

1.2

Linguistics

Paul Heggarty

Language lessons on the Andes–Amazonia divide

To other disciplines that seek to understand the human past, it is not always immediately apparent how our languages can have much to say. So the task of this chapter is to set out how linguistics can indeed inform our assessment of the Andes–Amazonia divide. It also aims to forewarn non-linguist readers, before they embark on the linguistics chapters in this book. It introduces the main concepts in language prehistory that are relevant to understanding any apparent Andes–Amazonia divide in linguistics, and seeks to head off certain common cross-disciplinary misunderstandings about what those linguistic concepts do or do not really mean for our purposes.

We begin with a foretaste of how languages on either side of the divide can shed light on the (pre)histories of the societies that spoke them through time, the inhabitants of the Andes and of Amazonia. Even from just the broadest overview, striking facts stand out. Arawak, for example, is a family made up of scores of languages that all unquestionably descend from a single common origin. Many lie within the core of the Amazon and Orinoco drainages, but other notable Arawak languages spread much further afield, too (see [Figure 1.2.1](#)). Moxo is spoken in the Llanos de Moxos in lowland north Bolivia. Taíno was the first native tongue of the Americas encountered by the Europeans in 1492, and was soon to become extinct from the many Caribbean islands where it had been spoken (although some deportee populations do still speak Garífuna along the continental coast of the Caribbean from Belize to Nicaragua). Other Arawak languages were once spoken even in parts of Paraguay and northern Argentina. In short, Arawak is the most expansive of all language families in South America, spread not just across Amazonia but far beyond. And yet there was one environmental gulf that it would not cross: the Andes–Amazonia divide. No Arawak language is spoken high in the Andes or on the Pacific coast.

In the Andes, meanwhile, the one family that approaches Arawak in the scale and environmental diversity of its expansion is Quechua. Its distribution has long

been observed to overlap fairly closely with that of the Inca Empire, although that parallel is a beguiling one that has also led to many superficial and anachronistic presumptions about Quechua's prehistory (see Beresford-Jones and Heggarty 2012b, 4–6). In the one respect most relevant to our theme here, however, the parallel does seem to hold. In pre-Columbian times, at least, Quechua did largely mirror a much-noted characteristic of the Andean societies that speak it: a reluctance to venture into Amazonia.

Indigenous languages can inform the Andes–Amazonia question, then, not least because they can be categorized, on specific linguistic criteria, into larger groupings of languages that go together in some way. One can then explore whether those entities or groupings have, through prehistory, either aligned with the Andes–Amazonia frontier, or crossed it. And for a further perspective on how meaningful any divide might be, one can also assess how far linguistic criteria define either just a single, coherent unit on either side of the divide, or multiple entities fragmented by further dividing lines within each region.

Also, as the structure of this chapter implies, it is not all about language families, like Arawak or Quechua. Families are just one of the *two* main levels – which moreover can crosscut one another – on which languages can be analysed into larger entities. Besides language families, the second level is that of 'linguistic convergence areas'. These are far less well known outside linguistics, and are often confused with families, when in fact for prehistory they mean very different things. A first indication is the contrast already evident between [Figure 1.2.1](#), which maps the main divergent language families in South America, and [Figure 1.2.2](#), which maps the main linguistic convergence areas.

Language families: Origins, expansions, migrations and divergence

So to begin with language families, what does a label like Arawak or Quechua really mean for our purposes here? The key is that any language family attests to a process of geographical expansion through time. By definition, every language family started out as a single ancestral language, from which all its 'daughter' languages descend. Spoken languages are always changing, however, incrementally through the generations. And if by some process of geographical expansion – demographic and/or cultural – a language comes to be spoken in different regions whose populations are no longer in constant contact, then from that point on, different changes can arise in different regions. These changes can affect all levels of language: vocabulary, sound system, grammatical system, and so on. Ultimately, so many changes accumulate, so different from one region to the next, that the original source language ends up effectively diverged into what have become its different 'daughter' languages. What also follows from this natural process of divergence, once a language is widely dispersed, is that the common ancestral 'proto-language' of any

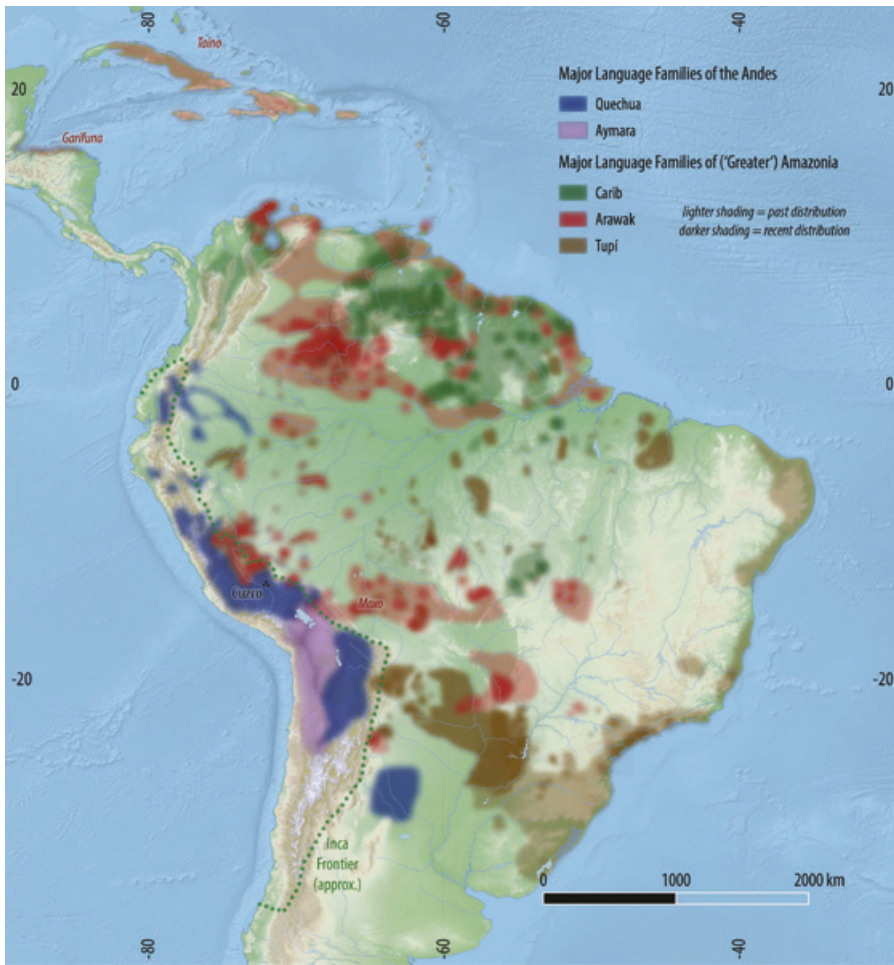


Figure 1.2.1 The main expansive language families of the Andes and Amazonia. © Paul Heggarty. For a closer view along the Andes–Amazonia transition, see Chapter 3.4, Figure 3.4.1.

family must originally have been spoken in just a relatively small region, and its divergence into a family came about in the first place only because of its expansion out of that homeland (see Heggarty and Renfrew 2014a, 23).

For a concrete illustration of how a language family arises by geographical expansion and divergence through time, the classic, historically known example is that of the Romance language family in Europe. In this case, the real-world driver that caused the family to come into being is very clear. The Roman Empire brought much of Europe to speak *Romanice*, ‘in the Roman way’ – in other words spoken, ‘Vulgar’ Latin. But once so dispersed, Latin was free to change in different ways in each new region. By today, the ‘neo-Latin’ spoken in those different regions has become so

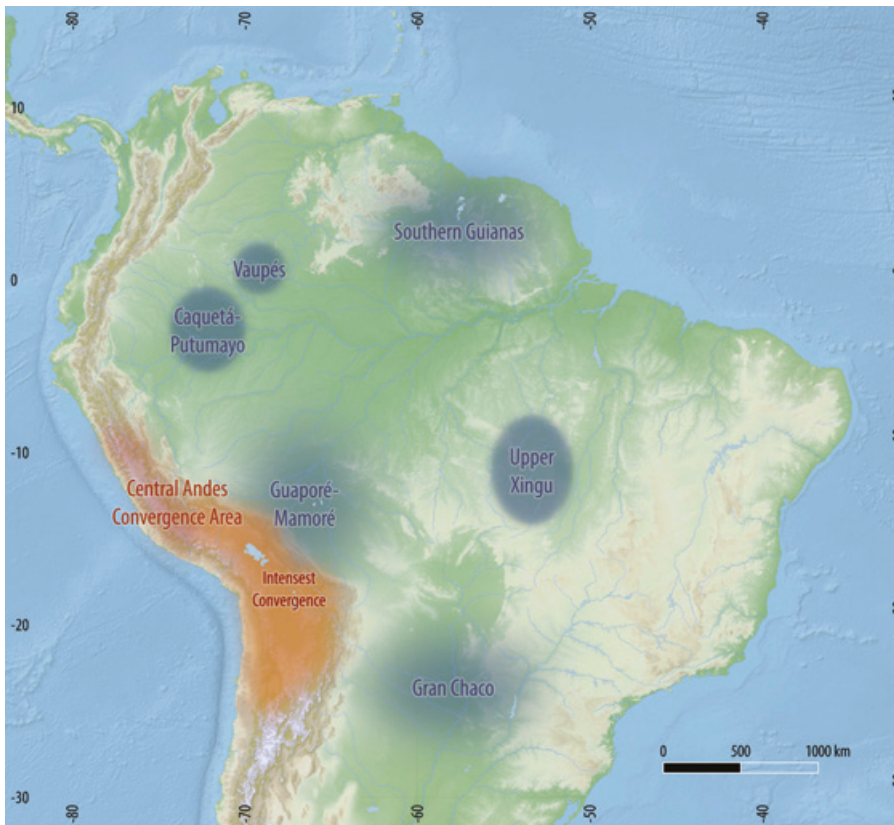


Figure 1.2.2 Zones of especially intense language interaction (‘linguistic convergence areas’) within South America, based on Beresford-Jones and Heggarty (2012b) for the Andes, and on Epps and Michael (2017) for the lowland languages. © Paul Heggarty. Earlier proposals of a looser convergence area stretching much more widely across most of Amazonia are increasingly challenged: see text, and Chapter 3.5.

divergent as to form the *family* of the various *Romance* languages. Amongst them are Romansch and Romanian, aptly named, but also Italian, French, Spanish, Portuguese and Catalan, and scores of lesser-known sister languages and dialects.

That Arawak is a ‘family’, then, also means that it is the set of languages that all go back to the same Proto-Arawak source language, but have long since scattered and diverged into significantly different languages, no longer mutually intelligible. Likewise for Quechua. Divergence within Arawak is actually somewhat greater than within Romance, whereas within Quechua it is if anything a little less. Since divergence is cumulative through time, the default implication is that Arawak has been dispersing and diverging for longer than the two millennia since the spread of *Roman(i)ce*, and Quechua for a little less than that. (Linguists have long been

dissatisfied with such impressionistic statements, of course, and have tried to put more precise, strictly cross-comparable numbers on degree of linguistic divergence. The nature of language itself, however, continues to pose serious methodological challenges to that goal.)

The key to what any language family means for prehistory, then, is that Romance did *not* ‘crystallize’ out of some process of convergence out of some diverse ancient speech already across Europe. On the contrary, Latin spread to *replace* almost all other language lineages previously spoken across much of continental Western Europe (the famous exception being Basque). Romance came about by a process of divergence, out of Latin, once it had dispersed. Until the rise of ancient Rome, Latin had been spoken only in that city and the province around it, Latium (modern Lazio), whence its very name, *Latin*.

Likewise, Arawak, as a language family, must originally have gone back to a much smaller homeland region, out of which it expanded. So too must Quechua. Each family must thus also have had reasons or ‘drivers’ for its geographical expansion – although by no means necessarily an empire like Rome, since many other processes can also drive demographic and/or cultural expansions that can take languages with them. Indeed, directly relevant to our theme is whether the expansions of the major language families in the Andes and in Amazonia were driven by similar types of demographic and/or cultural processes, or by very different ones on either side of the ‘divide’. If the two regions did indeed have radically different socio-political and demographic histories, then the processes that spread Arawak, for instance, might be expected to be correspondingly different to those that spread Quechua. Arawak may have no good analogues, then, for those late phases of Quechua expansion that seem to result from major, state-directed reconfigurations of Andean demography by the Incas. Certainly, languages do not necessarily require demographic dominance to spread. (That said, the languages of small demographic *elites* have typically fared badly before the modern era, except in particular ‘*primus inter pares*’ conditions: see Heggarty 2015, 622–3.) Quechua itself illustrates occasional expansions with precious little demographic trace, and precisely in the exceptional cases where it did spread down from the Andes into some parts of Amazonia, as explored linguistically in [Chapter 2.3](#), and genetically in [Chapter 3.3](#) by Barbieri. For, as in those cases, a particular socio-cultural context can confer utility on a language, making it a target for populations to switch towards. Still, that utility derives not from anything in the language *per se*, but from the scale, power and/or cultural prestige of the populations and cultures that (already) speak it. The language is carried along with a broader cultural package that is doing the expanding.

So it is not as if language families themselves have some innate and somehow ‘linguistic’ propensity to spread of their own accord. Their distributions stand very much at the *effect* end of a cause-and-effect relationship. Indeed, if language families can attest to the operation of expansive processes in prehistory in the first place – whether demographic and/or socio-cultural – then that is because they are the

direct results of those real-world processes (see Heggarty 2015, 600–2; Heggarty and Renfrew 2014a, 19–21). And in all cases, the basic principle remains: whichever particular expansive mechanisms lie behind any given language family, they are still expansive and divergent in nature, not convergent. The fundamental process that creates a family is still one of geographical spread, not convergence *in situ* in some form of network. It necessarily entails at least some migration of speakers, to carry the language lineage to other regions. This holds even if thereafter, in addition, locals may also switch to speaking the language of those incomers, for cultural and/or demographic reasons.

It is also these migrations, their directions, sequence and stages, that determine the structure of the ‘family tree’ of descent within each family, its branches and sub-branches. Those past processes thus remain encoded in that tree structure, hence the value for prehistory of recovering it by comparative linguistics. (Hence also the discipline’s near obsession with sound change laws especially, as the most reliable diagnostic for establishing those trees.) The Quechua of Cuzco and that of Bolivia, for example, share distinctive changes that define them together in the family’s far southern (or ‘QIIc+’) branch. These changes thus effectively prove that the Quechua of the Bolivian Altiplano can be derived from a movement of speakers southwards from the Cuzco region, at a relatively late stage, in the Inca and/or Spanish colonial period – and that the Quechua of Central Peru cannot.

The origins and main dispersals of the major language families of South America lie far back in prehistory; the shallow historical record here catches only their last phases. But this makes comparative/historical linguistics all the more valuable, because the discipline enjoys so many known historical test-cases, like Romance, that it has been able to develop and test its comparative methodology, and confirm the validity of its results against ancient written languages. By now, the same methods can confidently be applied without even requiring a historical record – and in some respects can even partly make up for the lack of one, in regions like South America.

Language families, then, can offer various perspectives on the Andes–Amazonia question. The first lies simply in how they map out across the continent, as we have already seen for Arawak and Quechua. That first illustration can seem unequivocal, in supporting the reality of a divide. On closer inspection, however, it turns out that the constraint not to trespass from the Andes into Amazonia does not hold up entirely, as explored in the ‘language families’ section of [Chapter 3.4](#). That chapter surveys what else *families* can tell us of the Andes–Amazonia divide in various other respects, too, beyond any such ‘trespassing’ taboo.

Contact and linguistic areas: Interaction and convergence out of diverse origins

In any case, there is plenty more that language can tell us about the reality or otherwise of an Andes–Amazonia divide, on another level that has nothing to do with

families. For relationships of common descent ('families') are only one way of looking at languages. The linguistic panorama includes another quite separate dimension that can cut across language family distributions, and indeed often does. This is only natural, in fact. For while a language *family* is the result of geographical expansion out of a single origin and the ensuing language divergence, it is hardly as if human societies only ever undergo processes that are expansive and divergent. On the contrary, groups with diverse origins can come into and remain in contact and interaction with each other. Intense and/or long-lasting interactions result in powerful processes of convergence. These too have their corresponding impacts in language – and most importantly, these impacts are not the same as the signals left by shared origin and divergence.

Languages can in fact display a whole scale of different degrees of intensity of contact effects upon each other (whether reciprocal or predominantly one-way). And for (pre)history, those different degrees of contact effects attest to different corresponding real-world contexts, of ever stronger interaction between the populations and societies that spoke them. For the purposes of this book, then, it is crucial to assess how intense was the level of past interactions between the Andes and Amazonia, as still recorded in their languages.

To start from the weakest indications, individual words may be borrowed from one language into another. Naturally, this happens especially with words for anything that is new to the speakers of one language, but already known and referred to by speakers of another. Just as European languages resorted simply to borrowing in words such as *llama*, *puma* or *coca*, it is natural that when people on one side of the Andes–Amazonia divide needed to refer to species or concepts typical of the other environment, they could simply borrow a word for it from one of the languages of that other environment, particularly an immediately neighbouring language along the divide itself.

Occasional loanwords for species or concepts 'alien' to the borrower language do not prove much more than the most limited interaction, however. On a greater scale are *Wanderwörter*, 'wandering words' that range far and wide, irrespective of language family, so much so that it can even end up unclear which family they actually originated in. For an idea of what these *Wanderwörter* can in principle tell us of the past, consider some well-known, long-range examples across Europe, such as words for *coffee*, *sugar*, *tea*, *potato*, or even *lion*, and mythical concepts such as *dragon*. These words in modern European languages even bear phonetic details indicative of which different external source they were loaned from, or indeed of how and when they were loaned serially from one language to another. (Note how English *café* differs from *coffee*; each tells a separate history. The former attests to French cultural influence in the late nineteenth century, and the latter to how the drink had first reached Europe some three centuries earlier, ultimately from speakers of Arabic, but only through speakers of Turkish as the intermediary traders.) Such *Wanderwörter* make for linguistic traces of the exchange routes of the corresponding real-world products, or the cultural networks through which concepts and

mythologies spread. South America offers plenty of intriguing examples, often the names for species, tradable items or cultural concepts (Epps 2017). A natural explanation is that such words spread through exchange networks. A caveat, however, is that far-flung *Wanderwörter* have often tempted unsuspecting scholars to read rather too much into them. They still need not mean anything more than a chain of local networks of ‘down-the-line’ trade, for example. Loanwords often spread through languages in series, just as *llama* reached Europe not by direct contact with the source language, but via Spanish (hence the *ll-*, even in the English spelling).

So however far they roam, individual loanwords remain only the most superficial form of language convergence, and they may result from limited exchanges involving just a few members of a community. Evidence for much more sustained and widespread interaction lies rather in whole swathes of loanwords that overtake significant proportions of the vocabulary, as with the flood of Norman French loanwords that reshaped much of the vocabulary of English. Even that, however, falls short of the next level up in ‘interference’ effects between languages, the quantum leap when those go beyond the vocabulary and encroach upon the sound and grammatical systems and structures of the languages involved. An example of such a ‘structural’ characteristic is how a language orders the components in a basic sentence, as subject-verb-object (svo, as in many European languages), subject-object-verb (sov, in many South American languages), or some other order. (Many other structural characteristics are illustrated by Van Gijn and Muysken in [Chapter 3.5](#).)

Where a language switches to adopt a deep structural characteristic of another language, this typically attests to a past phase of widespread bilingualism, if not multilingualism. Where such a phase ends up with a community switching from its original language to that of another population, then the contact effects can be particularly far-reaching. The generation(s) involved can carry over (unawares) structures from their original native tongue into the new language that they are (thus ‘imperfectly’) learning. At its most extreme level, the result is the wholesale restructuring of the sound and/or grammatical system of one language on the structural model of another. One such case arose between early forms of Quechua and Aymara, which has a bearing on the Andes–Amazonia question in ways taken up in [Chapter 2.3](#).

Moreover, language interaction need not involve only two languages. Indeed, the scale of the Andes–Amazonia question requires us to zoom out to look at how language convergence phenomena pattern much more widely. At the broad, multi-language level, linguistics employs a concept that is in many ways the antithesis of a language family, and of the process of separation and divergence by which that arises. On this other dimension, of contact and interaction, the basic concept is instead that of a ‘linguistic area’, shorthand for ‘linguistic *convergence* area’. This denotes a region across which multiple languages share certain structural characteristics, which, however, they did not all originally have, and have come to share only through contact and interaction.

To illustrate this more concretely, we take some of the evidence that Dixon and Aikhenvald (1999, 8–9) invoke to argue that Amazonia is a linguistic

convergence area (even if their case is today challenged; see [Chapter 3.5](#) by Van Gijn and Muysken, and [Chapter 3.4](#)). Amazonia is home still to scores of languages that are entirely mutually unintelligible and belong to dozens of different lineages with independent origins. Yet despite that, and irrespective of which family they come from, many languages here have (through interaction) come to share certain fundamental characteristics of language structure. Dixon and Aikhenvald (1999) list 15 of these, although here we illustrate only the less technical ones. The sound systems of Amazonian languages generally do not distinguish *r* from *l*, for example (as Chinese also does not, entirely coincidentally), and they typically have five basic vowels (*i, e, a, i, u/o*), as well as nasal vowels (as in Portuguese *São* or French *un bon vin blanc*). Their grammatical systems, meanwhile, have extensive gender systems, but few grammatical cases, and most allow prefixes.

The illustration becomes clearer still when Dixon and Aikhenvald (1999, 9–10) then look to the contrast with the opposing linguistic area of the Andes. Here, languages have converged instead on other structural characteristics, many of them diametrically opposed to the Amazonian ones. That is, their sound systems *do* distinguish *r* from *l*, but have only *three* vowels (*i, a, u*), and *no* nasal vowels. Their grammatical systems have *no* gender, *many* grammatical cases, and do *not* allow prefixes. Quechua and Aymara share all these characteristics, and more, making them very alike in the underlying nature of their sound and grammatical systems. They nonetheless remain utterly unintelligible to each other – inevitably so, because they are not of the same language family.

What defines a linguistic area, then, are effectively characteristics that are shared *not* because of common inheritance. Indeed, by default, a linguistic area spans languages from multiple *different* families and origins. When linguistics employs the term ‘areal’, then, tacit within that is the concept of (arisen by) convergence out of different origins.

To be clear, however, to avoid any dangerous misunderstandings: what emerges out of such convergence processes is *not* a new ‘hybrid’ language, and certainly not a *lingua franca*. Convergence can never go so far as to make two unrelated languages somehow become intelligible to each other, let alone identical. A linguistic area is nothing like this: it is merely a collection of unrelated languages, still radically different in countless ways, that have become alike only in certain deep structural features.

On this second main dimension of the linguistic panorama, languages in South America attest to interaction effects of all types, scales and degrees of intensity of interaction, from individual loanwords to full-blown structural remodelling. And there is interaction both between individual pairs of languages and across much wider linguistic convergence areas. For the Andes–Amazonia divide, the question is whether these convergence effects pattern geographically in ways that either respect or disqualify the idea of a divide. And, whether the convergence effects visible *within* the Andes and within Amazonia are far stronger than whatever convergence there has also been *between* the two regions. These are the themes taken up

in the third section of [Chapter 2.3](#) – where the overall picture does appear broadly compatible with an Andes–Amazonia divide, albeit with many attendant qualifications – and again in [Chapter 3.3](#).

Confusions and clarifications: Divergent families versus convergent areas

This fundamental contrast between language families and linguistic areas – between divergent *versus* convergent processes – helps to place the various linguistic contributions to this book in context, and to understand the different perspectives they give on the Andes–Amazonia question. Firstly, the families/areas contrast is the obvious criterion used to structure the overview, in [Chapter 3.3](#), of the broadest-scale patterns in the linguistic panorama with respect to the Andes–Amazonia frontier. [Chapter 3.5](#) (by Van Gijn and Muysken) focuses on linguistic areas, and presents a wide-ranging, quantitative assessment of the degree of convergence in structural characteristics between many languages of the Andes and of Amazonia. Most importantly, it also assesses differences within Amazonia, between languages nearer to and further from the Andes. [Chapter 4.1](#) (by Adelaar), meanwhile, looks at language families, but beyond the clearly established ones that do not significantly cross the Andes–Amazonia divide. It explores instead a hypothesis of an even wider, deeper relationship that would, if true, mean that one Andean language significant in prehistory (Puquina) might in fact have originated in a major lowland family. [Chapter 4.2](#) (by Zariquiey) also looks at a past hypothesis of a ‘long-range’ family relationship across the divide, only to debunk it. In the process, however, it finds evidence for a potential linguistic convergence area instead, and one that would indeed span the Andes–Amazonia divide.

Linguistics and genetics, classification and admixture

On this fundamental issue of distinguishing divergent language families from linguistic convergence areas, a clarification is needed to address a common misconception across the disciplines, in this case particularly with genetics. This is about what goes by the name of ‘language classification’. The defining criterion – tacit and understood in linguistics, and therefore potentially misleading to other disciplines – is direct *descent* of a language, in an unbroken chain of transmission and intelligibility through the generations, even as *modifications* do progressively build up. (Note the model of *descent with modification*: the process is best conceived of in terms of language *lineages*, more analogous to species, rather than in terms of discrete language units, as if they were individual organisms.) So by comparison with genetics, for example, there is nothing on the scale of the roughly 50–50 recombination of all autosomes with each new generation. On this criterion, it is a black-and-white ‘yes’ that English is of the Germanic family, because

it descends in an unbroken chain through the generations from Proto-Germanic. However much ‘admixture’ later came into it from Norman French, there never was a chain through the generations from English back to Latin. So English does not classify as a Romance language: again, a clear-cut ‘no’. Likewise, the classification of the Quechua of Ecuador and Bolivia is entirely clear-cut: both are of the Quechua family, transmitted through the generations from Proto-Quechua. Yes, Ecuador Quechua underwent convergence effects with other indigenous languages of Ecuador, as is perfectly well known to any linguist working on it. But such effects belong on the separate level of convergence; they are not part of the classification proper.

For a very rough analogy with human genetics, in linguistics it is as if it is both necessary and generally fairly easy to detect and exclude all impact of admixture (in autosomes), and as if classification were done entirely on the level of a uniparental marker that gives a clearer phylogeny of descent. Admixture effects are a key part of what we know of languages like Ecuadoran Quechua, but non-linguists should *not* expect to find them within the *classification* as Quechua. They are analysed on a quite separate dimension of contact and convergence effects, ‘despite’ the ancestry chain back to Proto-Quechua. Indeed, for the purposes of classification they are *confounds*, to be set aside to prevent them clouding the identification of direct descent.

This is hardly to say that contact effects are ignored by linguists – anything but. It is just that they (rightly) need to be kept separate from the task of classification into families. It is in fact a strength of linguistics that it has a developed methodology that generally does allow us to tease apart what is inheritance and divergence from what is contact and convergence. Geneticists would not confuse autosomal and uniparental markers, or assume that either will give the whole signal. Likewise, when comparing with linguistics, the different markers need to be compared independently with the different levels of language data – on convergence effects as well as on family classification – that correspond most closely.

Definitions and circularities?

The Introduction to this book identified how the very terms ‘Andean’ and ‘Amazonian’ can end up compressed and stretched, respectively, away from their basic geographical definitions. Linguistics seems particularly guilty of this, on both dimensions of divergent families and convergent areas. And this carries a risk that such malleability might end up in a self-fulfilling definition of a divide.

Perhaps more than in any other discipline, linguists have let their very data source shape their thinking towards a ‘Greater’ Amazonia. In lowland South America, the main language families spread far beyond Amazonia proper, through the Caribbean and much of Brazil beyond the rainforest. But those wider distributions are then what linguists have effectively taken to define an area of interest. Epps and Michael (2017, 935), for instance, put it thus: ‘Amazonia, which we define

loosely here as the lowland region drained by the Amazon and Orinoco Rivers and extending to the northern and eastern littorals of the continent'. They then cite other leading linguists of the region who do much the same: Dixon and Aikhenvald (1999, 4) and Rodrigues (2000, 15). This usage extends to the other dimension of a hypothesized Amazonian linguistic convergence area, too. Here, the dangers of circular definitions are even greater. For a language *family* does generally allow for a very clear-cut definition of which languages are or are not its members. Convergence areas, however, typically have a diffuse core-and-periphery structure and are defined by only partial overlaps in a bespoke collection of structural criteria, cherry-picked by researchers. Their exact geographical distributions, then, are much more malleable.

Conversely, and also as foreshadowed in the Introduction to this book, in linguistics as in some other disciplines, 'Andean' tends to be focused by default on just the *central* latitude band of the Andes. Again, this does not just *happen* to be the heartland of the two main families, Quechua and Aymara; rather, they have helped define that focus anyway. This narrow definition of Andean is reinforced on the convergence dimension, too, because Quechua and Aymara are the same two families that constitute the core of the 'Andean' linguistic area. Some of its defining structural characteristics actually begin to be lost even in the northernmost varieties of Quechua, in Ecuador and southern Colombia, through partial assimilation to local languages that are only peripheral, at most, to what is in reality mostly just a *Central Andean* convergence area.

In other words, linguists have conveniently stretched and compressed their Amazonia and Andes in line with known language patterns, in any case. The two regions are defined in part by the ranges across which the major language families have spread, and/or across which certain hand-picked structural characteristics are widely shared – and this in a context of widespread pre-existing conceptions of contrasting 'Andean' and 'Amazonian' realities. The effect can be to make the two regions appear as linguistically self-contained and coherent units that contrast with one another more starkly than they would if one kept to the stricter, geographical senses of the terms Andes and Amazonia (as discussed in the Introduction of this book). The impression can be further heightened because linguists use 'Andean' with a focus on those same central latitudes where the highlands abut onto the Amazon basin proper.

Other disciplines, of course, should also reflect on whether they too have preferred working definitions of Andes and Amazonia that risk turning the divide between them into a self-fulfilling prophesy.

The linguistic perspective: Potential, limitations and prospects

This chapter aspires to have clarified that linguistics has much potential – at least in principle – to help uncover the past, and to inform on the Andes–Amazonia question

specifically. There is one great proviso to this, however, that is particularly acute in South America. Before all else, the ability of linguistics to help is premised on having adequate language data in the first place. But documentation is still sorely lacking for many indigenous languages in Amazonia, which are dying out faster than a small band of fieldwork linguists can analyse them. In much of South America, it is already too late, including in the Andes of northern Peru, for example, a graveyard of languages that have vanished all but undocumented. Most of the indigenous linguistic diversity at first European contact is already long extinguished, and it is a race against the clock to record the little that remains. The result is that, for many a language in South America, for now we still have precious few clear answers on the where, when, how and why of its origins and expansions – and in some cases we will simply never be able to know. Similarly, as yet we have little in the way of consistent, large-scale databases of loanword and structural convergence across the continent, although ongoing work suggests improving prospects here, such as Epps (2017) or the database on which [Chapter 3.3](#) is based.

Another general proviso is that for all the strengths of linguistics in its internal methodologies, it is rather less straightforward to step from language family tree diagrams or statistical measures of convergence into the precise real-world contexts in prehistory that they might denote. Linguistics has developed various methods to try to bridge the gap from the prehistories of languages to those of their speakers, but most remain contested. A general exploration for non-linguist readers is Heggarty and Renfrew (2014a). Individual methods are set out in detail in many general works on historical linguistics, such as Campbell (1997), while Heggarty (2015) provides a briefer survey. Other introductions focus on South America in general (such as Heggarty and Renfrew 2014b), on Amazonia (like Epps 2009, and Epps and Michael 2017), or on the Central Andes (for example Heggarty 2007, 2008).

Obviously, the full details of those methods are beyond the scope of this chapter, which has focused instead on providing clarity on just the most basic linguistic concepts and principles that frame any attempt to learn about prehistory from linguistics. On the strength of this, it is hoped that readers from other disciplines are now better placed to approach the linguistics chapters within this book. [Chapter 3.4](#), particularly, will build on the general methodological background set out here, to offer a large-scale summary of what the great language families and linguistic convergence areas of South America mean in practice for the linguistic reality, or otherwise, of an Andes–Amazonia divide.