaries' (Sokal *et al.* 1988: 351–2):

We assume that this correspondence occurs because speakers representing a major language family, such as the Slavic or Celtic languages, trace their ancestry back in time to a relatively small group of persons who were geographically isolated for long periods from other nascent Indo-Europeans. As a result of their isolation and small numbers, they must have become linguistically and genetically distinct.

The chicken-and-egg problem noted above also applies to Barbujani and Sokal's demonstration — or perhaps claim — that 'Zones of sharp genetic change in Europe are also linguistic boundaries' (Barbujani & Sokal 1990). Their remarkable map of genetic boundaries (also in Cavalli-Sforza *et al.* 1994: 271) raises many interesting problems for the historical linguist. The genetic boundary between Ireland and Iceland (§8) is listed as linguistically 'English-Icelandic', but that between 'England and Wales vs. Ireland' (§11) as 'Celtic-Germanic, in part' (*sic*). So Ireland counts as an English- and Germanic-speaking country, a ruthlessly modern viewpoint. On the other hand, Iberia shows no genetic cleavage corresponding to modern Portuguese and Spanish, but it does show a cleavage between 'northwestern (*sic, recte* northern?) Iberia vs. the rest' (§21), labelled 'Basque-Romance, in part'. The genetic division between western and eastern Iceland is interpreted in terms of a thousand-year old differential settlement from the British Isles and Norway respectively (Barbujani & Sokal 1990: 1818, *cf.* Pálsson 1986: 290; Sigurðsson 1988: 27). In the selection of ancient and modern linguistic and historical data, there seems to be an element of having one's cake and eating it (Mallory 1992: 145–8). A further problem is that about two-thirds of the linguistic boundaries correspond with natural geographical barriers such as the North Sea. Of the third that do not (§§3, 4, 5, 7(?), 14, 15, 19, 28, 31), several are fairly arbitrary as linguistic boundaries (e.g. Dutch vs. German, central vs. southern Germany, and northern vs. southern Italian). Of those that remain (all involving Finno-Ugric), Piazza and colleagues (1995: 5840) comment that: 'It remains to be established . . . if linguistic bounda-

ries have generated or enhanced genetic boundaries or if both are the consequence of political, cultural, and social boundaries (as in the case of Lapps and non-Lapps) that have played a role similar to that of physical boundaries.'

These methods were applied to the British Isles in particular by Falsetti and Sokal (1993). They found 12 genetic boundaries, of which 'all but three (boundaries 6, 7 and 8) are substantial physical boundaries' (1993: 226–7). All three are hard to interpret. Boundary 6, which 'isolates the peninsula of Cornwall' (1993: 222, i.e. Devon and Cornwall), could be attributed to isolation rather than to its Dark Age Brittonic-speaking population; Falsetti & Sokal do not comment on it, but see Cole 1997. Boundary 7 'separates Suffolk into north and south regions' and Falsetti & Sokal suggest (1993: 227) that it stems from a difference in founding Anglian and Jutish populations, which is simply incorrect: the Jutes are only recorded in Kent and Hampshire (Bassett 1989). Finally, Boundary 8 'separates Yorkshire and Humberside from the East Midlands and East Anglia' (Falsetti & Sokal 1993: 222), a puzzling division for which Falsetti and Sokal offer no explanation. From a linguistic and historical point of view these results are disappointing. On the one hand, population genetics is an expensive way of telling us what we already knew (to paraphrase Peter Sawyer's famous jibe about medieval archaeology). On the other, it presents us with puzzles, such as Boundary 8, without the chronological clues needed to solve them.

In their subsequent research Sokal and his group have attempted to take account of ethnohistory, which in practice means the history or hypothetical prehistory of language families (Sokal *et al.* 1993: 58), reaching the conclusion 'that there is a correlation between ethnohistorical and genetic distances above that brought about by the common factor, geography' (Sokal *et al.* 1993: 62). This seems plausible — it would be surprising if ethnicity were wholly subjective — but the *degree* to which it is true cannot be estimated, given the inadequacy of their ethnohistorical data (as noted above, p. 506). This sort of work has been criticized by the archaeologist Mark Pluciennik (1996: 13) for 'historical and archaeological *naïveté*'. Pluciennik also complains about 'the lack of time resolution inherent in genetic studies of modern

populations'. Here one hopes that collaboration between archaeologists in the field and biologists in the laboratory will eventually prove decisive.

Some conclusions

One way forward is the analysis of ancient DNA, which has been producing remarkable results for some years (e.g. Handt *et al.* 1994; Richards *et al.* 1996: 193). Another is the expectation that some genetic variations will be interpreted in terms of mutations from a reconstructable ancestral starting-point, and the hope that some of these mutations may be *proved* to occur at a regular rate (*cf.* Goldstein *et al.* 1995; Sajantila *et al.* 1995: 50; Underhill *et al.* 1996: 199; Richards *et al.* 1996: 186, 197; Stenico *et al.* 1996: 1370). One hopes that such advances will eventually enable us to realize what Sir Walter Bodmer has described as 'the enormous power of combining humanistic studies of human populations with the scientific investigation of genetic differences' (Bodmer 1993: 57). As the many examples given above demonstrate, this potential may never be realized without more critical input from specialists in the humanities. To be credible, geneticists — and archaeologists and linguists — need to admit the uncertainty of our knowledge about the prehistory of even so well-known a linguistic family as Indo-European, and to detach themselves from the 'Neanderthal Loan-words in Proto-World' school of philology. Geneticists need to be beware of using linguistic and archaeological data as surrogates for ethnic data, and to throw off Primitivist assumptions about the relative simplicity of the past. The chronological ambiguity of gene-frequency maps needs to be reiterated, and apparent correlations of genetic and linguistic data need to be subjected to more stringent tests. Finally, the comparisons which have been made between linguistic and genetic trees should be recognized as flawed in principle and in practice. A convincing synthesis of genetics, linguistics and prehistoric archaeology is still some way off.

Note. This article is based on two lectures given at the Royal Irish Academy, one at the 1996 Colloquium of the Indo-European Sub-Committee of the National Committee for Greek and Latin Studies and the other at the 1997 Consultative Meeting of the Royal Irish Academy Advisory Committee on Genetic Anthropology.

References

- AMMERMAN, A.J. & L.L. CAVALLI-SFORZA. 1984. *The Neolithic transition and the genetics of populations in Europe*. Princeton (NJ): Princeton University Press.
- ANTHONY, D.W. 1990. Migration in archeology: the baby and the bathwater. *American Anthropologist* 92: 895–914.
- ANTHONY, D.W. & B. WAILLES. 1988. Review of Renfrew 1987, *Current Anthropology* 29: 437–68.
- BARBUJANI, G., A. PILASTRO, S. DE DOMENICO & C. RENFREW. 1994a. Genetic variation in North Africa and Eurasia: neolithic demic diffusion vs. paleolithic colonisation, *American Journal of Physical Anthropology* 95: 137–54.
- BARBUJANI, G. & R.R. SOKAL. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries, *Proceedings of the National Academy of Sciences* 87: 1816–19.
- BARBUJANI, G., R.R. SOKAL & N.L. ODEN. 1995. Indo-European origins: a computer-simulation test of five hypotheses, *American Journal of Physical Anthropology* 96: 109–32.
- BARBUJANI, G., C.N. WHITEHEAD, G. BERTORELLE & I.S. NASIDZE. 1994b. Testing hypotheses on processes of genetic and linguistic change in the Caucasus, *Human Biology* 66: 843–64.
- BASSETT, S. (ed.). 1989. *The origins of Anglo-Saxon kingdoms*. London: Leicester University Press.
- BATEMAN, R., I. GODDARD, R. O'GRADY, V.A. FUNK, R. MOOI, W.J. KRESS & P. CANNELL. 1990. Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language, *Current Anthropology* 31: 1–24, 175–83.
- BEEKES, R.S.P. 1995. *Comparative Indo-European linguistics: an introduction*. Amsterdam: John Benjamins.
- BELLWOOD, P. 1996. Phylogeny vs reticulation in prehistory, *Antiquity* 70: 881–90.
- BERTRANPETIT, J., J. SALA, F. CALAFELL, P.A. UNDERHILL, P. MORAL & D. COMAS. 1995. Human mitochondrial DNA variation and the origin of Basques, *Annals of Human Genetics* 59: 63–81.
- BIRKHAN, H. 1997. *Kelten: Versuch einer Gesamtdarstellung ihrer Kultur*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- BLENCH, R. & M. SPRIGGS (ed.). 1997. *Archaeology and language I: Theoretical and methodological orientations*. London: Routledge. One World Archaeology 27.
- BODMER, W.F. 1993. The genetics of Celtic populations, *Proceedings of the British Academy* 82: 37–57.
- BODMER, W.F. & L.L. CAVALLI-SFORZA. 1976. *Genetics, evolution, and man*. San Francisco (CA): W. H. Freeman.
- BOWCOCK, A.M., J.R. KIDD, J.L. MOUNTAIN, J.M. HEBERT, L. CAROTENUTO, K.K. KIDD & L.L. CAVALLI-SFORZA. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms, *Proceedings of the National Academy of Sciences* 88: 839–43.
- BROWN, T. (ed.). 1996. *Celticism*. Amsterdam: Rodopi.
- CARSTAIRS-MCCARTHY, A. 1997. Review of Merritt Ruhlen, *The origin of language: tracing the evolution of the mother tongue*, *Language* 73: 611–14.
- CARTER, H. 1986. Population movements into Wales: an historical review, in Harper & Sunderland 1986: 31–53.
- CAVALLI-SFORZA, L.L., P. MENOZZI & A. PIAZZA. 1994. *The history and geography of human genes*. Princeton (NJ): Princeton University Press.
- 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* 85: 6002–6.
- CHADWICK, D. & G. CARDEW (ed.). 1996. *Variation in the human genome*. Chichester: John Wiley and Sons.
- CHAPMAN, M. 1992. *The Celts: the construction of a myth*. Basingstoke: Macmillan.
- CHARLES-EDWARDS, T. 1995. Language and society among the Insular Celts, AD 400–1000, in Green 1995: 703–36.
- CHEN, J., R.R. SOKAL & M. RUHLEN. 1995. Worldwide analysis of genetic and linguistic relationships of human populations, *Human Biology* 67: 595–612.
- CLACKSON, J. 1994. *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.
- COLE, D. 1997. The Cornish: identity and genetics — an alternative view, *Cornish Studies new series* 5: 21–9.
- COLLINS, R. 1986. *The Basques*. Oxford: Blackwell.
- COLLIS, J. 1996. The origin and spread of the Celts, *Studia Celtica* 30: 17–34.
1997. Celtic myths, *Antiquity* 71: 195–201.
- COON, C.S. 1939. *The races of Europe*. New York (NY): Macmillan.
- CÓRTE-REAL, H.B.S.M., V.A. MACAULAY, M.B. RICHARDS, G. HARITI, M.S. ISSED, A. CAMBON-THOMSEN, S. PAPIHA, J. BERTRANPETIT & B.C. SYKES. 1996. Genetic diversity in the Iberian peninsula determined from mitochondrial sequence analysis, *Annals of Human Genetics* 60: 331–50.
- CUNLIFFE, B. 1997. *The ancient Celts*. Oxford: Oxford University Press.
- DANIEL, G.E. 1954. Who are the Welsh?, *Proceedings of the British Academy* 40:145–67.
- DAVIES, W. 1990. *Patterns of power in early Wales*. Oxford: Clarendon Press.
- DAWSON, G.W.P. 1964. The frequency of the ABO and Rh (D) blood groups in Ireland from a sample of 1 in 18 of the population, *Annals of human genetics* 28: 49–59.
- DIXON, R.M.W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- DRINKWATER, J.F. 1996. The 'Germanic threat on the Rhine frontier': a Romano-Gallic artifact?, in R.W. Mathisen & H.S. Sivan (ed.), *Shifting frontiers in late antiquity*: 20–30. Aldershot: Variorum.
- EHRET, C. 1988. Language change and the material correlates of language and ethnic shift, *Antiquity* 62: 564–74.
- FALSETTI, A.B. & R.R. SOKAL. 1993. Genetic structure of human populations in the British Isles, *Annals of Human Biology* 20: 215–29.
- FORSYTH, K. 1997. *Language in Pictland: the case against 'non-Indo-European Pictish'*. Utrecht: de Keltische Draak/Münster: Nodus Publikationen.
- FRANKLIN, M.J. (ed.). 1995. *Sir William Jones: selected poetical and prose works*. Cardiff: University of Wales Press.
- FRASER, A. 1992. *The Gypsies*. Oxford: Blackwell.
- GOLDSTEIN, D.B., A.R. LINARES, L.L. CAVALLI-SFORZA & M.W. FELDMAN. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans, *Proceedings of the National Academy of Sciences* 92: 6723–7.
- GREEN, M.J. (ed.). 1995. *The Celtic world*. London: Routledge.
- GUGLIELMINO, C.R., C. VIGANOTTI, B. HEWLETT & L.L. CAVALLI-SFORZA. 1995. Cultural variation in Africa: role of mechanisms of transition and adaptation, *Proceedings of the National Academy of Sciences* 92: 7585–9.
- HALL, J.M. 1997. *Ethnic identity in Greek antiquity*. Cambridge: Cambridge University Press.
- HAMP, E.P. 1982. [Review] "The Proto-Algonquians." By Willard Walker. *International Journal of American Linguistics* 48: 344–56.
- 1990a. The Indo-European horse, in Markey & Greppin (ed.): 211–26.
- 1990b. The Pre-Indo-European language of Northern (Central) Europe, in Markey & Greppin (ed.): 291–309.
- 1990c. carn, *Bulletin of the Board of Celtic Studies* 37: 111.
- HANDT, O., M. RICHARDS, M. TROMMSDORFF, C. KILGER, J. SIMANAINEN, O. GEORGIEV, K. BAUER, A. STONE, R. HEDGES, W. SCHIAFFNER, G. UTERMANN, B. SYKES & S. PÄÄBO. 1994. Molecular genetic analyses of the Tyrolean ice man, *Science* 264: 1775–8.
- HARDING, R.M., & R.R. SOKAL. 1988. Classification of the European language families by genetic distance, *Proceedings of the National Academy of Sciences* 85: 9370–72.
- HARPER, P.S. 1986. Mendelian disorders in Wales, in Harper & Sunderland (ed.): 273–89.
- HARPER, P.S. & E. SUNDERLAND (ed.). 1986. *Genetic and population studies in Wales*. Cardiff: University of Wales Press.

- HARRIS, A.C. 1990. Kartvelian contacts with Indo-European, in Markey & Greppin (ed.): 67–100.
- HEGEDŰS, I. 1997. Principles for palaeolinguistic reconstruction, in Blench & Spriggs (ed.): 65–73.
- HIGHAM, N. 1992. *Rome, Britain and the Anglo-Saxons*. London: Seaby.
- HINES, J. 1991. Review of Renfrew 1987, Mallory 1989, and Markey & Greppin 1990, *Norwegian Archaeological Review* 24: 49–54.
- HOPPER, P.J. 1989. Review of Sarah Grey Thomason & Terrence Kaufman, *Language contact, creolization, and genetic linguistics*, *American Anthropologist* 91: 817–18.
- HUTCHISON, W.G. (trans.). 1896. *The Poetry of the Celtic Races, and other studies by Ernest Renan*. London: Walter Scott.
- JACKSON, K.H. 1980. The Pictish language, in F.T. Wainwright (ed.), *The problem of the Picts*: 129–66, 173–6. Perth: Melven Press. 2nd edition.
- JONES, S. 1997. *The archaeology of ethnicity: constructing identities in the past and present*. London: Routledge.
- KITSON, P.R. 1996. British and European river-names, *Transactions of the Philological Society* 94: 73–118.
- LLOYD-JONES, J. 1928. *Enwau Lleoedd Sir Gaernarfon*. Cardiff: University of Wales Press.
- LOIRAT, F., S. HAZOUT & G. LUCOTTE. 1997. G542X as a probable Phoenician cystic fibrosis mutation, *Human Biology* 69: 419–25.
- LOYN, H. 1994. *The Vikings in Britain*. Oxford: Blackwell.
- LUCOTTE, G., & S. HAZOUT. 1995. Geographic and ethnic distributions of the more frequent cystic fibrosis mutations in Europe show that a founder effect is apparent for several mutant alleles, *Human Biology* 67: 561–76.
- MAC AN BHAIRD, A. 1991. Ptolemy revisited, *Aiam* 5 (1991–93): 1–20.
- MALLORY, J.P. 1989. *In search of the Indo-Europeans*. London: Thames & Hudson.
1992. Human populations and the Indo-European problem, *Mankind Quarterly* 33: 131–54.
1996. The Indo-European homeland problem: a matter of time, in K. Jones-Bley & M.E. Huld (ed.), *The Indo-Europeanization of Northern Europe*: 1–22. Washington (DC): Institute for the Study of Man.
1997. The homelands of the Indo-Europeans, in Blench & Spriggs (ed.): 93–121.
- MARKEY, T.L. & J.A.C. GREPPIN (ed.). 1990. *When worlds collide: the Indo-Europeans and the Pre-Indo-Europeans*. Ann Arbor (MI): Karoma Publishers.
- 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.
- MOORE, J.H. 1994. Putting anthropology back together again: the ethnogenetic critique of cladistic theory, *American Anthropologist* 96: 925–48.
1995. The end of a paradigm [review of Cavalli-Sforza *et al.* 1994], *Current Anthropology* 36: 530–31.
- NICHOLS, J. 1990. More on human phylogeny and linguistic history, *Current Anthropology* 31: 313–14.
1997. The epicentre of the Indo-European linguistic spread, in Blench & Spriggs (ed.): 122–48.
- NICHOLS, J. & D.A. PETERSON. 1996. The Amerind personal pronouns, *Language* 72: 336–71.
- NYBERG, T., I. PIÖ, P. MEULENGRACHT SØRENSEN & A. TROMMER (ed.). 1985. *History and heroic tale: a symposium*. Odense: Odense University Press.
- Ó CORRÁIN, D. 1985. Irish origin legends and genealogy: recurrent aetiologies, in Nyberg *et al.* (ed.): 51–96.
- O'RAHILLY, T.F. 1946. *Early Irish history and mythology*. Dublin: Dublin Institute for Advanced Studies.
- PÁLSSON, J.Ó.P. 1986. Die ethnogenese der Nordgermanen aus der Sicht anthropologischer Studien an Lebenden, unter besonderer Berücksichtigung der Isländer, in W. Bernhard & A. Kandler-Pálsson (ed.), *Ethnogenese europäischer Völker*: 285–96. Stuttgart: Gustav Fischer Verlag.
- PENNY, D., E.E. WATSON & M.A. STEEL. 1993. Trees from languages and genes are very similar, *Systematic Biology* 42: 382–4.
- PIAZZA, A., S. RENDINE, E. MINCH, P. MENOZZI, J. MOUNTAIN & L.L. CAVALLI-SFORZA. 1995. Genetics and the origin of European languages, *Proceedings of the National Academy of Sciences* 92: 5836–40.
- PINTO, F., A.M. GONZÁLEZ, M. HERNÁNDEZ, J.M. LARRUGA & V.M. CABRERA. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences, *Annals of Human Genetics* 60: 321–30.
- PLUCIENNIK, M. 1996. Genetics, archaeology and the wider world, *Antiquity* 70: 13–14.
- POKORNY, J. 1950. The river-name *Argita*, *Journal of Celtic Studies* 1: 130.
- POLOME, E.C. 1990. Types of linguistic evidence for early contact: Indo-Europeans and Non-Indo-Europeans, in Markey & Greppin (ed.): 267–89.
- POTTS, W.T.W. 1976. History and blood groups in the British Isles, in Sawyer (ed.): 236–53.
- PSICHARI, H. (ed.). 1948. *Oeuvres complètes de Ernest Renan, II: Essais de morale et de critique*. Paris: Calmann-Lévy.
- RAFTERY, B. 1995. Ireland: a world without the Romans, in Green (ed.): 636–53.
- RELETHFORD, J.H. & M.H. CRAWFORD. 1995. Anthropometric variation and the population history of Ireland, *American Journal of Physical Anthropology* 96: 25–38.
- RENFREW, C. 1973. Problems in the general correlation of archaeological and linguistic strata in prehistoric Greece: the model of autochthonous origin, in R.A. Crossland & A. Birchall (ed.), *Bronze Age migrations in the Aegean*: 263–79. London: Duckworth.
1987. *Archaeology and language: the puzzle of Indo-European origins*. London: Jonathan Cape.
1989. Models of change in language and archaeology, *Transactions of the Philological Society* 87: 103–55, 166–78.
1990. Archaeology and linguistics: some preliminary issues, in Markey & Greppin (ed.): 15–24.
1992. Archaeology, genetics and linguistic diversity, *Man* n.s. 27: 445–78.
1997. World linguistic diversity and farming dispersals, in Blench & Spriggs (ed.): 82–90.
- RICHARDS, M., H. CÔRTE-REAL, P. FORSTER, V. MACAULAY, H. WILKINSON-HERBOTS, A. DEMAINE, S. PAPIHA, R. HEDGES, H.-J. BANDELT & B. SYKES. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool, *American Journal of Human Genetics* 59: 185–203.
- RIVET, A.L.F. & C. SMITH. 1979. *The place-names of Roman Britain*. London: Batsford.
- ROBB, J. 1991. Random causes with directed effects: the Indo-European language spread and the stochastic loss of lineages, *Antiquity* 65: 287–91.
1993. A social prehistory of European languages, *Antiquity* 67: 747–60.
- ROOSENS, E.E. 1989. *Creating ethnicity: the process of ethnogenesis*. London: Sage.
- RUHLEN, M. 1994. *On the origin of languages*. Stanford (CA): Stanford University Press.
- SAJANTILA, A., P. LAHERMO, T. ANTINEN, M. LUKKA, P. SISTONEN, M.-L. SAVONTAUS, P. AULA, L. BECKMAN, L. TRANEBJAERG, T. GEDDE-DAHL, L. ISSEL-TARVER, A. DIRIENZO & S. PÄÄBO. 1995. Genes and language in Europe: an analysis of mitochondrial lineages, *Genome* 5: 42–52.
- SAMUELS, M.L. 1972. *Linguistic evolution with special reference to English*. Cambridge: Cambridge University Press.
- SAWYER, P. 1971. *The age of the Vikings*. London: Edward Arnold. 2nd edition.
- (Ed.). 1976. *Medieval settlement: continuity and change*. London: Arnold.
- SCHMIDT, K.H. 1990. The postulated Pre-Indo-European substrates in Insular Celtic and Tocharian, in Markey & Greppin (ed.): 179–202.

1996. *Celtic: a Western Indo-European language?* Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- SCHRIJVER, P. 1995. *Studies in British Celtic historical phonology*. Amsterdam: Rodopi.
1997. Review of J.F. Eska et al. (ed.), *Hispano-Gallo-Brittonica, Cambrian Medieval Celtic Studies* 34 (Winter 1997): 107–11.
- SCRIVER, C.R., S. BYCK, L. PREVOST, L. HOANG & THE PAH MUTATION ANALYSIS CONSORTIUM. 1996. The phenylalanine hydroxylase locus: a marker for the history of phenylketonuria and human genetic diversity, in Chadwick & Cardew (ed.): 73–96.
- SIGURÐSSON, G. 1988. *Gaelic influence in Iceland: historical and literary contacts. A survey of research*. Reykjavík: Bókaútgáfa Menningarsjóðs.
- SIMS-WILLIAMS, P. 1985. Some functions of origin stories in early medieval Wales, in Nyberg et al. 1985: 97–131.
1997. The Celtic languages, in A.G. Ramat & P. Ramat (ed.), *The Indo-European languages*: 345–79. London: Routledge. In press. Celtomania and Celtocepticism, *Cambrian Medieval Celtic Studies* 36 (Winter 1998).
- SOKAL, R.R. 1988. Genetic, geographic, and linguistic distances in Europe, *Proceedings of the National Academy of Sciences* 85: 1722–6.
- SOKAL, R.R., G.M. JACQUEZ, N.L. ODEN, D. DIGIOVANNI, A.B. FALSETTI, E. MCGEE & B.A. THOMSON. 1993. Genetic relationships of European populations reflect their ethnohistorical affinities, *American Journal of Physical Anthropology* 91: 55–70.
- SOKAL, R.R., N.L. ODEN, P. LEGENDRE, M.J. FORTIN, J. KIM & A. VAUDOR. 1989. Genetic differences among language families in Europe, *American Journal of Physical Anthropology* 79: 489–502.
- SOKAL, R.R., N.L. ODEN & B.A. THOMSON. 1988. Genetic changes across language boundaries in Europe, *American Journal of Physical Anthropology* 76: 337–61.
1992. Origins of the Indo-Europeans: genetic evidence, *Proceedings of the National Academy of Sciences* 89: 7669–73.
- SOKAL, R.R., N.L. ODEN, J. WALKER, D. DIGIOVANNI & B.A. THOMSON. 1996. Historical population movements in Europe influence genetic relationships in modern samples, *Human Biology* 68: 873–98.
- SOKAL, R.R., N.L. ODEN & C. WILSON. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion, *Nature* 351: 143–5.
- STENICO, M., L. NIGRO, G. BERTORELLE, F. CALAFELL, M. CAPITANIO, C. CORRAIN & G. BARBUJANI. 1996. High mitochondrial sequence diversity in linguistic isolates of the Alps, *American Journal of Human Genetics* 59: 1363–75.
- SUNDERLAND, E. 1976. Comment on 'History and blood groups in the British Isles' by W.T.W. Potts, in Sawyer (ed.): 254–61.
- SUPER, R.H. (ed.). 1962. *The complete prose works of Matthew Arnold III: Lectures and essays in criticism*. Ann Arbor (MI): University of Michigan Press.
- SZEMERÉNYI, O.J.L. 1985. Recent developments in Indo-European linguistics, *Transactions of the Philological Society* 1985: 1–71.
1989. Concerning Professor Renfrew's views on the Indo-European homeland, *Transactions of the Philological Society* 87: 156–65.
- THOMAS, A.R. 1986. The linguistic background, in Harper & Sunderland (ed.): 54–76.
- THURNEISEN, R. 1946. *A grammar of Old Irish*. Dublin: Dublin Institute for Advanced Studies.
- TOORIANS, L. 1990. Wizo Flandrensis and the Flemish settlement in Pembrokeshire, *Cambridge Medieval Celtic Studies* 20: 99–118.
1996. Flemish settlements in twelfth-century Scotland, *Revue belge de philologie et d'histoire* 74: 659–93.
- TRASK, R.L. 1996. *Historical linguistics*. London: Arnold.
1997. *The history of Basque*. Oxford: Blackwell.
- UNDERHILL, P.A., L. JIN, R. ZEMANS, P.J. OEFNER & L.L. CAVALLI-SFORZA. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history, *Proceedings of the National Academy of Sciences* 93: 196–200.
- UPTON, G.J.G. & B. FINGLETON. 1985. *Spatial data analysis by example I: Point pattern and quantitative data*. Chichester: John Wiley.
- VENNEMANN, T. 1994. Linguistic reconstruction in the context of European prehistory, *Transactions of the Philological Society* 92: 215–84.
- WADDELL, J. 1995. Celts, Celticisation and the Irish Bronze Age, in J. Waddell & E. Shee Twohig (ed.), *Ireland in the Bronze Age*: 158–69. Dublin: Stationery Office.
- WAIT, G.A. 1995. Burial and the otherworld, in Green (ed.): 489–511.
- WARD, R.H. & D. VALENCIA. 1996. Phylogeographic variability in traditional societies, in Chadwick & Cardew (ed.): 6–24.
- WARNER, R.B. 1991. Cultural intrusions in the Early Iron Age: some notes, *Emania* 9: 44–52.
- WARNOW, T. 1997. Mathematical approaches to comparative linguistics, *Proceedings of the National Academy of Sciences* 94: 6585–90.
- WATKIN, I. M. 1986. ABO blood group distribution in Wales in relation to human settlement, in Harper & Sunderland (ed.): 118–46.
- WENG, Z. & R.R. SOKAL. 1995. Origins of Indo-Europeans and the spread of agriculture in Europe: comparison of lexicostatistical and genetic evidence, *Human Biology* 67: 577–94.
- WHITTAKER, C.R. 1994. *Frontiers of the Roman Empire*. Baltimore (MD): John Hopkins University Press.
- WILLIAMS, E.M. 1986. Genetic studies of Welsh gypsies, in Harper & Sunderland (ed.): 186–211.
- WILLIAMS, W.R. 1986. Dermatoglyphic studies in the mid Wales borderland, in Harper & Sunderland (ed.): 212–21.
- WOOLLEY, V. 1986. Demographic studies in Pembrokeshire, in Harper & Sunderland (ed.): 236–50.
- WRIGHT, N. 1984. Gildas's geographical perspective: some problems, in M. Lapidge & D. Dumville (ed.), *Gildas: new approaches*: 85–105. Woodbridge: Boydell Press.
- ZEGURA, S.L., W.H. WALKER, K.K. STOUT & J.D. DIAMOND. 1990. More on genes, language, and human phylogeny, *Current Anthropology* 31: 420–26.
- ZIMMER, S. 1990. The investigation of Proto-Indo-European history: methods, problems, limitations, in Markey & Greppin (ed.): 311–44.
- ZSCHOCKE, J., J.P. MALLORY, H.G. EIKEN & N.C. NEVIN. 1997. Phenylketonuria and the peoples of Northern Ireland, *Human Genetics* 100: 189–94.
- ZVELEBIL, M. & K.V. ZVELEBIL. 1990. Agricultural transition, 'Indo-European origins' and the spread of farming, in Markey & Greppin (ed.): 237–66.