Linguistics for Archaeologists: Principles, Methods and the Case of the Incas

Paul Heggarty

Like archaeologists, linguists and geneticists too use the data and methods of their disciplines to open up their own windows onto our past. These disparate visions of human prehistory cry out to be reconciled into a coherent holistic scenario, yet progress has long been frustrated by interdisciplinary disputes and misunderstandings (not least about Indo-European). In this article, a comparative-historical linguist sets out, to his intended audience of archaeologists, the core principles and methods of his discipline that are of relevance to theirs. They are first exemplified for the better-known languages of Europe, before being put into practice in a lesser-known case-study. This turns to the New World, setting its greatest indigenous ‘Empire’, that of the Incas, alongside its greatest surviving language family today, Quechua. Most Andean archaeologists assume a straightforward association between these two. The linguistic evidence, however, exposes this as nothing but a popular myth, and writes instead a wholly new script for the prehistory of the Andes — which now awaits an archaeological story to match.

1. Archaeology and linguistics

Recent years have seen growing interest in a ‘new synthesis’ between the disciplines of archaeology, genetics and linguistics, at the point where all three intersect in seeking a coherent picture of the distant origins of human populations. Such research has been dogged, however, by interdisciplinary controversy, such as that surrounding Renfrew’s (1987) hypothesis that the Indo-European languages originated in Anatolia. The logic behind his proposal is that the dominance and territorial range achieved by this single language family are so striking as to demand explanation in terms of some ‘empowering’ cultural force of commensurate significance, whose impact should therefore be clearly visible in the archaeological record too. The motor that Renfrew looks to for the early expansion of Indo-European is indeed a simple, powerful one: agriculture. This association, hotly debated, illustrates one of the key levels on which linguistics and archaeology necessarily interact.

Controversies have continued into the latest phase of the new synthesis, which has seen phylogenetic methods from the biological sciences now being applied to language data. In a much discussed paper in Nature, Gray & Atkinson (2003) put forward a phylogenetic analysis of Indo-European languages, based on comparative vocabulary lists, which falls neatly in line with the ‘out of Anatolia’ hypothesis. In particular they propose a new dating mechanism which supports Renfrew’s chronology, much earlier than the one traditionally adhered to by most Indo-Europeanist linguists, who have duly reacted with considerable scepticism.

Recurrent in the controversies have been suggestions, from one side or the other, of a number of serious misunderstandings between the disciplines. In such a context, this article seeks to contribute one part of the story, by setting out to an audience of archaeologists the basic principles and methods of linguistics that are of relevance and value to their own discipline. We look at how it is that data from languages can be analysed so as to enable us to look into the past, specifically the prehistory of the human populations that spoke those languages. The focus is on the key questions of how, and how reliably, we can establish...
both time-frames and locations for the various stages in the expansions of major language families.

On a general level, then, this article aims to promote a greater understanding of the main issues to be faced whenever we seek to correlate linguistic and archaeological scenarios for the history of any populations. The discussion here is therefore relevant and applicable to all cases, including the best studied and most frequently discussed ones like the Austronesian and Indo-European expansions — the former a paradigm case of a relatively straightforward agreement between the two disciplines’ scenarios (or at least, the majority position in each), the latter much more controversial. More specifically though, this article aspires to be useful further afield, in other regions where new synthesis work is not yet so well advanced and where the relevant linguistic principles and methods have all the more still to contribute. As an ideal case study for illustrating them to a general archaeological audience, we turn to a part of the world where our two disciplines do still have a great deal to learn from each other: the rich archaeological hunting-ground of the central Andes, in particular Peru and northern Bolivia. This region has been conspicuous by its absence from previous collections of work on correlations between archaeology and language, such as Blench & Spriggs (1997; 1998). Here, in the absence of any linguistic or historical records from pre-Columbian times, the time-scales of archaeology and linguistics overlap particularly closely, and at a relatively recent time-depth over which linguistic methods are at their most precise and reliable.

And yet to archaeologists of the region, the linguists’ tale of the Andes invariably comes as quite a shock, if not to say a revelation. For it parts company radically with the straightforward assumptions still current among most researchers of other disciplines, even those who specialize in the Andes, as to which languages are associated with which cultures. Nowhere is this more apparent than in the case of the biggest surviving indigenous language family of the New World, Quechua, and its popular association with the most extensive and last of the great Andean civilizations, that of the Incas.

1.1. Associating archaeological cultures with languages: first impressions

It is true that in this case geography alone immediately suggests what looks like an obvious and straightforward correlation, as can be seen by comparing the extents of the Inca Empire in Figure 1 and the current distribution of the Quechua language family in Figure 2. There is a strikingly close match in that the regions where Quechua survives today (and those where it has become extinct since the European conquest) almost all fall within the extent of the Incas’ realm. It is widely known, moreover, that the
Incas instituted Quechua as the language of administration for their empire. There is even a ‘vertical’, topographical correlation too: Quechua, at least in its current distribution, is very much a language of the highlands, rather than the narrow strip of desert along the coast or the Amazon rainforest. So it seems to fit that the Incas’ capital city Cuzco lies high in the Andes, at 3400 m.

On the strength of all these correlations, which also fall in line with certain local ethnic and political interests, the standard popular assumption has been to associate Quechua directly with the Incas. Many if not most archaeologists of the Andes also hold to this view, seeing the Cuzco region as Quechua’s point of origin, its ‘homeland’, and the Inca period or Late Horizon as the timescale for its expansion (i.e. starting from as late as 1476, according to one commonly-used dating for the first major Inca expansion).

1.2. Quechua and the Incas? Exploding the myths
And yet these and a number of other widely-held assumptions are, for specialists in Quechua linguistics, patently mistaken. A number of popular myths urgently need to be debunked, among them the following:

• Quechua is by no means a single language, but a broader family of mutually unintelligible tongues.
• Much of Quechua’s territorial expansion owes nothing whatever to the Incas, but predates them by centuries in some areas, a millennium or more in others.
• The Quechua family appears to have originated nowhere near Cuzco, but far to the northwest, somewhere in central Peru. Its expansion from there took place over an extended time-scale of the order of a couple of millennia, and was largely complete well before the Incas’ relatively very late entry on the Andean scene.
• Quechua did not even reach the Cuzco region until shortly before the Inca period, when its gradual southward spread finally brought it as far as Cuzco itself.
• Perhaps most disconcerting of all, it may well be that the Inca nobility themselves originally spoke not Quechua at all, but a (now extinct) regional form of the other main surviving language family of the Andes, Aymara, and perhaps at an even earlier stage Puquina.

A host of similar misconceptions surround Aymara too, following a popular assumption that associates it with the Tiwanaku culture as the homeland and motor for its expansion, but which again seems quite mistaken on the linguistic evidence.

The Andes would indeed seem to offer a few object lessons, then, in both how and how not to go about bringing together linguistic and archaeological scenarios. Certainly, the linguists of the region appear
to have signally failed to make even the fundamentals of their field at all widely known outside it, with the result that most specialists in other disciplines are proceeding unaware of even the most basic linguistic facts. Many archaeologists have continued working with assumptions about the Andean languages and their associations with archaeological cultures which, while apparently entirely straightforward, are all the more deceptive for it. Languages’ histories are often much more complex than meets the eye: superficial assumptions, especially on the basis of where languages happen to be spoken today, are more likely to lead one astray than in the right direction.

So deep is the gulf to be bridged that we need to start right back at the fundamentals. This article therefore begins by reviewing, for the benefit of archaeologists, the basic principles of how languages change and diverge over time. It then surveys the various methods and analyses that are available for making use of such knowledge in order to look back into the past, to date and locate language expansions. Finally, we show how those principles and methods of comparative-historical linguistics can be applied in practice, to explore how (and how far) one can match up linguistic and archaeological scenarios for prehistory. The case of Quechua and the Incas has been chosen as an ideal illustration of how linguistic techniques can enable us to reconstruct the approximate time-scales and geography of a language expansion, and with how much certainty at each stage.

The role of the Incas is actually just one among a whole range of issues in the prehistory of the Andes, where archaeology and linguistics still have a great deal to learn from each other (including the question of Aymara and Tiwanaku, as already hinted). The next issue of this journal will therefore include a follow-up article ‘Linguistics for archaeologists: a case-study in the Andes’ to offer further illustrations of how language and archaeology can work together, looking beyond the Incas to set out a more comprehensive survey of the wider scenarios for the linguistic prehistory of the Andes.

1.3. How reliable are linguistic conclusions?

The linguistic scenario for Quechua and the Incas as outlined above represents such a striking departure from established assumptions still current in other disciplines that it is incumbent on us here to justify how linguistics can come to such bold conclusions, and to assess how firmly established they really can be.

Indeed, much of the controversy in the new synthesis goes back to specialists in other disciplines enthusiastically buying into hypotheses about deep, distant relationships between language families put forward by only a fringe minority of linguists whose supposed linguistic classifications are comprehensively dismissed by the linguistic consensus. Among these are notably Greenberg’s (1987) proposed three-way classification of the languages of the Americas, the Nostratic hypothesis and purported reconstructions of a ‘Proto-World’ ancestor language (e.g. in Bengtson & Ruhlen 1994).

For the linguistic mainstream it is a source of considerable frustration to see specialists in other fields continue to look to such claims as if they were some form of established linguistic data. One sometimes hears a defence on the grounds that such classifications are ‘all we have to go on’, as if they were a kind of first broad sketch, necessarily ambitious but at least a valid basic outline, even if the details remain to be filled out. For the linguistic consensus they are nothing of the sort: even the broad sketch conclusions have no real foundation or meaning, in that the data claimed as linguistic ‘evidence’ pointedly lack the due methodological rigour required to establish language relatedness. On the contrary, most linguists take these data signal failing to demonstrate that the languages concerned have anything more in common than would be expected by pure chance (as demonstrated in Ringe 1992, for example). Such ‘classifications’ are not approximate but simply vacuous. For further details on these issues a useful starting point is McMahon & McMahon (2005, ch.1.3).

So for our purposes here, as well as setting out what linguists have been able to work out about linguistic prehistory, it is equally important to be clear on the limits to what is known, and to what is knowable at all, on the basis of language data. The case of Quechua illustrates these issues well. On the one hand, there is general agreement among Andean linguists on the broad-brush scenario outlined above — the orders of magnitude of the time-scales, and the basic directions of expansion — for at this level the linguistic data are fairly unequivocal and convincing. On the other hand, in the Andes as in many other parts of the world, many of the more precise questions about dates, homelands and associations with particular archaeological cultures are open to debate. A number of more detailed scenarios have been put forward, but as yet they remain in the realm of speculation. To fill in these gaps and help assess the relative plausibility of the different proposals, it goes without saying that a contribution from archaeologists is indispensable — once they are better apprised of how things stand on the linguistic side.
2. Principles of historical linguistics — and common pitfalls to avoid

How is it, then, that linguistic knowledge can be used to investigate population pre-history? There are a number of specific methods by which languages can be analysed to this end, which we review in detail in §3. Before that, however, it would be well to pre-empt a number of common and damaging misconceptions, by clarifying certain basic principles fundamental to an understanding of how languages change and diverge through time. These will be illustrated first with more familiar European languages.

2.1. Language change and language divergence

The mechanism by which languages diverge, particularly when they expand over new territories, is in fact quite straightforward. Language always changes over time, and these changes are converted into language divergence by virtue of one simple fact: the same original language changes in all regions in which it is spoken, but in different ways in each different region. As time goes by and more changes accumulate, each region ends up with its own individual combination of changes: some shared with one or more other regions, others shared with other regions, some specific to one region alone.

So for example, starting out from the same original pronunciation [septem]: for the word seven, all Latin-speaking regions shared in the loss of the final \(-m\), hence [septe]. Or rather, this change had probably already happened in the popular (‘Vulgar’) form of Latin which was the more direct ancestor of Romance, even before it began to diverge. Yet once this had been implanted into many different regions as the Empire expanded, the various other changes that continued to arise no longer necessarily all occurred in all regions. Taken together, these changes have turned the original [septem] into a range of different pronunciations in what we now think of as the modern Romance languages: [japte] in the modern descendant of Latin as it is now spoken around Bucharest, i.e. what we now call Romanian; [sette] in the pronunciation now used in Florence; [sjete] in Madrid; [seta] in Lisbon (but [set[i] in Rio de Janeiro); and just [set] in Barcelona and Paris. Figure 3 sets out schematically the patterns of individual changes by which this divergence arose. Note here in passing a golden rule of linguistics: the spoken, not the written form, is primary. Meaningful, accurate comparisons require absolutely that we ignore the distorting mirror of different national orthographic conventions, including their misleading archaising spellings, such as the now ‘silent’ \(p\) in written French sept. What counts are actual pronunciations and/or the underlying structure of the sound system, i.e. phonetic and/or phonemic transcriptions respectively.

Some aspects of language change are effectively random, but others are anything but (see Heggart 2006, 187). In particular, change typically proceeds in certain known directions. Provided that one has detailed knowledge and experience of the potentially very complex and interdependent processes involved, it is often possible to work out with fair confidence what an original ancestral form must have been, by comparing the various modern variants that it has given rise to in different related languages.
A comparison of [sæptə], [sete], [sjete], [setə], [setʃi] and [set], for instance, by the ‘rules’ of typical sound change, would normally lead linguists to assume original */septe/. In this case we can independently confirm the veracity of this reconstruction against the Classical Latin written form <septem>, and documentary evidence of the final /m/ already being dropped in actual speech. More precise analyses still allow us to refine our transcriptions of the probable exact vowel sounds to *[septem] (see particularly Allen 1978).

The same analytical procedures can be applied to the surviving varieties of any language family. This is how linguists are able to come to, and assert with such confidence, their knowledge of the form that a particular word must have had in the original Quechua and Aymara languages, for example, as well as the basics of their grammatical structures, sound systems, and so on. It is true that we cannot work out all the details with complete certainty: how much and how reliably we can reconstruct depends very much on the feature in question. Moreover, the fewer surviving languages we have as data, and the further back in time we have to go, the more uncertain our reconstructions gradually become. Still, as we saw above with */septe/, where we do have sufficient data the analyses clearly can work well to a high level of detail, and our confidence in such reconstructions is well-placed.

Most important for our purposes here is that linguists can thus work out also each of the various individual stages in how an ancestral form turned into the modern ones descended from it. This allows us to reconstruct the corresponding stages in the history of how the single ancestral language diverged into its various modern descendants. These stages form our key linguistic data for working backwards in time to look into the prehistory of the populations who spoke the language at each point in its divergence history. It will be clear by now why the branch of linguistics concerned necessarily goes by the double-barrelled appellation comparative-historical linguistics, for the two aspects are inseparably intertwined.

2.2. ‘Old’ languages and ‘homelands’?
Since so much of our interest is in putting dates on languages and finding their homelands, from the outset we must be clear on two questions. What exactly are we dating in the first place? And what do we really mean by homeland?

A moment’s reflection on the mass of differences between any modern Romance language and Latin attests to another golden rule of historical linguistics, a simple fact of life of natural languages: they all, without exception, change through time. None remains exactly the same as it was, even a century or two earlier. It follows that popular romantic descriptions of a particular language as ‘ancient’ or ‘old’, as are often applied to Quechua, Aymara, and many another native language, make no real sense, and are best avoided altogether. No natural language spoken today is any older or younger than any other in any meaningful sense (we exclude here a number of artificial languages). It is hardly as if languages have some date of birth. They are not generated out of nowhere; on the contrary, they are always continuations of what one might call language lineages. These are continuously passed down through the generations, though they are constantly changing through time, not least in the very act of transmission, learning and reanalysing.

Extrapolating back in time one ends ultimately in the debate — or more realistically, speculation — as to whether the ultimate origins of all language lineages go back to a single monogenesis of language among humans, or whether language developed independently more than once. If the latter, then in principle some language lineages might well be ‘older’ than others, in this sense. In practice, though, this takes us so far back in time that it is utterly impossible to identify which (see below). For practical purposes, then, it is indeed effectively pointless to talk even of any particular language lineage as ‘ancient’ or ‘old’ relative to any other.

A complication in the transmission of language is that in particular socio-cultural circumstances a population group can of course become bilingual in more than one language lineage, by learning a second language from another group. This is known as the horizontal transmission of a language, as opposed to the usual vertical transmission from parents to children. Bilinguals may then — again, depending on the socio-cultural circumstances — choose to pass on to their children a different native language to the one that they themselves first acquired from their own parents, with fundamental consequences to which we return below (§4.2). Nonetheless, such bilingualism is still necessarily between contemporary languages, again neither of them older than the other. And even in less usual cases of transmission, where multiple languages contribute to a ‘new’ one, such as a pidgin or a so-called mixed language, it too is always formed out of contemporary stages of the two or more language lineages it is based on.

The epithet ‘old’ is sometimes loosely and unhelpfully applied to languages in two other senses. Firstly, while natural languages never attest to a zero rate of change, even over time-spans as short as a cou-
ple of centuries there can be enormous differences and instabilities in how fast different languages change. So among a family of contemporaneous languages all derived from the same ancestor (the so-called proto-language at the root of that family), some of the lineages may have changed much more slowly or quickly than others. If one takes the words for water in the Romance family, for instance, Latin acqua [ak*wa] has changed relatively little to give modern Italian acqua and Spanish [aywa] agua, but enormously in French to become [o] eau — yet over the same time-scale in each case, of course. The result is that among related modern languages, some can have come to differ from their common ancestor much more or less than others: e.g. French vs Italian respectively, or within the Germanic family, English vs Icelandic. In this sense, such languages can be described as relatively innovative or conservative. But this still does not make those modern languages any ‘younger’ or ‘older’ than each other: they are still contemporary stages of their lineages.

Secondly, different language lineages can have been established in a particular location from different times. That is, one lineage may have continued in the same location from an earlier date than other lineages: in this sense Quechua might be said to be an ‘older’ language lineage in Peru than is Spanish. But this hardly means that Quechua has remained as it was: its lineage too has been changing all the time that it has been there, all the while that one of the lines of Indo-European was changing into Latin, and then a line of that into what we now call Spanish. Similarly, the Basque language lineage may have been installed in Europe for longer than others such as Indo-European and Finno-Ugric, but these lineages too duly continue back to their own homelands. Modern Basque is no older than any other modern European language — they are all contemporaries. Likewise in the Andes we must dispose of the fanciful ideas that the modern Quechua or Aymara of any region could possibly be somehow ‘ancient’ or identical to the speech of some period centuries or millennia in the past; to imagine so would be as nonsensical as to pretend that modern French or Spanish is identical to Latin or even to Proto-Indo-European. No part of the world today speaks ‘Inca Quechua’ any more, far less ‘the language of Tiwanaku’, just as none speaks Shakespearean English, nor the Spanish of Cervantes.

All of this has crucial consequences for what we actually mean when talking of dating languages and locating their homelands or origins. In fact, it makes no sense to talk of dating a language, for any language is continuously changing, and any particular stage of a language is defined in any case only by its point in time. Strange though it may seem, it is better to think of Latin and modern Italian (or any other Romance language), not as ‘two different languages’ but as two different stages in time of one uninterrupted language lineage. And looking back in time from the modern Romance languages Italian, French, Spanish, Roman and so on, we see their separate recent lineages all converge at the stage of Latin.

What we date then, are not languages but specific ‘events’ in the continuous history of a language lineage. Nor do we date a single language, but the successive stages in the genealogy of a language family. The events to which we typically seek to assign dates are splits in the lineage, though serious misconceptions surround this concept too. In reality there is never any particular point in a divergence process at which we can say that one language has suddenly split into two. It is hardly as if there is ever any sudden break in intelligibility between one generation of speakers and the next. Even if one could find some way to define a threshold level at which we say that divergence has reached such a point that two or more related lineages now constitute separate languages from each other, when they did not before that date, any such cut-off point is necessarily vague and arbitrary: divergence must have first begun long before then, and will normally continue after it. The only meaningful, reasonably definable cut-off point that one can date a split to is the very first point at which a divergence into two or more sub-lineages first begins.

In practice this also corresponds to what is usually the clearest unequivocally datable point in the extra-linguistic context too: some population movement and settlement which brings about a significant geographical separation of a population into two or more groups, who remain largely self-contained thereafter. (It is true that in very recent times mass communication has reached such a point that geographical separation is no longer so powerful a determining factor for language divergence. For all the periods we are interested in here, though, this modern exception certainly did not apply.) Divergence is best taken as first beginning, then, effectively as soon as a language is first implanted into each new territory.

The term homeland, too, is to be taken normally as that of an entire family, back to the location from which it first started expanding and diverging into its descendant languages. Moreover, this concept of a homeland is necessarily accompanied by a crucial proviso: as far back as we can reliably trace that language lineage, geographically and linguistically. Like any language lineage, Quechua and Aymara did not appear out of
nowhere. Ultimately they must go back to the speech of the first human settlers of the Americas, and in turn to wherever that came from, right back to the first genesis of speech that gave rise to those lineages (and the monogenesis/polygenesis question once more). Such questions lie far beyond our purposes here, and we only can repeat to go so far back in time leads us into the realm of pure speculation, for a simple and ineluctable reason. It is the very nature of language to change, and even at its slowest ‘background’ rate the changes accumulate so fast as to ensure that our linguistic reconstruction methods eventually begin to break down as we look further and further into the past. We cannot take our knowledge of any language lineage back with any reliability beyond about 10,000 years, and even this only in ideal conditions: that is, where our data include very old written records and cover many surviving languages right across a single and widely diversified family. With the languages of the Americas, unfortunately those conditions do not apply.

2.3. Languages and archaeological cultures
2.3.1. Factors shaping language divergence

Another inherent characteristic of language that has fundamental consequences, not least for our purposes here, is its nature as a socio-cultural phenomenon. This is a vast topic, but here we can narrow down the discussion to what interests us for our purposes: a cause-and-effect relationship between certain socio-cultural forces that may be detectable in the archaeological record, and corresponding patterns observable in how languages change and relate to each other. Indeed, if language — and specifically, patterns of language divergence — can be so informative of history in the first place, then this is only because those patterns are themselves moulded by what one might call the ‘forces of history’.

What makes a language a function of the population history of its speakers are certain social, cultural, political and demographic factors such as social prestige, degree of contact with or isolation from other populations, population size, density, growth and expansion, and so on. Our interest here is especially in how certain languages expanded dramatically into broad families, spreading over large territories, at the expense of others that became marginalized and extinct. Such expansions call for explanations in the form of an association with some particularly potent force(s) able to account for this linguistic success. These forces either give the people who speak a particular language the wherewithal to expand into new territories and outpopulate speakers of the language previously spoken there; or they confer on their language itself such prestige and utility for neighbouring populations as to encourage them to learn it too, thus spreading it by horizontal transmission instead. A related issue is whether a language spreads to a new area relatively suddenly by conquest and elite dominance, or more gradually by population pressure and a wave of advance. (For a discussion of models for the various mechanisms of language expansion and their archaeological correlates see Renfrew 1987, 120–44.)

Proximate factors behind such language expansions typically include relative superiority in technology, military prowess, state organization and/or trade success. In the Andes another factor often invoked is religion: Torero (2002, 88–9, 93) suggests the radiating influence of the famous Pachacámac shrine (among others) as a key motor for Quechua expansion. Certainly, expansions in cult practice are very visible in the archaeological record of the Andes; yet in known linguistic expansions religion rarely appears to play a determining role by itself, but is usually just concomitant with one of the more potent factors listed above.

For the largest language family expansions one can look further back to more general factors still, what Diamond (1998, 369) calls the ‘ultimate driving-forces’ in his table 18.2 summarizing language expansions in the Old World. Seven of the ten (and all the major ones) are attributed to food production, two (both relatively minor ones) to horse-based pastoralism, while the biggest, Indo-European, is left undecided between the two.

The argument for agriculture as a prime motor for the expansion of many of the world’s biggest language families is much debated (see Bellwood & Renfrew 2003 for a bibliography of primary sources on this debate). The basic logic in its favour is that a technology that can make more food available (per unit of area) enables its users simply to out-populate smaller, more dispersed groups. The linguistic correlate would be that the language of the less productive group is pushed out of the picture, as the more productive group expands. Their language might also enjoy greater prestige, encouraging horizontal transmission too. In the initial expansions of agriculture that Diamond has in mind, then, the linguistic ‘losers’ would have been the neighbouring hunter-gatherers.

In the Andes, the first of the two factors, horse-based pastoralism, is obviously not applicable. The shift to food production, meanwhile, seems too early to fit with the relatively limited linguistic diversity within the language families identifiable in the region, even on a generous view of the bracket
of time-depths compatible with such diversity (see §3.1.4, §4.4 and §4.5.1 below). Instead, the follow-up article offers a much broader proposal on the linguistic correlate of the particular pattern by which agriculture first developed in the Andes.

For the relatively late expansions of the Quechua and Aymara families, then, it seems that explanations are to be sought more at the level of proximate factors. That said, the Andes do offer one interesting instance on the suggestion of agriculture as a force behind language expansions, in the shape of Wari, one of the two main Middle Horizon cultures, as we shall see in the sequel article. We expand there too on how Wari also illustrates a crucial distinction to be drawn when seeking to associate archaeology and language. Strictly, it is best to think in terms of associating a language spread directly with some particularly potent force(s) able to account for it, rather than necessarily with some specific socio-political entity empowered by such forces. The same socio-cultural forces whose discernible impacts in the archaeological record can lead us to identify a ‘culture’ (while still perhaps debating its exact political nature, as with Wari) may be sufficient in themselves as motors for language expansion, without requiring any particular vision of that culture as a political or military entity.

This cause-and-effect relationship between extra-linguistic forces and language expansions matters also for attempts at linguistic dating. For it is evident that the relative strengths of different cultures and extra-linguistic factors are themselves anything but stable over time; necessarily then, their impacts on language change and divergence are also correspondingly unstable through time. It is this key principle that is behind the rejection by mainstream linguistics of any dating mechanism that assumes that rate of change is somehow magically constant: it cannot be, because of the very nature of language as susceptible to unstable external forces, particularly socio-cultural ones.9

Indeed, given that dating is so problematic, this relationship of cultural causes with linguistic effects is all the more important in that it offers an alternative approach, not a purely chronological one, to how to link stages in language divergence to particular archaeological cultures. In many cases a great deal of effort has gone into detailed reconstructions of what languages were like at each stage of their history and divergence, but rather less into setting them into the external contexts that determined why and how certain languages expanded or retreated at particular points in history. Any major language expansion cries out to be explained satisfactorily in terms of some factor in the extra-linguistic context sufficiently powerful to account for its success. In cases like the Andes, it is naturally archaeology that is in the best position to assess the impact of any such forces in the past. Here again, the two disciplines clearly need each other.

2.3.2. Language expansions with or without population movements

To put any given language expansion (or indeed retreat) into its real external context, it is important for our purposes also to identify which, or which mix, of the two possible types of language transmission account for it, i.e. vertical or horizontal (§2.2). One scenario is all too straightforward, and unfortunately unduly popular for it: there is a significant movement of people, who bring their language with them into a new region, displacing the native population and their language(s). Such was the mechanism by which English expanded into North America, for example.

The alternative is the reverse: there is no wholesale population movement, i.e. the native inhabitants of a region stay put, but change their language. First they become bilingual in the language of another population alien to that area, usually that of prestigious neighbours, traders and/or a ruling class of conquerors. Gradually, eventually, they switch to that other language entirely, abandoning their original native tongue (§2.2). Such was the dominant mechanism in the spread of English across Ireland, for instance. This second mechanism has important consequences for the search for a multidisciplinary consensus on population prehistory, since it can bring about a complete mismatch between a given population’s human genetic lineage and the ‘genealogical’ lineage of the language they speak. In the case of Ireland, originally Celtic-speakers are now Germanic-speakers — though of course it is an open question whether those two linguistic terms correspond to any true ethnic and human genetic unit in the first place.

There is a tendency for such cases of language replacements to be considered something of an exception to vertical transmission taken as the norm. A strong case can be made, however, for horizontal transmission having been much more frequent through history and across the globe than is generally realized. There is a clear parallel with what archaeologists themselves have come to stress increasingly in recent years: while a grand, sweeping migration may offer an apparently straightforward and more ‘romantic’ scenario in the Völkerwanderung tradition, this alone hardly means that it is necessarily what really happened.
2.4. Mechanisms of language divergence: family trees vs dialect continua

As with language transmission, so too for language divergence there are supposedly two basic mechanisms, again one of them undeservedly popular simply for being easier for us to visualize.

In the well-known family tree model, the history of a language family shows a clean break when the original language lineage splits into two (or more) branches, and from that point onwards each branch develops its own essentially separate changes, perhaps itself splitting into further sub-branches. This typically happens when the population that speaks the original language itself splits, one part of it moving away to become geographically isolated from each other, such as the split between the speakers of early Germanic who left for Britain, and those who remained on the European mainland.

A quite different scenario is when the original language expands widely across a continuous swathe of territory. When this becomes so extensive that communication right across it is no longer significant, particularly if political changes fragment its cohesion as a single socio-cultural entity, then divergence typically occurs according to the alternative wave model. That is, different changes begin in different regions, and spread gradually in ‘waves’ to some neighbouring regions, but often not right across the whole contiguous speech area. This ends up as a mosaic in which each small territory has its own local speech, very slightly different to that of the neighbouring areas, as the different waves (isoglosses) overlap differently from one region to the next. The further one travels across the continuum, the more differences accumulate, such that speakers from points still within it, but quite far apart, can no longer understand each other. That is, they speak what are effectively different languages, even though there is no single, fixed line anywhere within the continuum where we can draw a clear-cut language boundary. A useful analogy is with a spectrum or palette of colours: we are happy to agree that red is a different colour to blue, even though there is no particular point in the spectrum at which one suddenly changes into the other. But should we also distinguish purple? If so, what about mauve or violet? Are they colours in their own right, or just shades of another colour? To understand a dialect continuum, just replace the terms colour and shade with language and dialect, and much the same could be said of Spanish, Portuguese, and Galician, not to mention Asturian, Cantabrian, Mirandese, Fronterizo, and so on (see Penny 2001, §4.12 & 163–6).

The histories of many language families involve something of both models. It has been typical, however, to pay nothing more than lip-service to the wave model, before falling back on the idealizations of binary splits and the family tree. These are more tractable, to be sure, but only because they are more simplistic. It is high time to redress the balance in favour of the wave model, and the more complex dialect continua that it results in, for in reality they are at least as common as splits. Take the four main language families of Europe, namely Romance, Germanic, Slavic and Celtic: each one in fact diverged essentially into a dialect continuum. Only in the last few centuries did new, extremely powerful socio-political forces emerge to distort and obscure that initial picture radically. The early rise of nation-states, and then mass literacy, education, communications, and even military conscription, all conspired to effect standardization around the speech of one particular region per nation-state which became elevated to the status of national ‘language’ (and this for purely socio-political, not linguistic reasons). All other non-prestigious varieties across the family continuum, intermediate between those chosen as the various national standards, were left to decline and often become extinct. The linguistic map of Europe is fast losing all its historical subtleties as the shades distil into the fewer, sharper colours of the national standards.

The result in modern Europe is that most of us recognize only a few well-known ‘languages’ from regions within the original dialect continuum that are far apart and whose speech is therefore quite distinct from each other’s: Lisbon, Madrid, Barcelona, Paris, Florence, and so on. This may make the picture look nowadays more like one of neat, early splits between distinct ‘languages’, but it does not change what actually happened in the divergence history of each language family, which was quite the opposite: a gradual diversification into a dialect continuum. It is no surprise that it has actually proved impossible to find an agreed internal ‘tree’ structure for the continuum that is Romance, for example. There are counter-examples to most of the proposed branches in the other families too. Readers whose own experience is that of modern, literate Western languages should beware that their perspective is actually distinctly misleading and unrepresentative of the linguistic picture over most of the rest of the world, and much of the rest of history. Again our case-study will illustrate this well, which should hardly come as any surprise, for throughout the history of the indigenous populations of the Andes, few of the conditions for standardization existed at all, and certainly not mass literacy.

Alongside the deceptive impression given by the surviving national languages of Europe, the other
reason why even most linguists have been tempted to stick to the tree-model is that binary branches are simply easier for us to understand. Unfortunately though, more often than not the idealized branch model is more misleading than helpful. As noted in §1, the latest phase in the new synthesis has seen phylogenetic analysis programmes — developed primarily for studies of speciation in the biological sciences — increasingly applied to modelling language divergence. One of the main objections in applying these to language is precisely that their models typically force on the data a representation in terms of a family tree structure in which only discrete binary branches are permitted. Many of the programmes effectively search through all possible binary branching trees and pick the ‘best’ one, though what this really means in practice is only the tree least incompatible with the data. The reality of language divergence is that, at least as often as not, it produces relationships between languages that are not compatible with any tree-only structure at all. It turns out that the application to language data of the latest phylogenetic methods of the new network-type which do not insist on tree-only representations can immediately challenge traditional family tree analyses, as has recently happened for Quechua, as we shall see below in §4.5.2.

3. Types of linguistic evidence and analysis methods

Before surveying the various types of evidence and methods for analysing them that are available to linguistics to provide clues to population prehistory, we should preface them all with a general and very important caveat: just like the data and methods of any other science, those of linguistics need to be analysed and applied with care and very considerable specialist knowledge. Linguists know only too well from experience that their discipline is arguably a special case in quite how beguiling and misleading its data can be to the untrained eye and ear (see for example Heggarty 2006, 168).

3.1. Documents and toponyms

There are some types of language data that can take us back in time more or less directly. Historical documents in or on the languages of interest are of immense value where they are available from far back in the past. That linguists have been able to work out so much about Indo-European is largely thanks to the availability of Latin, Greek, Sanskrit, Avestan, and even Hittite documents. In the case of Quechua we are much less fortunate, since no written records exist from before the Spanish conquest (i.e. the 1530s), and we have only a limited range of documents written on and/or in Quechua or other Andean languages since then. Still, provided that we are confident of how to interpret them — and this too is by no means always straightforward — then these documents can help us push back our analysis by up to five centuries.

Toponymy too can be of great value, in that a good proportion of placenames remain fixed (if somewhat distorted) even when populations and/or the languages they speak change. Again, a warning is very much in order that the interpretation of many individual placenames can be particularly involved and problematic. Still, when large numbers of toponyms in a region all point clearly to a particular language, they can stand as solid evidence of its being spoken there in the past.

3.2. Proposed linguistic dating methods

Other than such ‘direct’ historical sources, often few and far between, our evidence is effectively restricted to whichever language varieties have survived to the present-day. Still, it is the essence of comparative-historical linguistics that the analysis of languages even in their modern state can offer us insights into their past. What the linguistic toolbox does not come with, however, is any reliable absolute dating techniques. Methods have been proposed, two in particular, but they remain extremely contested and quite unproven.

3.2.1. Linguistic palaeontology

One of these methods goes by the name of linguistic palaeontology. Some linguists have set great store by this, particularly Indo-Europeanists for whom it appears to be an outlet and an application for their painstaking reconstructions of Proto-Indo-European vocabulary. The basic assumption is that if it is possible to reconstruct for Proto-Indo-European a word that now means wheel or plough, for instance, then those things must already have existed (in order for them to have already had names) at the time of the original Proto-Indo-European language. That language can therefore have begun diverging only at a time after that invention (which can be dated independently by archaeologists), for otherwise we could not have found the same word root shared in its different daughter languages.

Despite the apparent power of this technique to get straight to the heart of the crucial dating issue, there are critical objections to it. Indeed, for all the enthusiasm of Indo-Europeanists, there is a large body of historical linguists who are very far from considering linguistic palaeontology to be an established, reliable
dating method that has won any recognition as an orthodox tool of their discipline.

The treatment here will be limited to what is of most concern to archaeologists: essentially an overview of how opinions divide among linguists. Also, since the basic principle can on the surface appear to many non-linguists to be quite logical, it is important here at least to air the key objection. This is that the technique relies entirely on a basic assumption that overlooks the many other eminently plausible explanations for the proto-vocabulary into which linguistic palaeontology reads so much. The argument centres on the natural processes of meaning change that apply specifically to the coining of names for new technologies or newly-encountered species, and which can produce results that can be indistinguishable from shared origin. We can with confidence reconstruct a word in the Germanic languages for mouse, even in its modern technological sense, though this hardly means it is safe to assume that the Proto-Germans had computers, of course. Similar arguments can be applied to wheel and plough, for while we can reconstruct the (approximate) sounds in these words, we can be much less sure of exactly what they might have meant originally. In any case the Proto-Indo-European roots presumably existed in the language long before they were conscripted into their new technological meanings, so they were always there to be potentially reconstructible for us later, whenever the actual invention happened.

Linguistic palaeontologists have sought to respond to the criticisms with a number of caveats that surround the basic assumption, and detailed arguments as to the particular words, meanings and languages involved. Certainly they now seek to dissociate themselves from the simplistic application of the basic assumption, which led to the old vision that the Proto-Indo-Europeans had a king — since a root *h₂rēg-s (Beekes 1995, 39) is unquestionably reconstructible, and its derivatives eventually came to have this meaning in several branches of Indo-European. On the other hand, even the method’s detractors generally recognize certain instances where it does instil some confidence: the very strong reconstructions for some of the basic numerals up to ten, for example, together do strongly support certain deductions that the Proto-Indo-European language already had a decimal counting system.

Crucially though, the key words for locating and dating Proto-Indo-European are names for species and for new technologies; that is, precisely the fields where the natural semantic processes for word-coining muddy the waters the most. So here even the detailed defences mounted by the linguistic palaeontologists fail to convince. The Proto-Indo-European word *kʷe-kʷ₁l-o-) (Beekes 1995, 171), touted by linguistic palaeontologists as having meant wheel, goes back suspiciously to a derivation, to which one cannot necessarily assign any meaning more specific than ‘(any)thing that turns’, or to gloss this reduplicative root rather more literally, a turn-turner. There are in any case plenty of uses for wheels or ‘things that turn’ which do not involve vehicles. As archaeologists of Mesoamerica can confirm, a culture can perfectly well know of and make its own ‘things that turn’ long before it actually gets round to putting them on vehicles: witness the curious wheels and axles on toys made by the pre-Columbian peoples of the region.

Most historical linguists are at least conscious of the dangers in pushing linguistic palaeontology very far; and many consider it methodologically quite unsound. Anttilla (1989, 373), for example, in his authoritative Historical and Comparative Linguistics, devotes but a single page (out of 462) to this ‘doctrine’ which ‘has rarely enjoyed particularly high repute’. Among other objections, readers might consult Pulgram (1958, 146–7), Sims-Williams (1998, 510), Krell (1998, 279) and obliquely Dixon (1997, 49), while my own arguments appear briefly in Heggarty (2006, 188–92).

None of the objections to linguistic palaeontology in any way impugns the mainstay of comparative-historical linguistics, the so-called comparative method. This is indeed reliable and precise in identifying and applying known processes of sound change to reconstruct the phonological forms of words; when applied with due methodological rigour, no historical linguists object to this. On the level of meaning, however, things are quite different. While some correspondence in meaning is an intrinsic requirement of the comparative method, what it does not apply are any precise meaning change analyses. Famous rules like Grimm’s Law, Verner’s Law or Grassmann’s Law are ‘laws’ of sound change, not meaning change (witness the alternative name for the first of these: the First Germanic Sound Shift). The comparative method is not, therefore, a method for necessarily determining the exact original meanings of the words it reconstructs. Only linguistic palaeontology takes the further and speculative step of assuming it is safe to assert exact meanings too.11

In our Andean case-study it transpires that there is precious little to say of linguistic palaeontology in any case, though this in itself at least stands as a useful perspective on its status among historical linguists. For it is telling that the consensus scenario among linguists of the Andean languages is not based in any way on any dating of Proto-Quechua or Proto-Aymara
divergence by linguistic palaeontology. Rather, specialists in the field have preferred to give the method a very wide berth. The only exceptions are Torero’s (1995) very limited attempt, referring to very early periods, and Proulx (2005); neither has gained any serious acceptance.

3.2.2. Glottochronology
The terms lexicostatistics and glottochronology are often confused, yet the distinction between them is a straightforward and very important one. Lexicostatistics is a simple, basic approach to putting numbers on language divergence. Provided that two or more languages go back to a common ancestor, one can measure how much overlap now remains between their modern vocabularies. Lexicostatistics takes a list of either 100 or 200 basic word meanings and tots up the proportion of those meanings which are still expressed by the same inherited root or cognate in any two descendant languages. For the meaning root, for instance, French and Spanish use the words pied and its cognate pie respectively, both derived from the same ancestral form (Latin pedem); but not for the meaning head, tête vs non-cognate cabeza, derived from different roots (Latin testam and caput respectively).

Some decades ago there was a burst of enthusiasm not just for such lexicostatistical measures but also for attempting to derive from them a supposed glottochronology of languages, on the basis of an assumption that core vocabulary is replaced at a constant rate over time. This proposal has long been comprehensively discredited. Not only is it unwarranted in principle, but it is also wildly inaccurate in practice, as was noted as early as Rea (1958) for Romance, for example (to which we return in §3.1.4 below). Indeed many linguists openly assert the reverse, e.g. Dixon’s (1997, 9, 11) explicit assumption that ‘The rate at which a language changes is not constant and is not predictable’. Proposed alterations to the glottochronological methodology (e.g. Embleton 1986) have failed to address the root problem that its basic assumption is patently mistaken, and have thus quite failed to rehabilitate it in the linguistic consensus. Again our case-study in the Andes illustrates the reputation that this method now enjoys. Thirty-five years ago, glottochronology was tried here too, and the consensus has long been that it signally failed: the proposed dates are simply ignored, as we shall see (§4.4).

3.2.3. A recent proposal
Gray & Atkinson (2003) put forward a completely new dating technique, also based on lexicostatistical data but processed very differently by a phylogenetic analysis, coupled with a radically new dating technique which does away entirely with the old glottochronological formula and which strives to accommodate a certain amount of variation in the rate of change. This new approach does appear to have certain methodological attractions, and Gray & Atkinson’s results for Indo-European do seem remarkably stable under a range of different assumptions. Nonetheless, most linguists remain critical and far from convinced. Questions remain, for example, as to how they make use of the cognate sets in the linguistic data, and the concept of a split date that they use in their calibrations (§2.2). In any case, this technique has not yet been applied to our case study of the Andean languages, though it might be earmarked as one possible avenue for future research.

3.2.4. So what is possible in linguistic dating?
So despite the earlier enthusiasm of proponents of linguistic palaeontology and glottochronology, unfortunately linguistics still has precious little to offer by way of absolute dating techniques. So much so that McMahon & McMahon (2006) can bluntly entitle a paper ‘Why linguists don’t do dates’, while Dixon (1997, 49) muses that ‘What has always filled me with wonder is the assurance with which many historical linguists assign a date to their reconstructed proto-language ... it does seem to be a house of cards’.

Before we throw up our hands in despair, however, there is one vital principle that we can and should rescue from all this. For even if it is true that language change can proceed at very different rates in different circumstances, it is still the case that the more time elapses, the more change accumulates. This leaves at least a valid default assumption, which can be used as the basis for what one might call a very approximate technique of dating by degree of divergence. That is, for a family of related languages, if we can successfully measure how different the modern ones have now become from each other, then those measures can stand at least as broad-brush indications of the time-scale over which those differences arose, i.e. how long those languages have been diverging since their common ancestor. Such time-scales have to be assessed, naturally, with reference to other cases of language divergence for which we do know the dates from independent historical sources. That is, we ‘calibrate’ this technique against the known time-span that it took for Romance to diverge from Latin, for instance.

It goes without saying that there are patent dangers and difficulties with such an approach. For a start, there are grounds to question whether the European
demographic and historical contexts that determined the divergence of the calibration cases like Romance are necessarily representative of those in other parts of the world like the Andes (see §4.4 below). And this approach relies on a further assumption that external factors of unusual strength have not intervened to distort the default time-to-divergence equation. In reality, given how variable are rates of change, and how sensitive to external, non-linguistic forces, the gross correlation between divergence and time elapsed is a very blunt one. The greatest precision we can aspire to is to narrow down the windows of compatible dates to within a few centuries either side of any best-estimate absolute date, and this only for the most recent time-depths in well-studied cases. The further back we go in time, the wider such ‘margins of error’ become.

One has to be strictly realistic, then, about what can be inferred from measures of language divergence as to the time-spans that might correspond. There cannot be any magical formula by which measures of divergence can be converted into remotely precise absolute dates, so there is no question of this approximate technique being able to give us any. On the other hand, being realistic also entails that we should not throw out the baby with the bathwater either. What our general principle can provide is at least a default framework for which orders of magnitude of time-spans are or are not compatible with which orders of magnitude of divergence. To aspire to exact numbers and absolute dates is to misunderstand the nature of the approach. Its role is necessarily limited to a balance of greater and lesser plausibilities, but just because it cannot give an exact number or date hardly means that the technique is useless. In practice there are plenty of cases, when we have competing scenarios to assess, in which even approximate orders of magnitude can make a valuable contribution by helping tip the balance of probabilities more or less heavily one way or the other, as we shall see for Quechua in §4.1, and in our follow-up article also for Aymara. We must remain aware of the dangers and imprecision inherent in this technique, and in some instances it can be a difficult balancing act to judge quite how much confidence we can have. Yet even most linguists who are vehemently critical of glottochronology do nonetheless recognize that in the most clear-cut cases we can indeed be reasonably confident of the broad orders of magnitude involved.

Quite understandably, this apparent contradiction often strikes specialists from other fields as suspiciously unscientific. How can the same linguists who insist that absolute dating is not possible then proceed to announce fairly confident judgements about more approximate time-spans? The point of the above discussion is to try to explain how this can be so, and for good practical reasons.

Frustration is often expressed, too, that linguists are reluctant to give an idea of the margin of error associated with variation in rates of change and divergence from one language to the next. The main reason is simple, and lies in the methodological difficulties in putting numbers on language in the first place. The data available to linguists are not at all of the same type as data in the natural sciences. There simply are no relevant language data that lend themselves to absolute mathematical expression and the calculation of margins of error at all. Given the very nature of its data, linguistics is unfortunately set to remain at the stage that archaeology was at before its aptly-named radiocarbon ‘revolution’. One might ask what then were archaeologists’ precise margins of error in attempting to date by pottery styles, for instance?

What we can attempt here is to give at least an indication of the order of magnitude of the variation observed in rates of language change and divergence over time. The glottochronological rate of change formula was first worked out on the basis of a number of historically datable stages in language lineages from various families. These were used to calibrate a fixed lexicostatistical list of core vocabulary. Yet when Rea (1958) applied this rate to the Romance languages, for example, it yielded extremely late dates wildly out of accord with known historical fact. To bring them remotely into line takes multiplication by a factor of at least 2.2 (Kroeber 1958), though a coherent approach to the concept of a split date actually calls for a range of multiples rather higher still, and of course different multiples from one case to the next. This suggests that even within Romance, rates of change varied by a factor of three or so.

There are at least some limits to such variation in rates, however. To suggest that one language might have changed ten times more rapidly than another, for instance, would be quite exceptional in the experience of most historical linguists. Clearly what we need is some hard quantification of the range of variation in rates of change, but that entails that we first solve the problems in putting reliable numbers on language in the first place. Until that is achieved, a reasonable position would be that other than in particularly extreme cases one would not expect a relatively fast-changing language to change more than three or four times faster than a relatively slow-changing one, and typically rather less than this. It is true that for Australian languages Dixon (1997, 48–9) claims that ‘4,000 years
Finally, note that the basic principle discussed in this section can equally be applied the other way around. That is, if we know independently at which time-depth a given ancestor language began to break up into a family of descendant languages, but find one of them to be unexpectedly divergent, then this suggests that some particularly powerful external factor intervened in its case — something we should look out for in the historical or archaeological record when we seek to match it against the linguistic scenario. An example from Europe is French, which began diverging from the rest of Romance no earlier than any other member of the family, but which on many levels has gone through unusually far-reaching and rapid change. The explanation most often invoked is that the impact of the Franks must have been particularly powerful in northern France relative to cases of Germanic contact with Romance in other regions of the collapsing Roman Empire (the Lombards in northern Italy, the Visigoths in Spain).

### 3.3. Patterns in language relationships: tracing homelands and population movements

What linguists can also do, and with much more confidence and precision than they can quantify or date, is work out the detailed patterns of relationships between language varieties, using a number of linguistic analysis techniques that are proven and well-established (recall the */septe/* example in §2.1 above). Where enough data survive from a large enough number of language varieties that are sufficiently closely related to each other, linguists can establish in considerable detail how each one relates to all the others within a genealogical hierarchy and/or dialect continuum vision of that family. The linguistic data are unequivocal, for example, in establishing that all the Romance language ‘sisters’ are more closely related to each other than to their ‘cousins’ within any of the other sub-families of Indo-European such as Germanic, Slavic, Celtic or Indo-Iranian. Even within a single subgroup, one can establish that Portuguese and Spanish are more closely related to each other than either is to Romanian, for example, just as are German and Dutch as opposed to Icelandic and Faroese. One can thus also establish a corresponding hierarchy of chronological stages and/or degrees of separation between populations through the history of divergence of a language family. We can work out, for instance, that one of the subgroups that emerged at a relatively early stage from Proto-Indo-European was Italic (as opposed to Germanic, Celtic, etc.), and that then from within that emerged Latin (as opposed to other now extinct lineages within Italic, such as Oscan.
and Umbrian), which in turn then diverged into the Romance sub-family.

In clear-cut cases such as these, the conclusions are among the most secure that linguistics can offer, and serve as extremely robust data which can be used in various ways helpful for our purposes.

• The genealogical relationships between the sub-groups within a family can be analysed alongside their current geographical distribution. Working back in time one tries to uncover the scenario that most plausibly accounts for how the various subgroups came into their particular current constellation: in which order, and from which areas, they diverged and expanded to reach their modern locations, tracing them all the way back to the most likely original homeland for the entire family.

• On a finer scale, localized geographical discontinuities in the linguistic record — where a given area’s language is closer to that of far distant areas than to that of nearby ones — can suggest individual long-distance population movements. Such discontinuities turn out to be of special interest in the central Andes, in the light of certain known peculiarities in the population history of the region, the so-called resettlements.

Once we have identified genealogical and/or geographical subgroups within a larger family, one further potentially useful type of data is to be had by comparing the amount of internal linguistic diversity found within each subgroup. Again, there are various ways in which these data can be informative.

• Recalling the approach of approximate dating by degree of divergence within a subgroup, diversity measures can stand as a guide at least to the broad time-depth window that most likely corresponds, as we have seen (§3.1.4). Alternatively, stepping back from an attempt to establish absolute dates, by comparing the degrees of diversity of a range of different subgroups one can outline a default relative chronology of which subgroup most likely began diverging before or after which others.

• Diversity ratings can also usefully be set into a geographical perspective: the range of territory occupied by a given subgroup. That is, one can compare language diversity per unit of geographical area in each of the various regions where a language (or family) is spoken. Such data have often been looked to as pointers to the likely starting point for a language expansion. Within the English-speaking world, for example, the diversity in accents across the USA, particularly the more recently settled West, is considerably less than that found north–south along the East Coast, which was settled earlier, and much less than that across the much smaller area of the British Isles, even after the decline in traditional dialects there since mass literacy and standardization set in. Similarly, the diversity in the Hispanic languages now spoken from the Rio Grande to Tierra del Fuego is far less than that within Iberia itself. The rule of thumb, then, is that the starting point of a language expansion is normally expected to be in a region which shows relatively high, not low diversity per unit of area. We describe this as only a ‘rule of thumb’, however, for it is clear that this principle does not work automatically in all cases, and does not enjoy universal approval among linguists, though specialists do tend to consider it useful in the case of Quechua at least, for reasons we shall see in §4.5.1.

Both of these approaches are limited by the general obstacles to quantifying linguistic diversity, as already discussed in §3.1.4. And even for the few, tentative measures that we do have of the difference between any two language varieties, there is no established method for turning them into ratings of diversity for large sets of varieties, nor for sampling across a dialect continuum. Hitherto linguists have tended to fall back on rather impressionistic, unquantified judgements of such diversity, though in §4.5.1 we shall look at some first attempts to come up with actual numerical values for Quechua.

3.4. Evidence of external socio-cultural contexts

On a quite different level, detailed linguistic analysis can also uncover various types of particular ‘telltale features’ in a language that are typically associated with specific external contexts and extra-linguistic forces: cultural, socio-political, demographic, and so on. Since it is these forces that shape language divergence, we can work back from this linguistic record through the history of a given language family to infer the likely nature of the external factors that must have operated on it at the various stages in its development in order to produce those linguistic results.

Languages can show traces of having borrowed or copied certain words, sounds or structures, and in many cases it can be clear which language borrowed from which other, and at which particular points in their histories — i.e. before or after particular stages of change and divergence. Several of the traits which so set apart French from the rest of the Romance languages, for example, have been attributed to contact influence from the Germanic language originally spoken by the Franks. The technique of forming questions by inverting the order of subject and verb, for example, is not native to Romance and suspiciously
identical with Germanic: compare *Tu peux venir.* ↔ *Peux-tu venir?* with *You can come.* ↔ *Can you come?* (see Posner 1996, 248–9). In northern France it was the local Romance tongue which eventually won out, and all that remains of Frankish there are these faint superstrate imprints on French. English too is marked by superstrate influences (particularly loanwords) from Norman French, which help illustrate how language data can often help us date particular linguistic events. The fact that [s] is preserved in *castle* and *étrange,* for instance, shows that at the time English borrowed these words, Norman French must still have had the original Latin [s] sound, now lost in French *château* and *étrange* (remaining only as a written accent).

A language may also display features typical of having passed through a phase of horizontal transmission, an ‘imperfect’ acquisition by native-speakers of other languages. Among the characteristics distinctive of Irish English, for example, are that the sound spelt *th* in words like *three* and *tooth* can be pronounced in a way which to other English speakers can sound more like *tree* and *toot,* i.e. *[tʰuː]* and *[tʰu*t]*. This is most likely a throwback ultimately to certain pronunciation norms in Gaelic, an example of a substrate feature: a characteristic of an original underlying native language that its speakers carry over into the new language they learn. In this case we know independently that the population of Ireland who eventually learnt English had indeed originally spoken a Celtic tongue, but in cases where we might otherwise not be so sure such substrate features can offer further linguistic evidence of which languages were spoken where in the past. As with folk etymologies, however, great care has to be taken not to make superficial assumptions, jumping to conclusions that certain features are necessarily attributable to substrate influences. On closer inspection, the linguistic details are often far less straightforward than they appear at first sight, and many substrate claims turn out to be far from reliable.13

Finally, among the traits that can indicate horizontal transmission are some which do not necessarily go back to a single identifiable superstrate or substrate language. In circumstances where a common outside language is learnt as a means of basic practical communication between speakers of various different native tongues, particularly for the purpose of trade between them, each group typically speaks the common external language rather imperfectly, and differently depending on their own native speech. (Consider how foreign learners pronounce English in markedly different ways, determined by the different pronunciation norms in their various native languages: a French, Russian, Italian or German ‘accent’, and so on.) In a lingua franca or koine scenario, such differences tend to resolve quickly towards the least complex common denominator among the speakers, particularly by means of certain grammatical reanalyses, usually simplifications. So where a language shows a set of changes of these specific types at some point in its history, such data can suggest an external context in which it was taken up in the role of a lingua franca.

Linguistic indicators of all these sorts can point to a language having expanded into a given region principally by horizontal transmission. A matching archaeological scenario, then, need not envisage any significant population movement, but should instead offer evidence of extra-linguistic forces that would have been propitious for horizontal transmission, such as trading links or influences of other sorts able to confer particular prestige on an external culture, and by association on its language too. Similar linguistic traces can also suggest approximate relative time-scales not for language divergence events, but for periods of contact between the speakers of two or more language lineages. Again, indicators of both types prove helpful in our case-study on Quechua.

To recap, the last two sections have reviewed a range of techniques that do not involve any precise dating element, but which can offer alternative ways to match a linguistic scenario up against an archaeological one. That is, we can also look for how good a fit there is between such scenarios on the levels of their geographical and/or socio-cultural contexts, and by no means just in chronology.

4. Dating and locating homelands: the case of Quechua and the Incas

It is finally time to turn to our case-study, to see what all these linguistic principles, methods and sources of evidence can tell us in a specific real-world example: the ideal illustrative case of the main surviving indigenous language family of the Americas, Quechua.

4.1. Approximate dating by degrees of linguistic diversity

4.1.1. A recent expansion: the case of Quechua in Bolivia

We start with a fairly clear-cut case: the relationship between the Quechua of Cuzco — the Inca capital — and that of Bolivia. All Quechua-speaking regions from Cuzco southwards, including all parts of Bolivia where it is spoken, are generally described as belonging to the same Cuzco-Collao or Cuzco-Bolivian dialect of Quechua. Nonetheless, a number of local
differences have arisen across this large territory, and a few of these are significant enough to mean that speakers from different regions within it can no longer understand each other perfectly. Notably some areas of Bolivia show particularly heavy influence from Spanish, and a decline in the use of certain significant suffixes retained in Cuzco (e.g. the native noun plural ‘kuna and assertive ‘mi). Nonetheless, most differences are at a relatively minor accent level, and the regional varieties within Cuzco-Bolivian do remain all mutually intelligible to a very high degree.

For an external perspective, the degree of diversity across all Quechua from Cuzco southwards is broadly comparable with (if anything, slightly greater than) the diversity that has developed within Spanish over the last five centuries since it first began expanding across the New World, i.e. the differences between the Spanish of Iberia and that of Puerto Rico, Mexico City, Lima, Cuzco and Buenos Aires. The limited diversity within Cuzco-Bolivian Quechua is thus eminently compatible with a time-depth of the same order as Spanish in the New World, or perhaps just a little older than that. In the Andean context, this means the Late Horizon, the time of the Inca Empire. Certainly, such chronological indications are by no means precise enough to exclude a time-frame somewhat later or earlier, but our scenario is supported on other linguistic grounds too. Within the family as a whole, the Quechua of Bolivia bears a strikingly close relationship to that of Cuzco, so its ‘genealogy’ too is very much compatible with seeing its origins in populations who moved from the Inca heartland of Cuzco into Bolivia, as the Empire expanded there.

Certainly the Incas garrisoned the area around Cochabamba (e.g. at Inkallaqta), presumably with loyal troops from the wider Cuzco region; though in fact it appears that the Incas were far from the only agents responsible for the spread of Quechua across Bolivia, for it continued gaining ground there in later times too, well into the colonial period. For over three centuries (up until the 1780 rebellion under Túpac Amaru II) the Spaniards themselves all but promoted Quechua, as a vehicle of communication with the native population, not least for the purposes of evangelization. More important still, their labour drafts for the silver mines in Potosí, which transported into southern Bolivia workforces drawn from a vast catchment area that included a broad swathe of southern Peru: the bishopric of Cuzco.

4.1.2. An early expansion: the case of Quechua in north-central Peru

This observation that the Quechua of Bolivia is broadly very similar to and generally mutually intelligible with that of Cuzco has been enough to create a widespread impression of the Cuzco variety as somehow the one established reference form of ‘the Quechua language’. With this has gone an assumption that as well as the Bolivian variety, the Quechua spoken in many other regions of the Andes must likewise be just a derivative of Cuzco Quechua, and still essentially the same language. Despite the wide currency of this assumption, even among specialists in Andean history and archaeology, this is in fact anything but the case.

A few examples will suffice. The basic word where? is [majpi] in both Cuzco and Bolivia, but the
corresponding term in central Quechua is so different as to be unrecognizable: [meítʃo]. For the numeral one, regional pronunciations vary from [ʃux] to [huk] to [øy], and so on across a range of further variants; while the pronoun we is [noqañiʃiʃ] in Cuzco, but in other regions [nɔwañʃiʃ], [nukanʃiʃ], [nocantsexʃ] and even [jaʔaʃiʃ], respectively in Huancavelica, the Ecuadorian Amazon, Chavin (the Ancash region) and Huancayo — see the more close-up map of Quechua varieties in Peru in Figure 4.

It is in particular these last two regions of central highland Peru that are home to varieties of Quechua that show the deepest gulf with Cuzco Quechua on multiple levels, not just in sound structure but also in word-structure, grammar and vocabulary. This realization proves crucial to an understanding of Quechua and its origins. For as even this briefest experience of them suggests, the Quechuas of Ancash and Huancayo are far from mutually intelligible with that of Cuzco, and not even easily with each other (see Torero 2002, 58, 85). That is, even with just these three regions — Cuzco, Ancash and Huancayo — we have what are effectively three quite different languages.

In reality then, the term Quechua refers to a linguistic unit that is parallel not to Spanish but more to the Romance languages as a whole. That is, Quechua is by no means a single monolithic language, but a fairly broad family of several different languages. Indeed, the different regional forms of Quechua have variously been described as more or less equally different to each other as are Spanish, Portuguese, Italian and other Romance languages (e.g. by Parker 1976, 24). Perhaps the closest parallel in terminology is with the popular use of the overarching term Chinese: thus Ancash vs Cuzco Quechua would be akin to Cantonese vs Mandarin. Indeed as with Chinese, the question of exactly how many different Quechua languages one might identify is a notoriously difficult one. Much of the vast Quechua-speaking area is a broad dialect continuum without clear boundaries (§2.4), but whose poles are nonetheless very different. As an indication, a government survey in 1976 chose to produce separate grammars and dictionaries for six different Quechuas in Peru alone, and even this left out many other regions where the local Quechua was already moribund.

What of all this for dating, then? The degree of divergence within the Quechua family is not greatly different from that of Romance, particularly if one excludes the case of French with its atypically rapid change. Linguists generally agree that this divergence is of a far greater order than is compatible with a time-span so shallow as that of the Incas. Especially by comparison with Bolivian Quechua, it is all the more obvious that the Quechua of north/central Peru must have been diverging from that of Cuzco for many centuries more than that. Some indicative figures might help: on the measures in Heggarty (2005, 13), Cuzco rates at between 86 and 90 per cent similar to most Bolivian Quechua, but only in the low 70s per cent to the Quechua of central and northern Peru, and in one case as low as 64 per cent. The rough default assumption would be that we should be thinking not in terms of the half-millennium or so since the Inca expansion into Bolivia, but a time-scale of the order of some low multiple of that; recall that Romance has been diverging for a little over two millennia.

To overturn this assumption would call for some external factor that could have accelerated language change everywhere but in Cuzco and Bolivia in some spectacular and exceptional fashion. There seems little evidence of any factors of sufficient force, nor is there any need to invoke any when a much more plausible picture presents itself in any case. The divergence of the Quechua language family goes back far earlier than the Late Horizon, a scenario strongly supported by the array of all other forms of linguistic evidence available.

Firstly, we know that Inca rule was relatively short-lived, lasting well under a century in almost all areas of the Empire. Such a time-span is unlikely to have been remotely sufficient for an imposed administrative language successfully to displace the local speech. On the contrary, it seems rather that Quechua was all the more viable as an easy lingua franca precisely because so many of the areas that the Incas conquered already spoke their own form of it. There would still have remained enough common ground between the different regional forms that speakers from any of these regions would find Cuzco and/or the main northern variety of Quechua still similar enough to their own speech that they could more or less understand either, or at least pick them up relatively easily. A modern parallel would be how even today speakers of Italian, Spanish and Portuguese can ‘still’ communicate with each other, albeit only at a rather basic level, each in their own native form of ‘modern Latin’.

Secondly, the only way in which a language can establish itself over so short a time-span as the Incas ruled for is by wholesale settlement of relatively unpopulated areas; it seems clear from history and archaeology that this did not apply to most territories north of Cuzco. To the south, though, some such settlement might well account for the otherwise unusual discontinuous geographical distribution of what is linguistically a relatively homogeneous dialect by the standards of the Quechua family: Cuzco-Bolivian. This seems to have leapfrogged from the Cuzco/
Puno region straight into southern Bolivia, beyond the extensive and long-settled regions of Titicaca, Tiwanaku and northern Bolivia where Aymara was spoken (alongside other indigenous languages now either endangered or recently extinct: Uru-Chipaya and Puquina). If archaeologists and/or geneticists can estimate population densities for all these regions before and after Inca conquest, this might help challenge or support this linguistic hypothesis.

Thirdly, a wide range of documents from the early years of Spanish rule provide evidence of the most reliable kind that even at the time of the conquest, very significant differences already existed between the Quechua spoken in different regions. One of our earliest and most reliable sources, the Quechua grammar by Domingo de Santo Tomás ([1560] 1995, 18), begins by explicitly warning readers that ‘The Indians of one province pronounce many terms differently from those of another’15, before listing a number of such differences (see the interpretation in Cerrón-Palomino [1987] 2003, 84–5). Many texts mention specifically a ‘Chinchaysuyu’ form of Quechua, named for the vast province which covered almost all of the Empire north of Cuzco, though in practice this seems to have been a catch-all term applied to any of a number of heterogeneous regional varieties of Quechua, on the sole criterion of being clearly distinct from the Quechua of Cuzco. All this evidence sits very ill with the suggestion that these forms of Quechua, so diverse already by the time that the Spaniards arrived, could have been a mere recent imposition of the Incas (see also Cerrón-Palomino [1987] 2003, 326).

The rough match between the extent of Quechua-speaking areas today and that of the Inca empire, then, is largely coincidental. An analogy may help put in perspective the scale of the misapprehension in imagining that the spread of Quechua derives from the Incas: it would be equivalent to claiming that the Romance languages of Europe derive, say, from the Empire of Charles V in the sixteenth century, rather than from the Latin of the Roman Empire. Over much of their territory, the Incas conquered peoples already speaking Quechua, albeit a very different form from their own, which they could no longer well understand (Cerrón-Palomino [1987] 2003, 325).

4.1.3. An intermediate case: the case of Quechua in Ecuador

In the clear-cut cases of Bolivia and Central Peru the technique of broad-brush dating by degree of linguistic divergence is clearly of value, then. We now move on to a more borderline case, however, which reveals the serious limits to its precision. The illustration is provided by Ecuador Quechua (more commonly known there as Quichua), whose degree of similarity with Cuzco Quechua is intermediate between the two preceding cases: 78 to 79 per cent in the measures in Heggarty (2005, 13). Moreover, the Quechua of Ecuador is ‘genealogically’ more closely related to that of Cuzco than to that of central Peru. This immediately poses a geographical and archaeological conundrum: the two extremes of Quechua-speaking territory, Cuzco/Bolivia in the south and Ecuador in the north, speak varieties more closely related to each other than either is to the Quechua of central Peru that lies between them...

On the default assumptions at least, the lineage of Ecuador Quechua would have diverged from that of Cuzco clearly before Bolivian Quechua did, i.e. a few centuries before the Late Horizon. But even at such a shallow time depth (about 800 or 900 years bp), given the range of variation in rates of divergence over time, we cannot exclude a margin of error of the order of a few centuries either side, and the later end of this bracket would therefore include the Late Horizon, and the Incas.

On this evidence alone, linguists would be uncertain about when and how Quechua reached Ecuador. It might have been brought with the Incas, though in this case we would still have to explain why it has nonetheless come to differ from Cuzco Quechua markedly more than has Bolivian Quechua, over the same time-frame. Which factor might have intervened in Ecuador to accelerate the rate of divergence sufficiently? Rather more plausibly, Quechua would have reached Ecuador some centuries before the Incas, though now we are left having to explain which cultural factor (and possibly also a population movement) took it there. As this example makes clear, in the absence of clear-cut, precise linguistic dating, in intermediate cases we are left with just a balance of greater and lesser plausibilities. To progress, we need to look to our alternative level, that of associations with the demographic and social factors that shape language divergence, and to the other available sources of evidence for them.

In this case too, old Spanish colonial documents turn out to be of particular help, not least their numerous references to the speech of ‘Chinchaysuyu’ (which includes modern Ecuador), and specifically to this as a form of Quechua, but one already considerably different to the speech of Cuzco. One indication of quite how different they were even by the time of the Spanish conquest is to be found in the Historia General del Perú (book I, cap. xxv) by the Spanish/Inca chronicler Garcilaso de la Vega, commented on extensively by Torero (2002, 100). The last Inca Emperor, Atahualpa,
chooses to speak to the Spaniards’ interpreter Felipillo ‘in the idiom of Chinchaysuyu, which the interpreter understood better, since it was more common in those provinces [Cajamarca?] than that of Cuzco’. As a ‘prince’ of the Incas but one who had spent most of his youth and adult life in modern Ecuador, Atahualpa would have known both the Cuzco and Chinchaysuyu forms of Quechua. Felipillo, a trader from some northern region, appears not to have known the Cuzco variety well.

It is highly implausible that the northern and Cuzco varieties of Quechua could have diverged so much as to impinge on mutual intelligibility for Felipillo over only the six decades or so of Inca presence in these northern areas. Moreover, the details of the genealogical affiliation of Ecuador Quechua set it firmly within a quite separate sub-group to Cuzco-Bolivian Quechua. Important for our search for the origins of Ecuador Quechua are some important features that it shares with the Quechua once spoken in the central-southern coast of Peru, as documented in the early grammar by Santo Tomás ([1560] 1995), and which are markedly distinct from Cuzco Quechua (see Cerrón-Palomino 2003, 245).

On a range of linguistic evidence, then, much the more plausible scenario is that some other form(s) of Quechua had already reached Ecuador, and indeed northern regions of Peru such as Cajamarca, at least several centuries before the Cuzco variety ever appeared there with the Incas. What this still leaves us needing to explain, however, is how — by what socio-cultural force — it got there.

4.2. Language expansions without population movements

Ecuadoran Quechua also serves as a good illustration of how other types of linguistic analysis can offer clues to the operation of such external forces. Muysken (1977; 2000) identifies a number of features in Ecuadoran Quechua that seem typical of a lingua franca (§3.3), i.e. a language learnt by a number of existing populations, each of which originally spoke a different native tongue, in order to facilitate communication amongst themselves (see also Cerrón-Palomino [1987] 2003, 343–4.) Ethnological clues suggest such a history too: strong ethnic identities and tribal names have carried on even among Ecuadoran populations who now all speak Quechua (for a map of pre-Quechua languages in the Ecuadoran highlands see Adelaar & Muysken 2004, 166). Different substrate languages in different regions would also help account for Ecuadoran Quechua’s fairly high degree of fragmentation into a dialect continuum.

So in large part at least, Quechua seems to have spread in Ecuador by the existing local populations just switching language to Quechua, a process which appears to have continued over many centuries, into Inca and colonial times. It follows that in the archaeological record we need not necessarily look for a significant population movement to have brought Quechua-speakers from elsewhere into Ecuador. Rather, we should look for evidence of some factor that could have motivated the local populations to switch to an outside language originally from further south. Torero (2002, §3.3.4, 91–105) posits that Quechua’s attraction was precisely that of a lingua franca for the purpose of trade, by which it spread northwards from central coastal Peru into northern Peru and Ecuador several centuries before Inca times. Many linguists take this as the most plausible proposal so far, though we have here a clear instance of how the new synthesis disciplines can inform, corroborate or challenge each other’s scenarios. How valid it is to posit the expansion of northern Quechua as a trading language is necessarily to be assessed on the strength of archaeological evidence for the trading relations into which Torero reads so much. Were they really intense enough for Quechua to have held out such a strong attraction as to be adopted in preference to native ethnic languages? And which precise regions did such trade link, by which routes? Torero (2002, 96) follows Rostworowski (1970; 1975, 340–42) in envisaging maritime trade; but Hocquenghem (1993) counters that it must have been primarily overland. We return to these issues in more detail in our follow-up article.

4.3. Discontinuities in dialect geography: minor population movements?

In other Quechua-speaking regions there are other types of linguistic evidence that do indeed suggest particular population movements: discontinuities in dialect geography (§3.2). In certain localized pockets, the local Quechua shows unusual features more typical of the speech of other distant regions than of nearer ones. This suggests individual population movements that uprooted a small population and removed it far from its original homeland. Such movements are a known feature of the Andean context, not least of the Late Horizon. The Incas’ policy of planned mitma resettlements saw populations moved wholesale around their Empire, particularly to ensure a presence of loyal subjects in newly-conquered areas. The suddenness of the Inca collapse then left other populations stranded far from home too, including both their mit’a labour drafts for major public works, and armies on campaign in the Inca war of succession at the time.
In a number of areas where such discontinuities have been detected in the local Quechua, resettlements or armies have duly been invoked as the explanation. The southernmost outpost of Quechua, for instance, is the Santiago del Estero province in northwestern Argentina (where again it is known locally as Quichua). This lies far beyond and cut off from the nearest other Quechua regions, which speak the Cuzco-Bolivian variety (see Fig. 2). It was traditionally assumed that the Quechua here was simply a continuation of that of Cuzco and Bolivia, and settled from those regions. In more recent research, however, Adelaar (1995) and de Granda (1999) have highlighted a number of features atypical of Cuzco and reminiscent instead of the Quechua of both northern and central Peru. The suggestion — in Adelaar (1995, 46–7), for example — is for a significant input of Quechua speakers from these regions too, whether at the time of the Inca expansion itself, or even as native troops accompanying the first Spanish expeditions into Santiago del Estero. It should be cautioned that Adelaar and de Granda’s linguistic arguments are not universally accepted, but the debate nonetheless illustrates how linguistic data can contribute to our view of historical geography and population movements.

Conversely, there are also discontinuities in the shape of features that do hark back to Cuzco Quechua, and yet are found in distant varieties not otherwise closely related to it. A small number of these have been identified in some Ecuadorian Quechua, such as the occasional aspirated stop consonants heard in the central and southern highlands of Ecuador, but absent from all other areas north of the Cuzco region. Some authors have claimed such correspondences as evidence that it was after all the Incas who took Quechua to Ecuador. The features that they identify correspond only sporadically to the Cuzco equivalents, however, and are of little significance in relation to the mass of other linguistic data that unmistakably attest to a quite different lineage for Ecuador Quechua. Most linguists consider other explanations far more plausible: the correspondences are either common survivals from early Quechua, lost in all other regions; or more likely they are superstrate features, late local ‘imitations’ of the Cuzco form of Quechua, encouraged by the prestige and influence that it would have enjoyed during the brief period of Empire.

It is appropriate here also to mention the unusual position of the varieties of Quechua spoken in northern Peru, found only in a number of small, geographically isolated pockets (see Fig. 4), where, moreover, their survival is now in doubt. The Quechua of some parts of northern Peru, particularly in the Amazon lowlands, was clearly brought there from Ecuador in relatively recent immigrations; our interest here is in the highland areas further west where the language has been long established, yet within larger regions that do not seem ever to have been predominantly Quechua-speaking (see Adelaar & Muysken 2004, 172). This applies notably to the pockets around Cajamarca, and to those of Inkawasi and Cañaris (in the mountains of the Ferreñafe district, Lambayeque department), as well as perhaps to those further east in the Chachapoyas region.

In some cases local folklore claims an origin as a resettlement or an army, an explanation occasionally entertained by linguists too (Quesada 1976, 28; Taylor 1984b, 18, 42). Whatever demographic explanation is proposed, though, it will have to reflect a number of linguistic facts, above all that these North Peruvian Quechus too are by no means directly descended from Cuzco Quechua. Their linguistic characteristics do bring them together into a loose overall regional grouping, but quite where this fits into the family as a whole is unclear. For beyond some individual characteristics of their own, the North Peruvian Quechus show a mixed bag of features held in common with one or other of the main dialect groups: some with Southern Quechua (Cuzco-Bolivian), but many others with Ecuadorian, others still with Central Quechua, and some even with the Quechua once spoken on the south-central coast of Peru (see Landerman 1991). Most plausibly, then, the source of these North Peruvian varieties too goes back to well before the Incas, and to a region where some intermediate variety of Quechua was once spoken — perhaps the central coast of Peru, where it has long since died out (for more details, see Cerrón-Palomino [1987] 2003, 237–46).

4.4. A time-frame for the Quechua family

The Quechus of different regions are so radically different to each other, then, that the time-frame for the first significant divergence in Quechua must go back much, much deeper than the Incas. But how far back, precisely?

In the heyday of glottochronology, Torero ([1970] 1972) sought to apply it to Quechua and came up with initial calculations of the order of eleven centuries for the family as a whole. Above all else, though, this application to the Andean languages serves to illustrate how the linguistic mainstream considers the whole methodology discredited. Specialists in Andean linguistics have always viewed these glottochronological dates (and those for Aymara) with the utmost scepticism, and before long Torero himself had abandoned them in his own revisions of 1983 and 1984. They are...
generally felt to be serious underestimates of the true time-depths that seem likely on all other evidence (see Cerrón-Palomino [1987] 2003, 329–33; 2000, 282, 287).

We have to fall back instead on an estimation of a broad window of possible time-depths, as indicated by the degree of divergence across the Quechua family. Starting out from the fact that it shows diversity of an order comparable to that between some of the major Romance languages, there is no particular objection to a similar time-frame for the divergence of Quechua. This would take it back to the Early Horizon in the archaeological chronology of the Andes, set at 900 bc to AD 200 on one common dating, and centred on the culture of Chavin de Huantar far inland in the highlands of north-central Peru (see Fig. 4), which would therefore seem a plausible candidate as motor for that first language expansion.

We need to be wary, however, of an evident bias in the evidence that we typically look to in order to calibrate our impressions of what we suppose is a median rate of change. The experience of the vast majority of researchers is heavily tilted towards Indo-European languages, particularly those of Europe. And it is in Europe, and over only the last two or three millennia, that are concentrated the vast majority of our ‘calibrations’ too: those instances of language divergence for which we do have reliable dates on independent non-linguistic evidence. Of the thirteen calibrations for glottochronology in Lees (1953, 118), only two were non-European: one for Egyptian, the other for Chinese. There are good grounds for arguing that this region and period may have been atypically tumultuous by the standards of other eras and other parts of the world (see Dixon 1997, 4–5), in which case our Eurocentric calibrations might well represent an unusually rapid pace of linguistic change and divergence, rather than a median reliable for other periods or regions of the globe. Here again, archaeologists and geneticists could contribute to the linguistic debate with an assessment of how the population prehistory of the Andes over the last few thousand years might compare to that of Europe in this respect.

All this leaves us with a still distinctly broad window of possible time-frames for the Quechua expansion, centred on a time-depth of the order of two millennia bc — or perhaps rather more. Beyond this, the paucity and vagueness of linguistic dating techniques forces us to look to other levels on which we might correlate more precisely the main stages in the divergence of Quechua with scenarios from other disciplines which do have reliable dating techniques. It is to these other levels that we now turn. We shall approach things first from a geographical rather than a chronological framework, to see what clues linguistics can give us here as to the most likely locations for the origins of each family, and each of the stages of its divergence.

4.5. Locating homelands: the case of the Quechua family

For a start, since most of the spread of Quechua cannot be attributed to the Incas at all, the question of where it first started expanding from is thrown wide open. At the time of the Incas themselves, the people of the Andes would hardly have laboured under the modern misconception of placing its origins necessarily in Cuzco. Writing in 1560, Domingo de Santo Tomás ([1560] 1995) was certainly under no such illusions: ‘Quechua was not the general language of the Incas and their empire; Quechua arrived in Cuzco, as an end point; but it did not start out from there to reach the coast and all parts’. He even goes on to offer a strong view on where Quechua did originate: ‘The first who spoke it lived on the coast, in Chincha, a very ancient and most powerful kingdom’. This presumably reflects a traditional local belief, so not one whose veracity we can rely on, but which is certainly of interest in the light of modern proposals, as we shall see below and in the follow-up article. For a detailed exposition of the history of the various proposals for the geographical origin of Quechua, see Cerrón-Palomino ([1987] 2003, 328–49). Here we look only at the main aspects that best illustrate the relevant linguistic principles and methods.

4.5.1. Locating homelands by degree of linguistic diversity per unit area?

Recall from §3.2 the suggestion that measures of language diversity per unit of geographical area can provide a default indication (though not a guarantee) of the likely starting point for a language expansion. For Quechua, it has long been claimed that the region of Central Quechua (i.e. north-central Peru, see Fig. 4) is characterized by particularly high internal dialectal diversity per unit of area, considerably higher even than in the Quechua-speaking regions to its north (Ecuador) or south (Ayacucho-Cuzco-Bolivia). These claims derive from the observations of specialists in the comparative linguistics of Quechua, who almost all concur in the key conclusion that the Quechua of the Central region is particularly diverse compared to that of the Southern region. Nonetheless, those judgements remain unquantified and impressionistic, and the sceptic might well question what hard quantitative data have been put forward to support these claims as to relative degrees of diversity.
Unfortunately, such hard quantifications are all but entirely wanting, and for two practical reasons. Firstly, as already observed (§3.1.4, §3.2), putting numbers on aspects of language like relative dialectal diversity is a difficult enough task in itself. Secondly, the measures we need in this case call for data at a detailed level of dialectal coverage across the Quechua family which are simply not yet available. The only existing study which approaches a workable coverage is that of Torero ([1970] 1972), though unfortunately it suffers from some significant drawbacks. Firstly, Torero published not his full data but only his results, in the form of glottochronological calculations, which as datings have long been abandoned, as we have seen. If one refrains from dating and from any assumption about rates of change, however, one can recover Torero’s original measures of degree of divergence in vocabulary (however quickly or slowly that arose). This still leaves a second caveat that even these figures cannot be taken with too much confidence in their precision, since serious questions remain about lexicostatistics as a technique for converting language data meaningfully into numbers (§3.1.4). Nonetheless, most linguists are at least prepared to take such figures as of some value at least as approximations, especially in cases such as Quechua where they are all we have available so far at the required level of dialectal coverage.

Once we have turned language into numbers like this, there are various ways in which we can process those figures to produce measures of the degree of diversity within particular groups of dialects. As a first simple statistical approach one can compare the divergence ratings for all pairs of dialects within each of the various groups. For Central Quechua Torero’s figures show a mean divergence of 0.046 (where identity = 1 and total difference = 0), and for Southern Quechua 0.043. The standard deviations within those groups are 0.0146 and 0.0143 respectively. The lexicostatistical figures recovered from Torero’s dates can also be input to certain phylogenetic analysis programmes, including one of the new network-type: NeighborNet developed by Bryant & Moulton (2002). This yields the representation in Figure 5, which helps visualize better the extent of the diversity within each of the various dialect groups.

Together, these analyses indicate that the diversity across Central Quechua has perhaps been somewhat overstated hitherto; it appears in fact to be only slightly greater than that across Southern Quechua. Nonetheless, the latter spans a much larger expanse of territory than the former (from Huancavelica all

---

Figure 5. NeighborNet of the relationships between 37 regional varieties of Quechua based on the glottochronological measures of Torero ([1970] 1972) reset to undated lexicostatistical distances.
the way to Santiago del Estero in Argentina, see Fig. 2), so per unit area diversity is indeed higher in Central Quechua. In this sense, then, these first tentative measures of dialectal diversity in the different Quechua-speaking regions do seem to fall generally in line with unqualified specialist opinion.

Recall that we cautioned that this principle of locating homelands by degree of linguistic diversity per unit area is at best a rule of thumb, and by no means always reliable. Nonetheless, in the case of Quechua the overall patterns in the linguistic data do seem to justify some assumptions along these lines, if only in that the conclusions conform closely to the rather stronger ones arrived at independently using the main alternative approach to locating homelands, i.e. by dialect geography and classification (§4.5.2 below). On the strength of these combined arguments, it is reasonable to adhere to the consensus among specialists which sees Central Peru as by far the most plausible candidate for the overall region from which Quechua must have begun expanding.

Beyond that general statement, however, more precisely where within this still extensive region might have been Quechua’s original homeland remains very much open to debate. Diversity per unit area is not of much further help, for it seems to be high right across the span of Central Quechua, from its northernmost limit in northern Ancash, down to its southernmost limit at the one fairly sharp linguistic boundary that does exist within Quechua, between the cities of Huancayo (Central Quechua) and Huancavelica (Southern Quechua) - Figure 4. Only one detail will turn out to be significant: that there is one small Quechua-speaking area, the Yauyos province in the highlands just inland from Lima, which despite its very small geographical extent shows an unusually high level of diversity for such a small area. We shall return to this shortly, but in the meantime there are more clues to be had in other data and in the final analysis technique in our linguistic armoury.

4.5.2. Locating by dialect geography and classification

More useful and convincing than locating by degree of diversity per unit area is the main alternative method for locating a possible homeland, which is to study the detail of the dialect geography of a language family. For Quechua this is a field in which a great deal has changed in recent years, indeed it represents a clear case where the branching family tree has been increasingly exposed as an unjustifiable idealization as a model for the divergence of a language family which can only lead us astray in seeking to reconstruct its true history.

The traditional classification of the Quechua family drew a fundamental two-way split between a Central Quechua branch (termed QI) and another branch (QII) which brought together all the rest, itself divided into further sub-branches. This family tree structure has come under sustained challenge since the reporting of a series of Quechua varieties inland in the Lima department, notably that of Pacaraos as described by Adelaar (1982; 1987), and a string of local varieties in what Taylor (1984a; 1987) calls the ‘dialectal microcosm’ of Yauyos, as well as those of Northern Peru (Taylor 1984b). These varieties, highly divergent among themselves, all sit uncomfortably on one side or other of the supposedly fundamental QI-QII split, on precisely those features in phonology and word-structure that had been taken as the defining ‘diagnostic’ criteria on which the traditional family tree model for Quechua was originally based (see particularly Landerman 1991).

The challenge to that model is fully supported by my own recent applications of quantitative methods and phylogenetic analyses to divergence measures on other levels. On the level of vocabulary I have used both Torero’s recovered lexicostatistical measures to yield Figure 5, and the more finely-calculated results from my own method and separate fieldwork data set (see §3.1.4 and Heggarty 2005), to yield the Neighbour-Net in Figure 6. Ongoing work (Heggarty in prep. a,b) reveals a similar picture on the level of phonetics too. All of these graphical outputs look nothing like neatly branching trees, but webs suggestive not of some radical early split in Quechua but a gradual expansion into a broad dialect continuum. Indeed the varieties of Northern Highland Peru, supposedly QIIa, i.e. a sub-branch of QII, in fact appear much closer to QI than to the rest of QII.

The (re-)classification of Quechua and its consequences for our vision of the population prehistory of the Andes — that we should no longer look for some sharp and deep historical break between populations — are taken up more in depth in our follow-up article. For our current purposes the key observation, clearest in Figure 5 thanks to its fuller coverage of the central regions, is that one group stands as unmistakeably intermediate between the two traditional branches: the Quechua of the same Yauyos region that has already deserved mention both for uncomfortably straddling the supposed primary split within the family, and for the striking degree of diversity it shows for its small area.

Yauyos is a province in the southeast of the Lima department, in the highlands a short distance inland from the coast of central Peru. The other varieties intermediate between the two extremes, moreover,
Paul Heggarty

are those of neighbouring provinces: Jauja and Chongos Bajo in the Junín department, and continuing further north a number of other areas known to be problematic for the traditional classification, through the so-called Yaru Quechuas as far as Pacaraos in the northeast of the same Lima department. Together these form a band of intermediate varieties stretching some 200 kilometres north to south, immediately inland from the coast, through the mountains of the Lima department and into neighbouring parts of Junín (see Fig. 4).

For our search for the Quechua homeland, it is very suggestive that we have now seen two key characteristics distinctive of this same region: a classificatory status intermediate with respect to the rest of the Quechua family; and particularly high internal diversity per unit area. Both characteristics suggest some crucial role for this region in the earliest stages of Quechua divergence. Either it was the homeland from which Quechua first began to expand — but in that case what would be a suitable candidate culture within this area? — or it was at least one of the first areas that Quechua spread into, and through which it continued diverging further afield, both north and south.

At first sight this location might seem to support the idea of a homeland for Quechua necessarily in the highlands, though we should not let this fool us. First, we should dispose immediately of the spurious ‘vertical’ correlation idea once also used to argue for Cuzco as a suitably high-altitude homeland (§1.1). That is of course a red herring: the same can be said of no end of important sites in the archaeology of the Central Andes, not least the foci of the Early and Middle Horizons: Chavin, which stands at 3150 m; and both Tiwanaku (c. 4000 m) and Wari (2700–3100 m).

Secondly, that Quechua survives today only in highland or jungle regions is by no means a guide to where it was spoken in the past, but simply a reflection of the topographical remoteness of these areas, which allowed the language there to resist for longer the encroachment of Spanish, which began earlier and most powerfully on the coast. That the intermediate varieties known today are inland ones says nothing about the neighbouring Quechua varieties once spoken on the coast, none of which have survived into our modern data sets. We at least have the grammar of coastal Quechua by Santo Tomás ([1560] 1995), and the closely related language of the Huarochirí manuscript of c. 1608 (see Salomon & Urioste 1991, and par-

![Figure 6. NeighborNet of the relationships between 15 regional varieties of Quechua based on the difference ratings in lexical semantics in Heggarty (2005) for 150 basic meanings.](image-url)
ticularly Taylor 2001). Both of these suggest that the Quechua of this region too was ‘transitional between the two major branches of Quechua’, i.e. QI and QII, as Cerrón-Palomino ([1987] 2003, 245–6) puts it. As for how diverse these coastal varieties were amongst themselves, we shall doubtless never know, for they are long extinct. In fact, even further research on their counterparts just inland, so critical to our understanding of the origins of Quechua, is a race against time in the face of their own imminent, or in many cases already consummated, extinction.

A highland origin, then, is by no means certain. Indeed when Torero (2002, 46) assessed the likely origins of both Quechua and Aymara, he for one had little hesitation in putting them both on the coast: for him the two families ‘originated, respectively, on the central coast — the area of the formation of proto-Chavin cultures — and on the southern coast — the area of the Paracas culture’. Such ambitious precision seems rather precipitate, however, in the eyes of most researchers, for whom Torero’s proposal remains no more than one possibility among others. Cerrón-Palomino ([1987] 2003, 22), for instance, favours a location somewhat further inland. He notes that even the very early pre-ceramic site of Caral (3000–1600 BC) has been suggested as a possible ultimate homeland of the ancestor language, though he hastens to qualify that even if so, this could only have been a ‘Pre-Proto Quechua’, i.e. corresponding to a period long before any break-up began. In other words, Caral is not yet seriously entertained as a motor for any significant territorial expansion and divergence of Quechua. Even to suggest it as a putative ultimate homeland remains, as Cerrón-Palomino rightly observes, ‘rather premature, if not to say speculative’. Still, that such authorities on Quechua even entertain the possibility of a link between the two is testimony to how uncertain our ‘linguistic datings’ are, and how far things have changed since the discrediting of glottochronology and its claims of a split some three millennia later.

5. Beyond Quechua and the Incas...

Our examination so far of the prehistory of Quechua and the role of the Incas has sought to provide an illustrative survey of how the principles and methods of comparative-historical linguistics can shed light on issues critical to our understanding of human prehistory. In this case, they comprehensively discredit the popular assumption that the spread of Quechua was the work of the Incas: both the timescale and the geography of the language expansion signal fail to match.

But if not the Incas, then who? Which archaeological culture(s) are to be identified as the motors for the expansion(s) of Quechua, and indeed of Aymara? Aside from the few pointers that have come into the discussion so far, these questions have had to remain largely unexplored within the scope of this first article. A further and fuller look at the Andes, however, opens up many more avenues for linking linguistic and archaeological scenarios with greater precision.

To close, then, readers with an interest in following up the issues discussed here are directed to a shorter sequel article ‘Archaeology and Language: a Case-Study in the Andes’, to appear in the next issue of this journal. There, the origins and expansion of Quechua are set into a wider context, not least the prehistory of the other main indigenous language family of the central Andes, Aymara, with which the story of Quechua is inextricably intertwined. Moreover, it looks beyond the Incas to the other major cultures of the Andes (including Tiwanaku, Wari, Chavin and Pachacamac), to assess the roles that they may or may not have played in the various stages of these linguistic expansions. Archaeologists who specialize in the Andes will find that the linguistic story once more turns out to be starkly at odds with a number of assumptions current in their field. For a wider archaeological audience too, the full Andean case-study serves equally well, as a further exploration of how the principles and methods of linguistics can be applied in practice to link language expansions with evidence in the archaeological record.

Paul Heggarty
McDonald Institute for Archaeological Research
Downing Street
Cambridge
CB2 3ER
UK
Email: paulheggarty@yahoo.fr

Notes

1. For their detailed and constructive comments on this manuscript, and wider discussion of the issues involved, my thanks and acknowledgements go to a number of specialists in the Andes, from various disciplines: particularly to the archaeologists David Beresford-Jones and Kevin Lane, as well as Alexander Herrera and Anne-Marie Hocquenghem; linguists César Itier and Rodolfo Cerrón-Palomino; the historian Adrian Pearce; and two anonymous and excellent referees.
2. All phonetic transcriptions here are in the International Phonetic Alphabet, which can be can be consulted at: www.arts.gla.ac.uk/ipa/fullchart.html.
3. In fact the -m was lost most likely through a sequence of two separate changes. Even Classical Latin authors observe that words written with final -m were no longer actually pronounced with an audible consonant [m] at the end of the word (i.e. without closing the lips), but just a nasal quality on the remaining vowel: original [septem] had already become [septë], see Allen (1978, 30–31). In Vulgar Latin not only had this first change probably already happened before much divergence began, but so too had the next stage, the dropping of the nasal quality on the final vowel too, i.e. [septë] had changed again to become [septë].

4. The actual pronunciations of these cognates in the various Romance languages can be heard online at my website: www.languagesandpeoples.com/Romance.

5. Non-linguists typically underestimate the degree of complexity and number of variables involved in processes of sound change. In particular, change in any one sound is, in the majority of cases, conditioned by a complex combination of criteria in that sound’s phonetic ‘environment’. This conditioning can include some or other of the phonetic characteristics of the immediately preceding and following sound (or sounds); the nature and position of stress in the word; rules of syllable and word structure (which vary from language to language); and so on. Moreover, every change in any given sound has its own knock-on effects, in that it in turn redefines the phonetic environment for the same surrounding sounds that conditioned the first change. It is no surprise that a single language like Latin can end up spawning many hundreds of clearly distinct local varieties right across the Romance dialect continuum.

6. The term natural language is used very deliberately here, and in a specific linguistic sense which excludes a number of apparent exceptions to the idea that no modern language can meaningfully be described as older or younger than another. Firstly it holds to a basic definition of language as the transmission of meaning by means of the sounds produced in the vocal tract, which therefore excludes deaf sign languages. The term natural, meanwhile, is taken to exclude both artificial languages such as Esperanto, and cases where the transmission of a language has been ‘unnatural’: by formal education, for example, rather than from parent to child as everyday speech. It thus excludes also cases such as modern-day uses of Latin or Sanskrit, where a particular group has sought to preserve unchanged a particular ‘fossilized’ form of a language in some limited (especially ritual) use. In fact such supposed exceptions serve only to prove the rule that all natural languages change, for even as these forms of language were being preserved, their natural equivalents did indeed change (and in some cases died out). As soon as Hebrew, for instance, became effectively ‘resurrected’ as a living language, it once more began to change. Even artificial and sign languages, too, develop linguistic changes once they are in natural use.

7. The term mixed language is another technical linguistic term reserved for a very specific concept. By no means should it be misconstrued and misapplied to any language which happens to share occasional traits that seem to go back to more than one other language. Non-linguists should be particularly wary of the temptation to ‘eyeball’ such traits without specialist knowledge, and to read far too much into them. Above all, one must beware of falling for common simplistic assumptions about particular languages as ‘mixtures’ of others. While one can often hear Catalan, for example, described as a ‘mixture’ of Spanish and French, such a description is quite mistaken and entirely misleading. Catalan is intermediate between them, yes; but to imagine it as a mixture of them is to fail to grasp the real nature of how languages diverge from and relate to each other.

8. Not that this has stopped the proponents of dating techniques — whether linguists or not — trying to define splits in other ways. This can be seen in the dates they choose for ‘calibrating’ their methods, as for example in Kroeber (1958, 456) for glottochronology, and Gray & Atkinson (2003, supplementary information). In a further article now in preparation I argue that their visions of language splits, and thereby also their calibrations, are seriously flawed.

9. This is not to claim that language change is determined only by socio-cultural forces — there are certainly other forces at work too, not least language-internal ones (see Heggarty 2006, 187–8). Nor does it argue that all aspects of language are to be considered ‘socio-cultural’. What does claim, to be precise, is that change in all aspects of language is susceptible to socio-cultural forces. How can this be asserted so boldly? The case of contact between languages suffices to prove the principle. While it is clear that such contacts initially have more impact on some aspects of language (notably vocabulary) than on others (syntax), it is increasingly accepted in historical linguistics that particularly intense contact between languages does appear capable of causing change in any aspect of a language — nor is there a priori any reason why any part of language should be immune to this. So, given that contact can affect change in all aspects of language, and that it is primarily external forces that determine which languages come into contact with which, it follows that such external forces do indeed have the potential to determine changes in all aspects of language. There is no better example than Quechua itself, which has seen myriad changes (particularly loan-words, syntactic calques, and so on) that it certainly would not have undergone in precisely those ways were it not for its contact with one other specific language: Spanish. The event and particular conditions of that contact were, par excellence, determined historically by forces nothing to do with language itself.

10. Archaeologists should note that the wave model of language divergence is not to be confused with the wave of advance model of demographic spread (see e.g. Renfrew 1987, 126–31): the two may go together, but languages can also diverge by the wave model without any population growth or movement.
11. Apologists for linguistic palaeontology sometimes claim that one cannot impugn their method without invalidating the comparative method itself. This is rather disingenuous given the great consensus about the latter, and widespread scepticism about the former. And it simply does not follow, given the fundamental differences in the assumptions behind the two methods as to meaning shifts and how to rule out borrowings. I take up these issues in more detail in an article in preparation specifically on linguistic palaeontology.

12. For measurements, data, recordings and forthcoming publications on this see my website: www.soundcomparisons.com.

13. One popular claim, for example, surrounds French numerals from 70 to 99, based on Latin words but combined in very non-Latin ways: 72 is literally sixty-twelve, 82 four-twenties-two, 92 four-twenties-twelve, and so on. This too has been claimed to be a substrate feature, harking back to a supposedly Celtic system of ‘counting by twenties’ as confirmation that the population of Gaul who eventually learnt Latin had originally spoken a Celtic language. On closer inspection, however, many of the linguistic details are far from clear-cut and suggest other explanations as equally or indeed more plausible; this claim is hotly disputed.

14. The actual pronunciations of these words as said in the various Quechua regions can be heard online at my website: www.quechua.org.uk/Sounds.

15. Where original citations are in Spanish the translation is my own in each case.

16. Supplementary information on the data referred to in this section may be consulted at www.quechua.org.uk/supplinfo.htm. This includes: the full grid of Torero’s original glottochronological results; the lexicostatistical counts recovered from them by undoing the application of the glottochronological formula; and the definitions of Quechua subgroups used in calculating the statistical counts of degree of divergence.

17. Statistics calculated by Dan Dediu at Linguistics and English Language, University of Edinburgh.

References


Heggarty, P., in prep. a. Getting the measure of Quechua in space and in time: pointers to the geography, history and archaeology of a language family (provisional title), to be submitted to Revista Andina.


**Author biography**

*Dr Paul Heggarty*, a comparative/historical linguist at the McDonald Institute for Archaeological Research in Cambridge, works at the intersection between these two disciplines, seeking to bring together their independent scenarios for human prehistory, especially in Europe and the central Andes. His own approach is to develop new methodologies for quantifying language divergence, so as to open up this field to powerful new statistical and phylogenetic analyses.