Pearce, A.J., Beresford-Jones, D. G. and Heggarty, P. (eds.) 2020. Rethinking the Andes–Amazonia Divide: A cross-disciplinary exploration. London: UCL Press.
https://doi.org/10.14324/111.
9781787357358

3.5
Highland–lowland relations: A linguistic view

Rik van Gijn and Pieter Muysken

Introduction

It has long been the prevalent view in ethno-history, archaeology and linguistics that the Andean and Amazonian cultural spheres form separate worlds, with little interaction between them. Some scholars, however, most notably in anthropology, have voiced different opinions, as expressed particularly in Chapters 1.4 and 1.5 in this volume, and in the extensive discussion of these contrasting visions in the introduction to this book. Among the best-known analyses suggesting that the separation between highland and lowland cultures was not always as evident as it appears to be today is that of Renard-Casevitz et al. (1988). Based on ethno-historical and (to a lesser extent) archaeological evidence, they argue that a lively trade existed in pre-Columbian times. In their view, the gradual decline of highland–lowland interactions is connected to the disintegration of the Wari cultural complex and the subsequent turbulent period in the lowlands, where local feuds and migrations had rendered the lowland polities less reliable allies for highland peoples. From then on, highland expeditions into the lowlands (and vice versa) slowly decreased in number, but in fact contacts persisted until well into the Inca era. Highland–lowland interactions probably took place predominantly in different directions in different periods. Earlier on, lowland groups possibly helped shape highland cultures. A case in point is the role that Arawakan cultures possibly played in the creation of complex highland societies, as in the case of Tiyawanaku, which through one of its main languages, Puquina, may be linked to the so-called Arawakan matrix (Santos-Granero 2002) although the evidence for this is indirect (for more detail, see Chapters 4.1 and 4.3 for an archaeological perspective). Later on, in the centuries preceding and following the Spanish conquest, highland cultures influenced the lowlands. Linguistic evidence for this comes in the form of Quechua varieties spoken in the lowlands, and the loanwords from Quechua into many languages of the eastern slopes and Amazonia proper.

Here we adopt the methods of linguistic typology, which means that we systematically compare features across languages, rather than primarily looking at family relationships (see Chapters 1.2 and 2.3 for more on this general distinction within linguistics, and what it means for interpretations for prehistory). The study of language structure (that is, the grammatical 'architecture' of languages)[1] has lagged somewhat behind other disciplines in recognizing the more intricate and gradual transition between the highlands and lowlands; a number of linguistic overviews of the area are based on the presumption of a sharp distinction (Torero 2002; Adelaar 2008, 2012a; Derbyshire and Pullum 1986; Dixon and Aikhenvald 1999). This distinction has the virtue of clarity, but it is ultimately not very helpful as it is too simplistic. There is now a large literature on the broad outlines of the geographical distribution of grammatical characteristics of South American languages, which suggests a rather different picture. Generally speaking, the following broad conclusions can be drawn.

1. There is wide typological diversity among the languages of the continent. However, it has been repeatedly observed that a number of grammatical characteristics are shared by many South American languages over large geographical areas, and across language families (see for example, Van Gijn 2012, 2014a, 2016, for studies of such widely shared individual features). In a global study based on the data provided in Dryer and Haspelmath (2013a, 2013b), Dediu and Levinson (2012) conclude that the language families of South America are somewhat more similar to each other than those of other continents, in that they seem to share partial profiles.[2]
2. There is a central Andean cluster (termed CAC here), encompassing the two language families most widely diffused in the Andes, namely Quechuan and Aymaran. Morphological and phonological evidence would suggest that Aymaran was the original model (Adelaar 2012a; Muysken 2012b), given that it appears more irregular and complex than Quechuan. Puquina and Uru-Chipaya are also influenced by this cluster, but show features of their own, while Mochica on the north coast of Peru, for example, was very different (Kerke and Muysken 2014).
3. More broadly, several families in the western part of South America, such as Barbacoan (with languages spoken in western Ecuador and south-western Colombia) and Jivaroan (with languages spoken in northern Peru), vaguely resemble the languages in the CAC (Muysken et al. 2014b).
4. Languages in the foothills may tend more towards the CAC profile or to an Amazonian profile, but most show a mixed signal in their structural characteristics (Van Gijn 2014b).
5. In terms of grammatical language profiles, there is indeed a broad east–west division in South America (Krasnoukhova 2012, 2014; Birchall 2014a, 2014b). In these studies, the dividing line between the two regions does not, however, coincide with that between the Andes and their foothills with Amazonia. Where broad generalizations can be made, the foothill languages

resemble their Andean neighbours structurally more than the more easterly Amazonian languages.

6. Overall, the languages in the western part of the continent show less diversity than those in the east (Muysken, Hammarström, Krasnoukhova et al. 2014), broadly speaking. The similarities of the languages in the west may be leftovers from very old relationships, too deep to be detectable by orthodox methods of recovering shared descent (see Chapter 2.3), or may result either from long-standing interaction zones, or from recent convergence due to ethnic reshuffling in the wake of the European invasions.

In this chapter we zoom in on the transition area between the Andes and Amazonia: the upper Amazon area. This is defined here as a broad strip of land between the Andes to the west and Amazonia to the east, and roughly between the Putumayo River that separates present-day Ecuador from Colombia in the north, and the savannahs of the Gran Chaco in Paraguay and northern Argentina in the south (see Figure 3.5.1).

Structural features are shared or differ between the languages of the highlands and lowlands in a complex and multi-layered network; to represent it fully will ultimately require the concerted effort of specialists from several subdisciplines. Dixon and Aikhenvald (1999, 10) mention that 'there is no sharp boundary between the Amazonian and Andean linguistic areas: they tend to flow into each other'.[3] The goal of this chapter is to come to a more refined picture of how these areas 'flow into each other', by focusing on how specific structural features are distributed geographically across the languages of the upper Amazon and adjacent areas in Amazonia and the Andes, building on an approach developed by Van Gijn (2014b). In particular, we will be concerned with the role of elevation differences in shaping the distributional patterns. In the next part of this chapter we introduce the language sample and the choice of linguistic features; following this we discuss the patterns that emerge and what these mean. In further work we will also try to explore the region through a fine-grained analysis of the individual river systems, but this chapter presents a more global exploration, building on Van Gijn (2014b).

Approach

The upper Amazon is characterized by the many rivers that rise in the Andes and come together further eastwards to form the great Amazon River. The sediments of this abundance of rivers, in combination with the differences in elevation between the Andean slopes and Amazonian lowlands, create a landscape of great ecological diversity, which is matched by the cultural-linguistic diversity in the region. The western part of South America is among the linguistically most diverse zones in the world in the diversity of independent language lineages (Dahl et al. 2011).

For specific parts of the eastern slopes it is also structurally highly diverse (Dahl 2008).[4] In particular, both the northern edge of the upper Amazon, in Ecuador and northern Peru, and the southern edge in Bolivia, are extremely diverse.

Figure 3.5.1 Map of the upper Amazon. © Rik van Gijn and Pieter Muysken.

Sample

Given this diversity, and because we are especially interested in local patterns, we have sampled as densely as possible, wherever languages are well documented enough for us to include them. We have also included languages spoken in the adjacent parts of Amazonia and the Andes, to gain a more complete picture. The sample is presented in Figure 3.5.2 and Table 3.5.1 (affiliations and locations are based on Hammarström et al. 2015).

A reviewer correctly notes that the locations of specific languages have changed over time, and that taking present location as a point of reference may thus

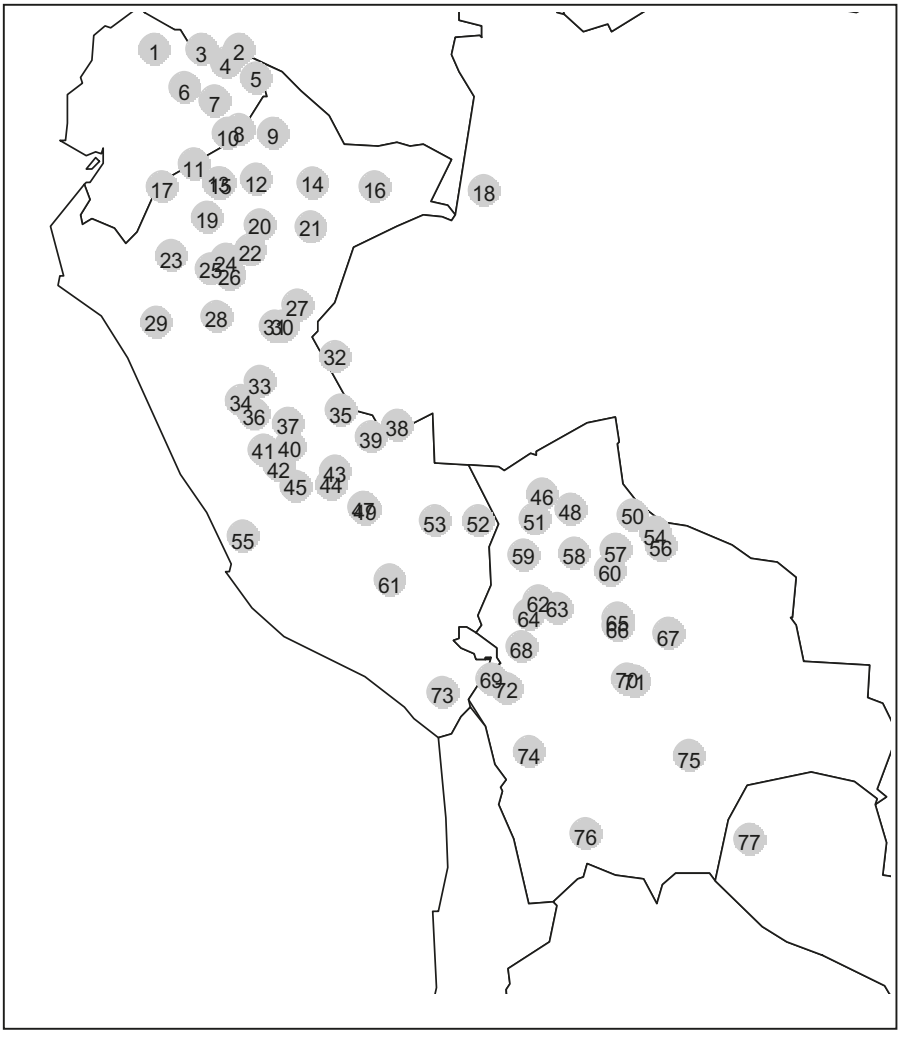

Figure 3.5.2 Map of well-documented languages of the Andes and upper Amazonia covered in this study. © Rik van Gijn and Pieter Muysken.

Table 3.5.1 Sample languages, affiliations, ISO codes, and main sources.
© Rik van Gijn and Pieter Muysken, based on Hammarström et al. 2015.

No.	Name	Affiliation	ISO	Main source(s)
1	Imbabura Q	Quechuan[a]	qvi	Cole (1982)
2	Siona	Tucanoan	snn	Bruil (2014)
3	Cofán	Isolate	con	Borman (1962); Fischer and Van Lier (2011); Tobar (1995)
4	Napo Q	Quechuan	qvo	Mercier and Marcos (1979)
5	Secoya	Tucanoan	sey	Johnson and Levinsohn (1990)
6	Tena Quechua	Quechuan	quw	fieldwork notes Muysken for Arajuno
7	Waorani	Isolate	auc	Peeke (1973, 1991); Saint and Pike (1962)
8	N Pastaza Q	Quechuan	qvz	Nuckolls (2010)
9	Arabela	Zaparoan	arl	Rich (1999)
10	Záparo	Zaparoan	zro	Peeke (1991)
11	Achuar	Jivaroan	acu	Fast and Fast (1981, 1996)
12	Taushiro	Isolate	trr	Alicea Ortiz (1975a, 1975b)
13	Andoa	Zaparoan	anb	Peeke and Sargent (1959)
14	Iquito	Zaparoan	iqu	Eastman and Eastman (1963)
15	S Pastaza Q	Quechuan	qup	Landerman (1973)
16	Yagua	Peba-Yaguan	yad	Payne (1985, 1986)
17	Shuar	Jivaroan	jiv	Saad (2012)
18	Omagua	Tupian	omg	Michael and O'Hagan (2016)
19	Candoshi Shapra	Isolate	cbu	Anderson and Wise (1963)
20	Urarina	Isolate	ura	Olawsky (2006)
21	Kokama	Tupian	cod	Vallejos Yopán (2011)
22	Chamicuro	Arawakan	ccc	Parker (2010)
23	Aguaruna	Jivaroan	agr	Overall (2007)
24	Jebero	Cahuapanan	jeb	Valenzuela (2012)
25	Chayahuita	Cahuapanan	cbt	Rojas Berscia (2015)
26	Muniche	Isolate	myr	Michael et al. (2009, 2013); Michael p.c.
27	Capanahua	Panoan	kaq	Loos (1969); Loos and Loos (2003)
28	San Martin Q	Quechuan	qvs	Coombs et al. (1976)
29	Cajamarca Q	Quechuan	qvc	Quesada (1976)

Table 3.5.1 Continued

No.	Name	Affiliation	ISO	Main source(s)
30	Shipibo	Panoan	shp	Valenzuela (2003)
31	Panobo	Panoan	pno	Gomes (2010)
32	Shanenawa	Panoan	swo	Cândido (2004)
33	Cashibo	Panoan	cbr	Zariquiey Biondi (2011)
34	Cholón	Hibito-Cholon	cht	Alexander-Bakkerus (2005)
35	Ucayali-Yurúa Ash	Arawakan	cpb	García Salazar (1993)
36	Huallaga Q	Quechuan	qub	Weber (1989)
37	Ajy Apurucayali	Arawakan	cpc	Payne (1981)
38	Yaminahua	Panoan	yaa	Faust and Loos (2002)
39	Amahuaca	Panoan	amc	Osborn (1948); Hyde (1980); Sparing-Chávez (2012)
40	Pichis Ash	Arawakan	cpu	Payne (1989)
41	Yanesha	Arawakan	ame	Duff-Tripp (1997)
42	Ashéninka	Arawakan	prq	Mihas (2010)
43	Yine	Arawakan	pib	Hanson (2010)
44	Caquinte	Arawakan	cot	Swift (1988)
45	Nomatsiguenga	Arawakan	not	Shaver (1996)
46	Ese ejja	Tacanan	ese	Vuillermet (2012); Vuillermet p.c.
47	Nanti	Arawakan	cox	Michael (2008)
48	Chácobo	Panoan	cao	Córdoba et al. (2012)
49	Machiguenga	Arawakan	mcb	Snell (1978, 1998)
50	Itene	Chapacuran	ite	Angenot-de-Lima (2002)
51	Araona	Tacanan	aro	Emkow (2006, 2012)
52	Iñapari	Arawakan	inp	Parker (1995)
53	Amarakaeri	Harakmbut	amr	Helberg Chávez (1984)
54	Itonama	Isolate	ito	Crevels (2012a)
55	Jaqaru	Aymaran	jqr	Hardman (1983, 2000)
56	Baure	Arawakan	brg	Danielsen (2007)
57	Cayubaba	Isolate	cyb	Crevels and Muysken (2012)
58	Cavineña	Tacanan	cav	Guillaume (2008)
59	Tacana	Tacanan	tna	Ottaviano and Ottaviano (1965)
60	Movima	Isolate	mzp	Haude (2006)
61	Cuzco Q	Quechuan	quz	Lefebvre and Muysken (1988); Cusihuamán Gutiérrez (2001)
62	Mosetén	Mosetenan	cas	Sakel (2004)

Table 3.5.1 Continued

No.	Name	Affiliation	ISO	Main source(s)
63	Reyesano	Tacanan	rey	Guillaume (2012)
64	Leco	Isolate	lec	Kerke (2009)
65	Ignaciano	Arawakan	ign	Ott and Ott (1983); Olza Zubiri et al. (2004)
66	Trinitario	Arawakan	trn	Rose (2014)
67	Sirionó	Tupian	srq	Firestone (1965); Priest and Priest (1965); Gasparini (2012, p.c.)
68	Callawaya	Mixed	caw	Muysken (2009)
69	Uru	Uru-Chipaya	ure	Hannss (2008)
70	Yurakaré	Isolate	yuz	Van Gijn (2006)
71	Yuki	Tupian	yuq	Villafañe (2004)
72	Aymara	Aymaran	ayr	Hardman (2001)
73	Southern Aymara	Aymaran	ayc	Coler-Thayer (2010)
74	Chipaya	Uru-Chipaya	cap	Cerrón-Palomino (2006)
75	Canichana	Isolate	caz	Crevels (2012b)
76	Bolivian Q	Quechuan	quh	Plaza (2009)
77	East Bolivian Guaraní	Tupian	gui	Dietrich (1986)

[a]We use the ending *–an* to refer to language families, such as Quechuan and Tucanoan. Q stands for Quechua.

present an incorrect picture. There have been attempts, such as Eriksen (2011), to map the precise locations of all languages at the time of contact with the Spanish and Portuguese invaders. We have chosen to use present locations for several reasons. First, the information available for the contact period is not always complete. Second, that is also just a snapshot of a specific moment. Ethnicities would have been moving constantly in the pre-Columbian past as well, and we cannot say what was the relevant precise moment for changes to have taken place. Needless to say, however, more focused micro-studies of sub-regions of the area surveyed here are urgently needed, with the largest possible time-depth, taking demographic, ecological, cultural, archaeological and ethno-historical data into account. Such studies may help explain specific sub-patterns within the overall patterns we focus on in this chapter.

Features studied

The methodology used in this chapter analyses a list of individual properties of language structure (in the sound system, word structure, and sentence syntax). Each

property is 'coded' as a binary opposition, that is, either present or absent in each individual language in our sample – or in some cases, as a three-way opposition. The codes are assigned by analysing published language descriptions, and in exceptional cases on our own field notes. Most sources are modern comprehensive grammars (for example, Sakel 2004; Overall 2007; Guillaume 2008; Zariquiey Biondi 2011), but in a few cases we had to resort to older and/or less comprehensive descriptions. Sometimes this coding is fairly straightforward, as in 'does language X have a central high vowel?', but sometimes it is fairly complex, as in 'does the adjective follow or precede the noun?'. The reason is that all languages have vowels, but not all have adjectives in exactly the same way, and adjectives may precede *and* follow the noun, as in Spanish (for example, *un gran amigo* but *una casa grande)*. The data are sometimes hard to interpret, then; also, data are sometimes simply lacking.

Any study that is based on comparing structural features has to select those features on the basis of a certain rationale. The underlying principle in this chapter is to consider features that have already been proposed by various authors as either typical of Amazonia or of the Andes, and therefore attesting to convergent processes at play right across each region. This approach, and the justification of the features, is discussed more extensively in Van Gijn (2014b), so for this chapter we confine ourselves to mentioning the sources and briefly describing the features.

Table 3.5.2 describes the linguistic overview studies of the Andean and Amazonian regions that are the sources consulted in drawing up our list of features. It lists the source reference in the first column, an abbreviation code by which we refer to those publications hereafter, a brief description of the feature, and the macro-area (Andean or Amazonian) to which it applies.Table 3.5.3 lists the 23 structural features coded for all languages in our sample.

Results and discussion

Figure 3.5.2 summarizes the degrees of difference between all languages with respect to all features in this section of the chapter in the form of a NeighbourNet graph (Bryant and Moulton 2004).[5] The three best represented families are additionally indicated by a square (Quechuan), circle (Arawakan), or a rhombus (Panoan). The languages taken together roughly divide into three groups, which can be characterized areally:

1. An Andean subgroup, which contains all the Quechuan and Aymaran languages, as well as – more distantly – the Uru-Chipaya languages, the Tacanan languages, Jebero (Cahuapanan), and the isolates Candoshi and Leco.
2. A northern upper Amazon subgroup, bringing together all Panoan, Jivaroan and Tucanoan languages in our sample, the northern Tupí-Guaraní languages Kokama and Omagua, the other Cahuapanan language Chayahuita, and the northern (semi-)isolates Cofán, Waorani, Taushiro, Yagua and Urarina. Two

Table 3.5.2 Survey of linguistic studies of the Andean and Amazonian areas. © Rik van Gijn and Pieter Muysken.

Source	Code	Description	Area
Büttner (1983)	B	Comparison of languages from the central Andes in lexis, and in broad typological features in phonology and language structure.	AND
Derbyshire and Pullum (1986)	DP	Survey of a number of morphosyntactic 'areal typological similarities', based on a sample of 20 languages.	AMZ
Derbyshire (1987)	D	Report based on a sample of 40 languages, which reconfirms some of the Amazonian features mentioned in DP.	AMZ
Payne (1990)	P1	Survey of morphological characteristics for a sample of selected Amazonian languages.	AMZ
Dixon and Aikhenvald (1999)	DA	List of features encountered across families in the whole of Amazonia.	AMZ
Payne (2001)	P2	Review of Dixon and Aikhenvald which criticizes their list of Amazonian features and proposes a number of additional ones.	AMZ
Torero (2002)	T	List of 40 features for the central Andean area, ranging from northern Peru to north-east Argentina and Chile; includes proto-languages and extinct language data; also includes some data from languages of the foothills.	AND
Adelaar (2012a; 2012b)	A	Overview of the language situation in the central Andes, focusing on structural and lexical traits of the Aymaran and Quechuan language families.	AND

unexpected languages in the 'northern' cluster are Amarakaeri (Harakmbut) and Mosetén (Mosetenan).[6]

3. A southern upper Amazon subgroup, with all Arawakan languages, the southern Tupí-Guaraní languages Sirionó, Yuki and east Bolivian Guaraní, Chapacuran Itene, and the southern and central (semi-)isolates Cholón, Itonama, Cayubaba, Movima, Yurakaré and Canichana. Surprising languages in the southern cluster are Zaparoan Arabela and Záparo, and the isolate Muniche.

The general picture that emerges is one of areal contact-induced convergence effects, as well as genealogical relatedness in language families. Contact effects can arguably account for the closeness of Tacanan languages to Uru-Chipaya languages, as well as that of Urarina, Leco and Jebero to the Quechuan and Aymaran

Table 3.5.3 Linguistic features studied in this chapter. © Rik van Gijn and Pieter Muysken.

	Feature	amz	and
1	Phonemic central high vowel	Y	N
2	Phonemic mid vowels	Y	N
3	Phonemic nasal vowels	Y	N
4	Phonemic palatal nasal consonant	N	Y
5	Phonemic velar-uvular opposition for stops	N	Y
6	Phonemic retroflex affricates	N	Y
7	More phonemic affricates than fricatives	Y	N
8	Single liquid phoneme	Y	N
9	Proportion of consonants permitted in syllable coda	A	C[a]
10	Presence of morphophonemic nasal spread	Y	N
11	Presence of phonemic glottalized stops	N	Y
12	Presence of phonemic aspirated stops	N	Y
13	Presence of prefixes	Y	N
14	Identical markers of possessor and of core verbal arguments	Y	N
15	Elaborate case-marking system	A	C[b]
16	Presence of core case markers (ERG, ABS, NOM, ACC)	N	Y
17	Accusative alignment in simple clauses	N	Y
18	Dependent marking for possession	N	Y
19	Presence of noun class or gender systems	Y	N
20	Object before subject in basic main clause constituent order	Y	N
21	Basic adjective-noun order within the noun phrase	N	Y
22	Presence of indigenous numerals higher than 9[c]	N	Y
23	Presence of an ideophone word class	Y	N

[a] Three-way distinction based on the percentage of phoneme consonants that can occur in coda position, ranging from 0 to 100, divided into three groups: A: 0–30, B: 31–60, C: 61–100.
[b] Three-way distinction (A) small set of case markers or no case marking (0–4), (B) medium set of case markers (5–6), large set of case markers (>6)
[c] Not counting obvious loans from Spanish, Portuguese, Quechua, or Aymara.

languages. The split of the Tupí-Guaraní languages between northern (Kokama, Omagua) and southern (Sirionó, Yuki, East Bolivian Guaraní) is also suggestive of contact effects, as is the presence of the southern and northern isolate languages in the southern and northern clusters, respectively. Areal effects seem nonetheless outweighed by language genealogy (inherited structures from a common ancestor), across most major families – Arawakan (except Chamicuro [ccc]), Panoan, Quechuan, but also smaller families like Jivaroan [jiv, agr, acu] and Aymaran [jqr, ayr, ayc] – since each of these clusters relatively homogeneously.

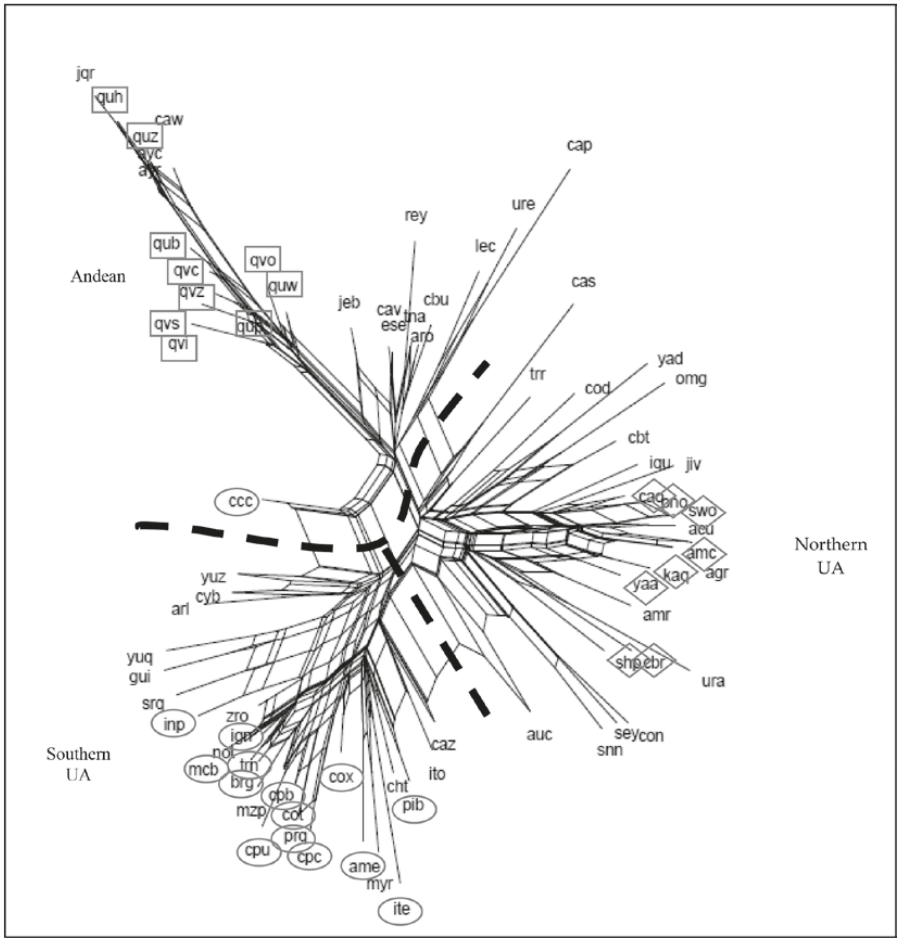

Figure 3.5.3 Neighbour-Net of typological differences between all sample languages (all features). © Rik van Gijn and Pieter Muysken.

The areal effects suggested by Figure 3.5.3 call for a closer look. In the remainder of this chapter, we concentrate on the distribution of individual features: in phonology (that is, the sound system – see section on 'phonological features', below), morphology (that is, word structure, see 'morphological features'), syntax (that is, clause structure, see 'syntactic features') and lexis ('lexical features').

Phonological features

Figure 3.5.4 shows the approximate geographical distributions of the four features to do with vowels. The x-axis in each of the plots shows latitude from south (left) to north (right); the y-axis shows elevation from low (bottom) to high (top).
The first vowel feature is whether each language has a central high vowel – a sound intermediate between Spanish /i/ and /u/. As can be seen, the central high vowel is

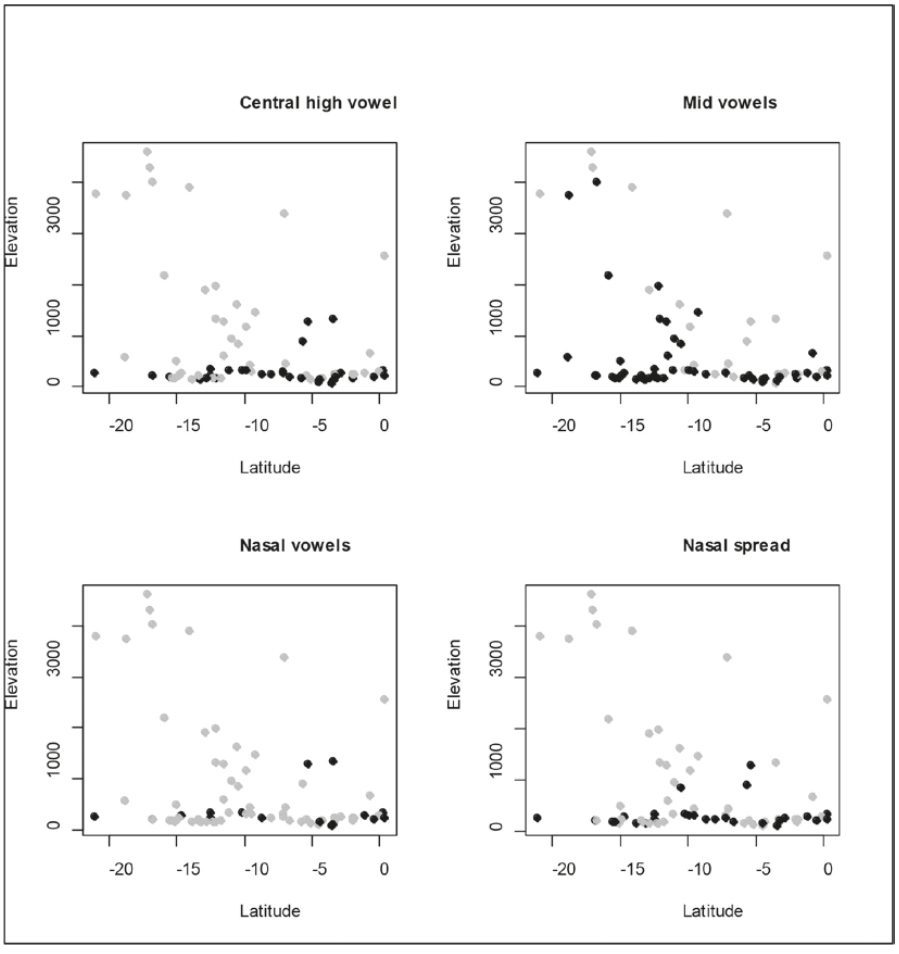

Figure 3.5.4 Distribution of four vowel features by latitude and elevation in the languages of the Andes and upper Amazonia. © Rik van Gijn and Pieter Muysken.

clearly a lowland feature: all the black dots, for languages that do have the feature in question, are at low elevations; the dots for the highland languages, meanwhile, are all grey, showing that they do not have this feature. The central high vowel is found over the entire north–south span of the upper Amazon (though it is slightly less frequent in the south). The three languages spoken at slightly higher altitudes and that also have a central high vowel are Chayahuita (of the Cahuapanan family), and Shuar and Aguaruna (both languages of the Jivaroan family; the third Jivaroan language in the sample, Achuar in the lowlands, also has a high central vowel). Both the Cahuapanan and Jivaroan territories stretch from higher altitudes eastwards to lower altitudes. Nonetheless, there are also many lowland languages that do not have the central high vowel. Interestingly, although it is assumed that proto-Arawakan did have a central high vowel (Aikhenvald 1999, 76), most

modern Arawakan languages in the sample do not (Baure, Ignaciano, Trinitario, Ashéninka, Nomatsiguenga, Ucayali Yurúa Ashéninka, Nanti, Machiguenga, Yanesha'), implying that it must have been lost, perhaps under the influence of highland contact.[7] Furthermore, the central high vowel is not found in any of the Tacanan languages, suggesting that their common ancestor did not have it either. Alternatively – given the putative deep genealogical connection with the Panoan languages, which do generally have the high central vowel – this phoneme was perhaps lost before the Tacanan languages dispersed.

Mid vowels are pronounced with the tongue at a mid height in the mouth, for example, /e/- and /o/-type vowels, rather than 'high' /i/ and /u/, or 'low' /a/. In the upper Amazon, mid vowels show a less clear-cut pattern by elevation: they seem almost omnipresent in the lowlands, but are certainly found at higher altitudes as well, notably in the Uru-Chipaya languages, in some of the higher Campan Arawakan languages (Nanti, Matchiguenga, Nomatsiguenga, Ashéninka Perené, Pichis Ashéninka, Caquinte),[8] and in some of the (semi-)isolates spoken at higher altitudes (Kallawaya, Cholón, Leco, Canichana). Tena Quechua, one of the lowland Quechuan languages, has also developed phonemic mid vowels (unlike most highland Quechua varieties). This distribution suggests an important role for genealogy, since there are very few clear examples of mid vowels being acquired (other than in unadapted loanwords), while they were perhaps lost (and both low and high vowels were retained) in some of the Arawakan languages, such as Yanesha and Ajyíninka Apurucayali. The same important role for genealogy can be observed in the lowlands of the central upper Amazon, where Panoan languages generally do not have mid vowels.

Phonemic nasal vowels (Figure 3.5.5) are independent vowels of the same general type as those pronounced in French *un bon vin blanc* (where the written <n> is no longer pronounced as a consonant n at all), or written with a tilde as in Portuguese *São Paulo*. Nasal spread refers to a more automatic process in some languages, where one or more of the vowels in a word acquires a nasal pronunciation automatically, if that word also contains a nasal consonant (*n*, *m*, and so on). Taking these together (that is, whether phonemic or not), nasal vowels seem to be a clear lowland feature in the sense that they are hardly ever found in the highlands (except in Jivaroan languages) – although that does not mean that they are omnipresent in the lowlands. In particular, phonemic nasal vowels seem relatively rare, and concentrated mostly in the northern upper Amazon, which thus potentially constitutes a minor areal pattern spanning the Tucanoan languages Secoya and Siona and the isolates Cofán and Waorani, concentrated along the Aguarico River (the northernmost group on Figure 3.5.5).

Nasal spread is more common, and may follow areal patterns, expanding along rivers: the Aguarico/Napo in the north, Marañón in northern Peru, Ucayali in central Peru, and the Mamoré in Bolivia (see Figure 3.5.6).

To summarize, the central high vowel and nasal vowels seem to be lowland features, and the range of the latter especially (phonemic or not) seems to have

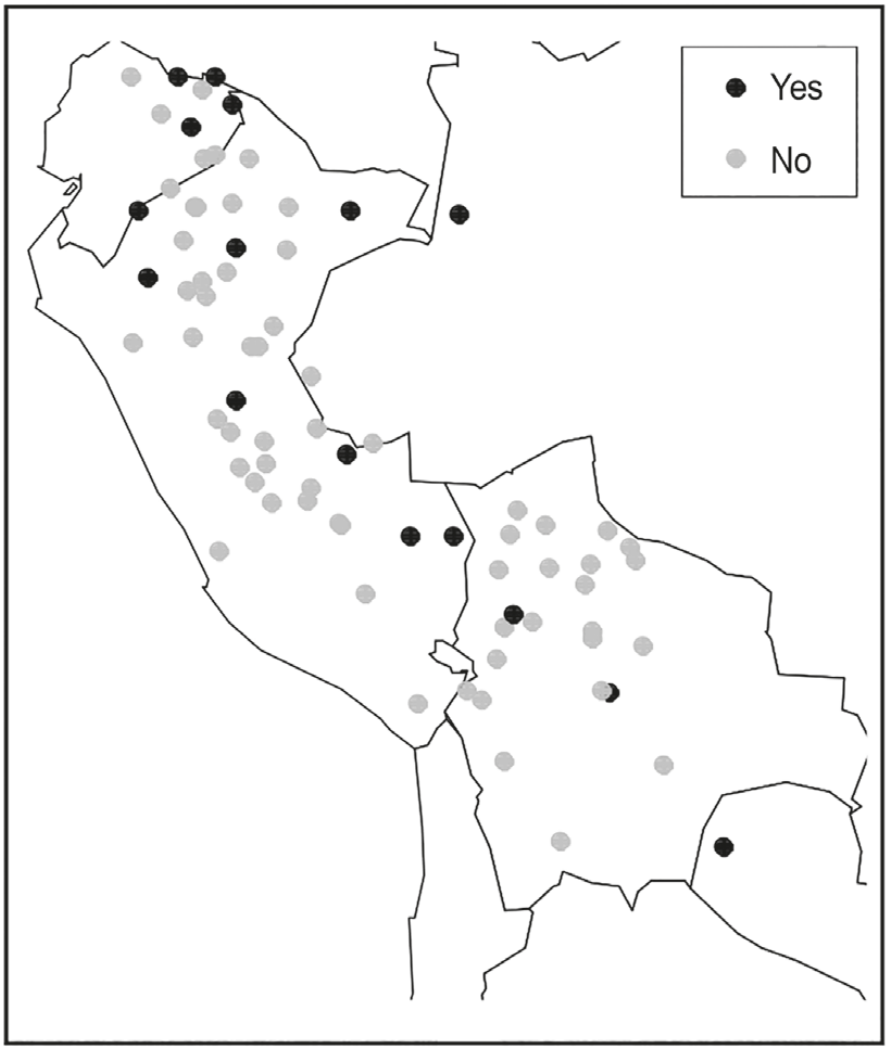

Figure 3.5.5 Map showing the presence or absence of nasal vowels. © Rik van Gijn and Pieter Muysken.

been expanding, possibly through contact. The presence of mid vowels seems to be determined mainly by language affiliation, but their absence may be a contact effect, especially in some of the Campan languages.

Moving on to the consonant features, Figure 3.5.7 shows the geographical distributions of the presence of:

1. a phonemic palatal nasal (the sound spelt <ñ> in Spanish, and <nh> in Portuguese);
2. a retroflex affricate (that is, a sound of the type spelt <ch> in Spanish and English, but pronounced retroflex, with the tongue curled back);

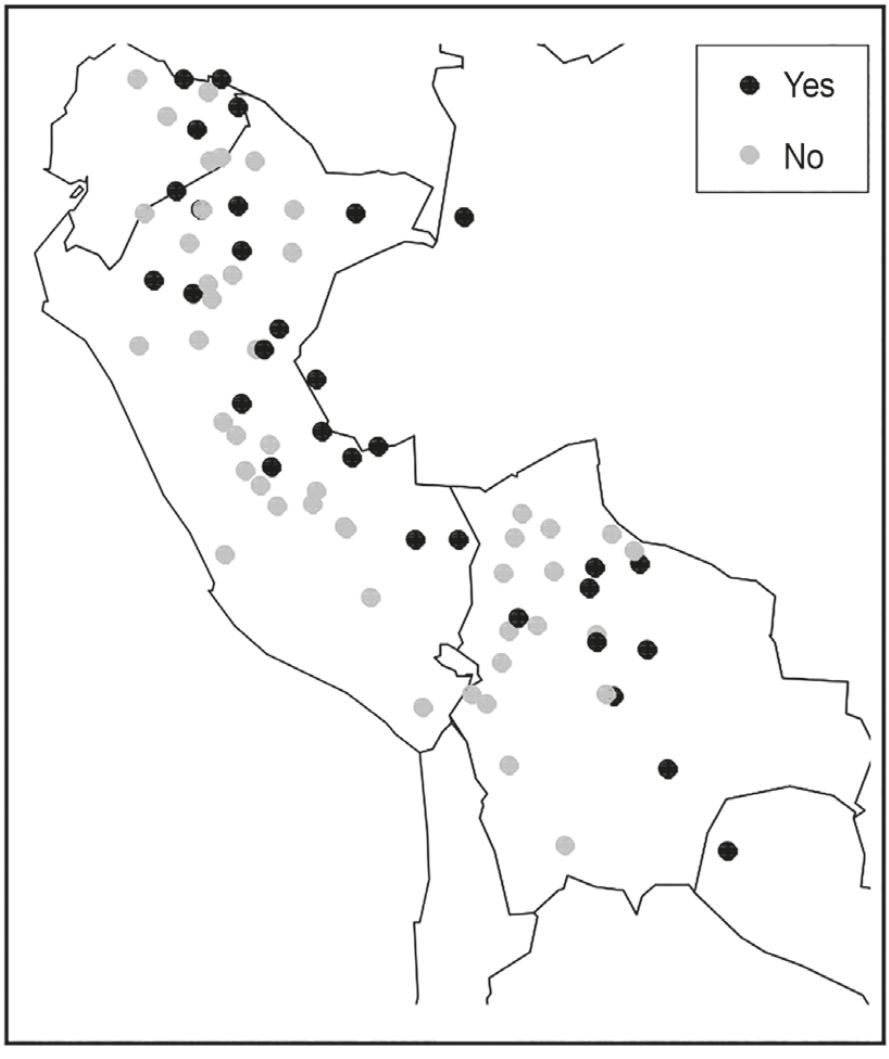

Figure 3.5.6 Map showing the presence or absence of nasal spread. © Rik van Gijn and Pieter Muysken.

3. more affricate than fricative phonemes (that is, more sounds of the type spelt <ch>, <dg> or <ts> in English, than of the type spelt <sh>, <z>, <s>, <th>, <f>, etc.);
4. only a single liquid phoneme (that is, not both *r* and *l* sounds, but just one, undifferentiated *r/l*).

The distribution of the palatal nasal may have areal dimensions, as it occurs in the Aguarico, Santiago and Marañón areas, as well as in the upper Ucayali, Madre de Dios, and part of the Mamoré. More broadly speaking, and especially in the central

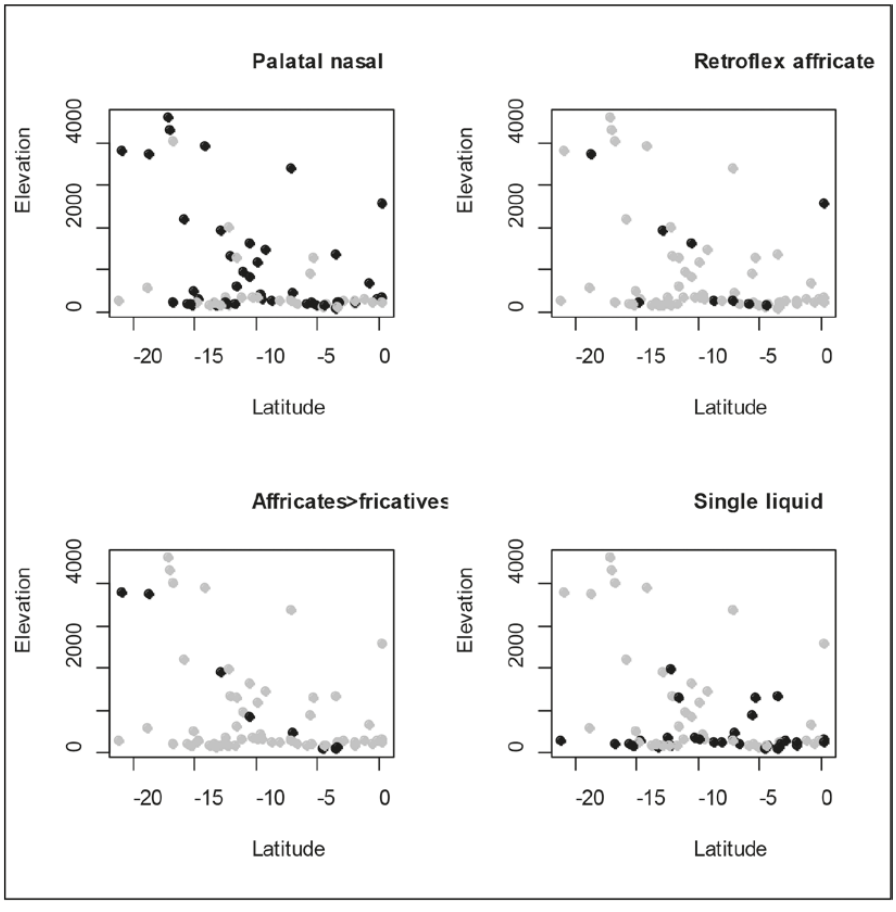

Figure 3.5.7 Distribution of the presence or absence of a palatal nasal by latitude and elevation in the languages of the Andes and upper Amazonia. © Rik van Gijn and Pieter Muysken.

upper Amazon, languages closer to the Andes more often have a phonemic palatal nasal than those further east, which again may point towards highland–lowland interactions (see Figure 3.5.8).

The retroflex affricate and cases of affricates outnumbering fricative phonemes, are rare in the entire area, as well as in the adjacent Andean languages. They do not seem to be particularly associated with either the highlands or lowlands, nor with particular river systems or sub-areas in the upper Amazon.[9] In fact, it is rather surprising to find the retroflex affricate in so many lowland languages (Urarina, Muniche, Cashibo, Shipibo, Reyesano), and to find affricates outnumbering fricatives in highland languages (Bolivian Quechua, Chipaya, Jaqaru). Just a single liquid phoneme, meanwhile, seems to be a lowland rather than a highland feature, although it is also found in some scattered lowland languages, with potential diffusion areas in northern Peru and central Bolivia in particular.

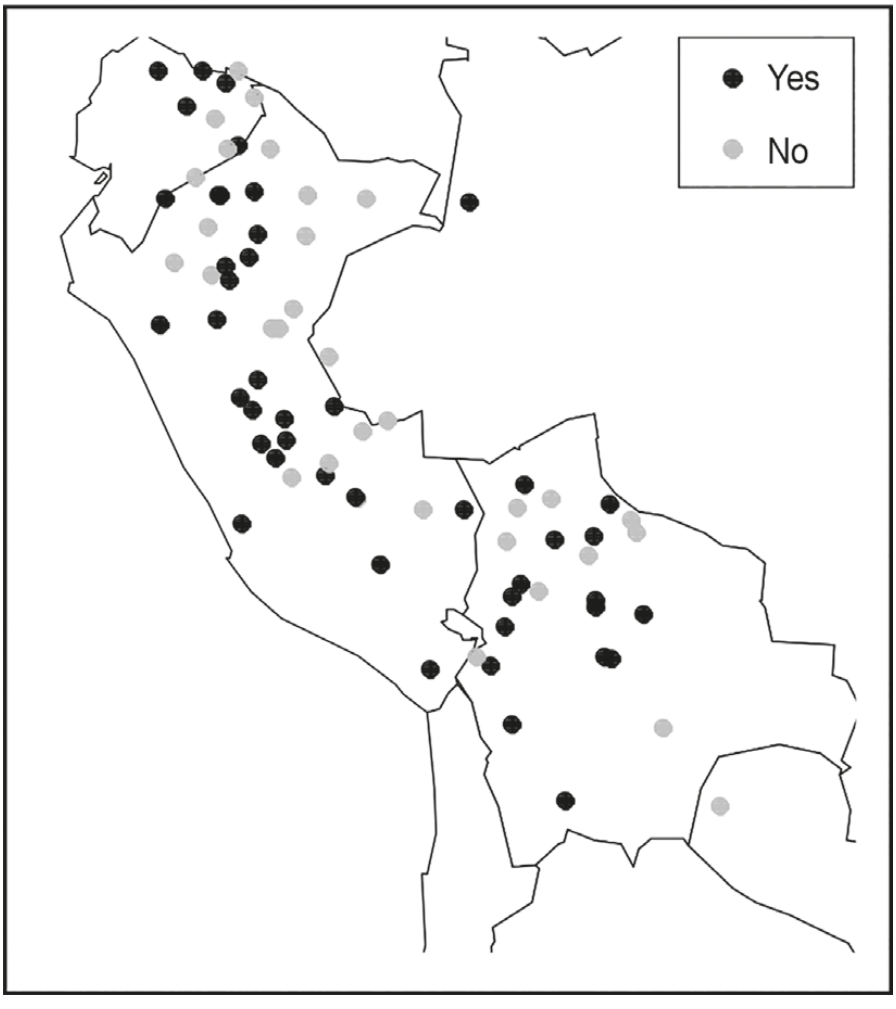

Figure 3.5.8 Distribution of four consonantal features by latitude and elevation. © Rik van Gijn and Pieter Muysken.

These consonant features, in other words, do not pattern clearly by elevation; the palatal nasal and a single liquid phoneme show distributions that may be connected to river-based expansions.

Figure 3.5.9 shows three features related to the pronunciations of stop (or 'plosive') consonants (that is, those of the type /p/, /t/, /k/ and /b/, /d/, /g/). All three stop features have been associated, in published areal studies, with the Andes, or perhaps more narrowly with the Quechuan and Aymaran families. These features are whether a language has distinctions between:

1. velar versus uvular stops, that is, the contrast between sounds spelt <k> and <q> respectively, in modern indigenous orthographies for Quechua

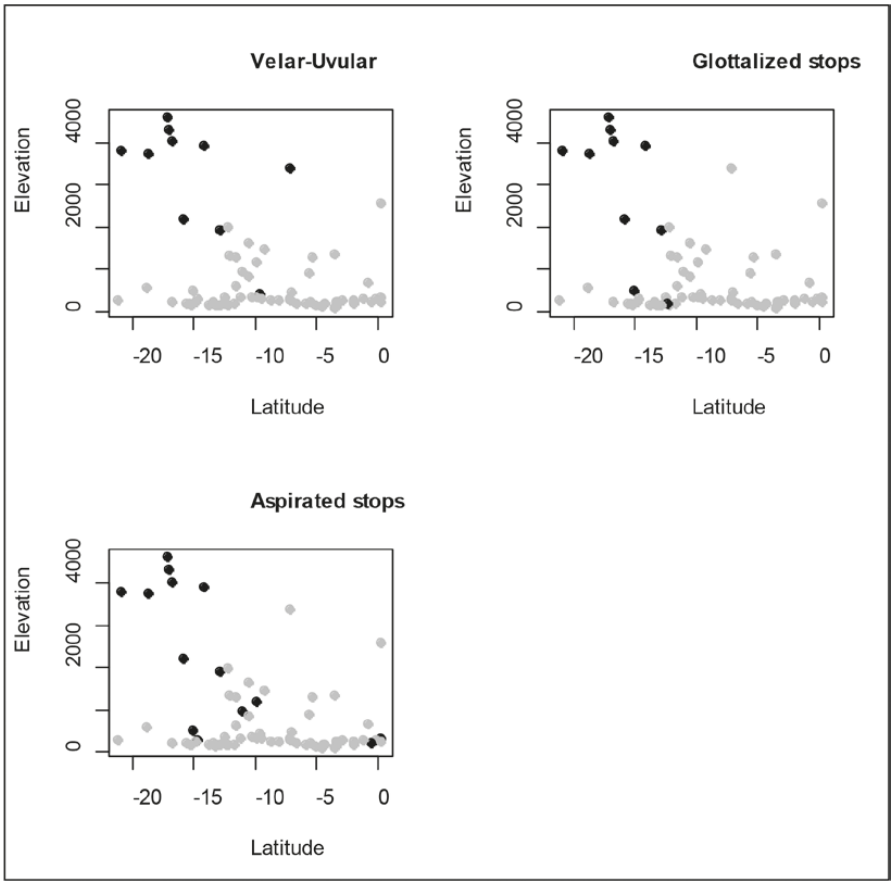

Figure 3.5.9 Distribution of three stop features by latitude and elevation.
© Rik van Gijn and Pieter Muysken.

and Aymara (for example, the k in *piki* and the q in *llaqta*, in the name of the well-known archaeological site of Pi̲killaq̲ta, 'flea town');

2. normal versus glottalized stops (the latter spelt with an apostrophe, for example, in P'i̲saq);
3. normal versus aspirated stops (the latter spelt with a following <h>, for example, *khipu*).

All three stop features are fully present in the Bolivian and south Peruvian highlands (in the Quechuan,[10] Aymaran and Uru-Chipaya families), but glottalized and aspirated stops are lacking in the more northerly Quechuan varieties of Imbabura, San Martín, Napo and Cajamarca Quechua. The velar-uvular distinction has also been lost in Imbabura, Napo and San Martín Quechuas. Leco (an isolate) also has glottalized as well as aspirated stops, undoubtedly under the influence of a Southern Quechuan and/or Aymaran language. Itene (Chapacuran)

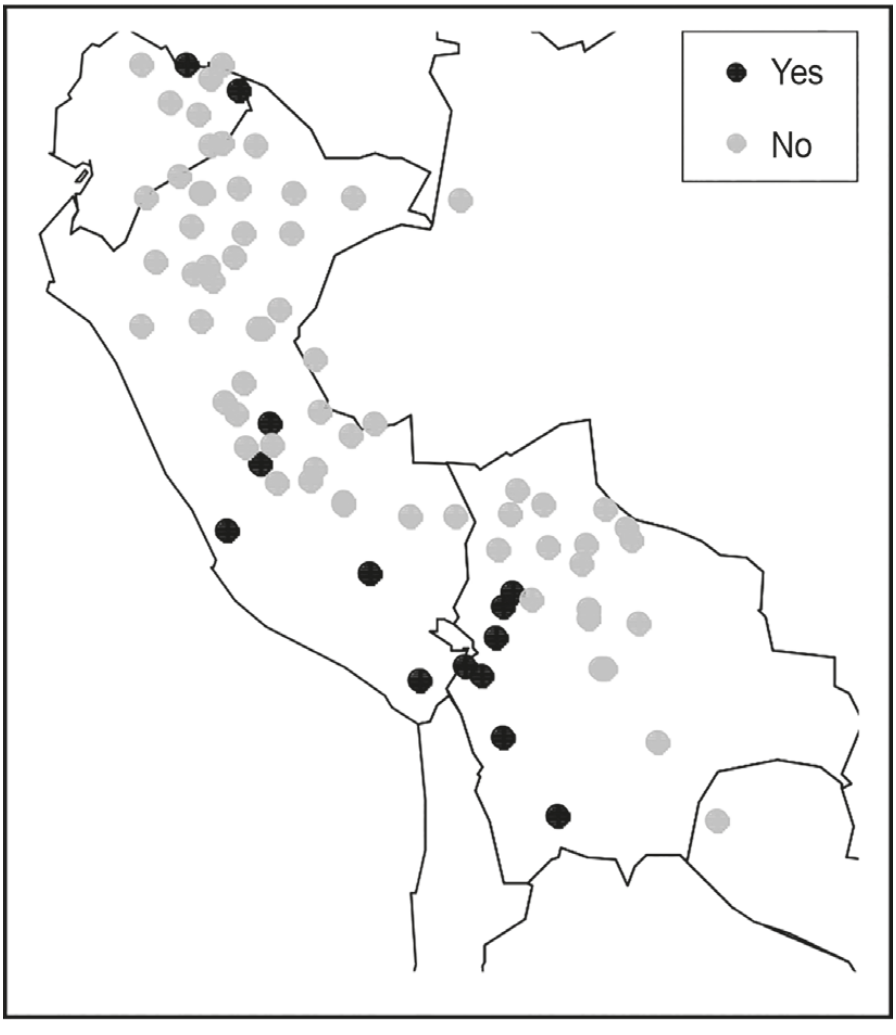

Figure 3.5.10 Map showing the presence or absence of aspirated stops.
© Rik van Gijn and Pieter Muysken.

has pre-glottalized stops, which seem unrelated to the Andean type of glottalized stops, given both the geographical distance and their contrasting phonetic realizations. Secoya (Tucanoan), Mosetén (Mosetenan), Cofán (isolate), the Arawakan languages Ashéninka Perené and Ajyíninka Apurucayali, as well as the isolate Leco, are all lowland languages that do have aspirated stops. There are two regions in particular – around Lake Titicaca, and also in central Peru – that seem to be diffusion areas for aspirated stops (Figure 3.5.10): they came into Quechua from Aymara, and seemed to have expanded eastward into the lowlands.

There is some leakage of these typical Andean stop features into languages of the foothills. In particular, aspirated stops seem to have diffused to languages

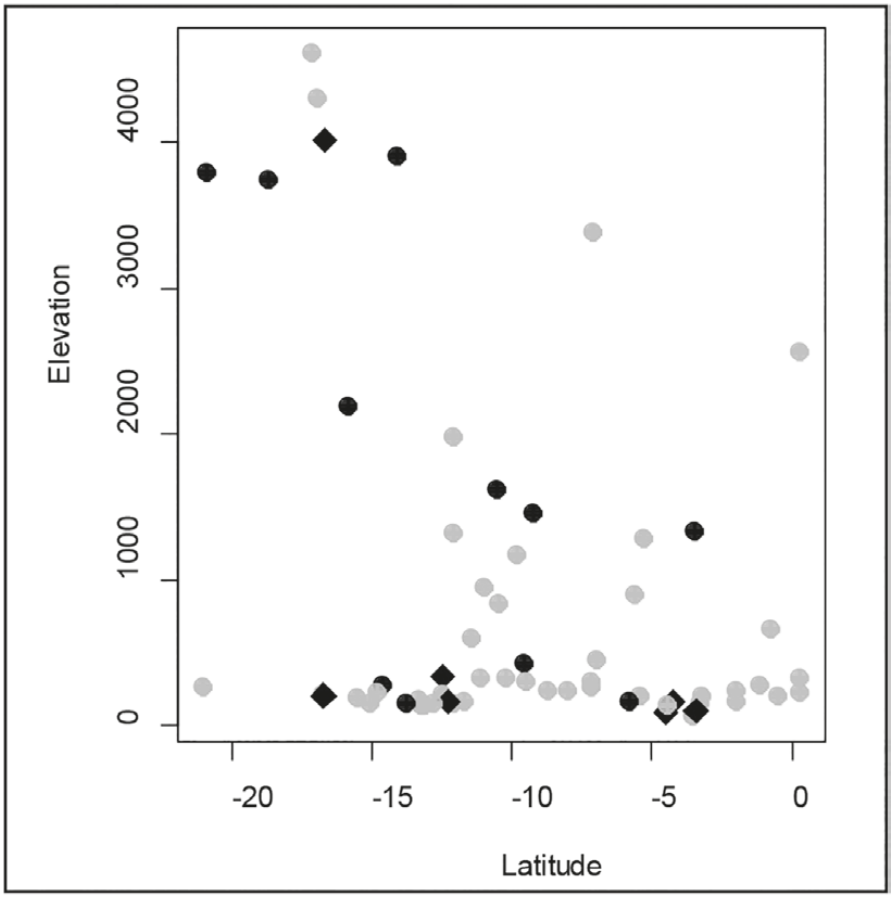

Figure 3.5.11 Distribution of closed syllables by latitude and elevation.

spoken at lower altitudes. At the same time, many northern Quechua varieties in Ecuador (although not Imbabura) surprisingly do have the aspirated/non-aspirated distinction in stops (but not the glottalized/non-glottalized contrast), possibly due to Cuzco Quechua adstrate or superstrate[11] in the Inca period.

As a final illustrative feature in phonology, Figure 3.5.11 looks at closed syllables, that is, those that do not end in a vowel, but have a consonant immediately following it. Specifically, Figure 3.5.11 asks what proportion of the consonant phonemes in a language are permitted in (underlying) coda position, that is, at the end of a syllable, after the vowel (for example, the two /n/ sounds in English *London*). The grey circles are languages with the most restrictions, those that allow less than a third of their consonants to stand in coda position. The black circles are languages with the least restrictions, allowing over two-thirds of consonants in codas; and the black diamonds are the intermediate cases. Of the highland languages, southern

and central Quechuan varieties generally do allow many of their consonants to stand in the syllable coda, as do Uru-Chipaya languages (although with slightly more restrictions in Uru). Aymaran languages, however, have more restrictions, at least underlyingly (that is, before suffix combination rules allow some vowels to be dropped), as do the northern Quechuan languages. Other languages at mid-elevations that put few restrictions on the coda are Yanesha', Shuar, Callawaya and Cholón. Lowland languages with few to intermediate restrictions on the coda are Amarakaeri, Mosetén, Yurakaré, Candoshi, Itene, Muniche, Movima, Yagua and Kokama. The foothill languages Mosetén and Yurakaré, as well as the languages at mid-elevations, may have been influenced by Andean languages.

Morphological features

An important typological characteristic of Andean languages is that they tend to be exclusively suffixing, whereas many Amazonian languages have (person) prefixes. Figure 3.5.12 indicates that although prefixes certainly tend to become less

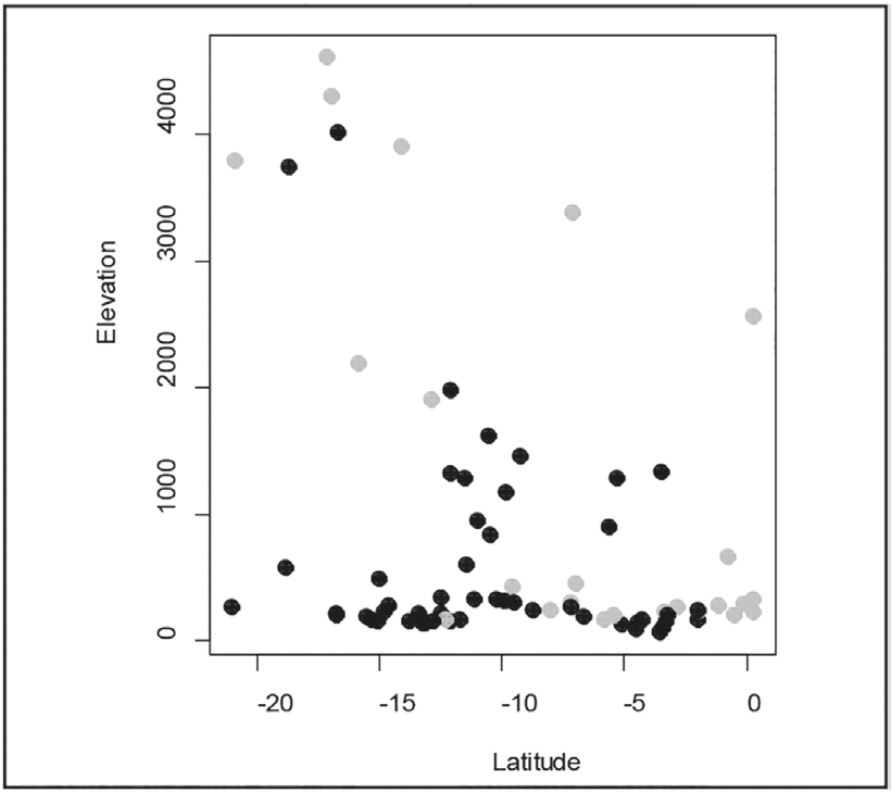

Figure 3.5.12 Distribution of presence of prefixes by latitude and elevation.
© Rik van Gijn and Pieter Muysken.

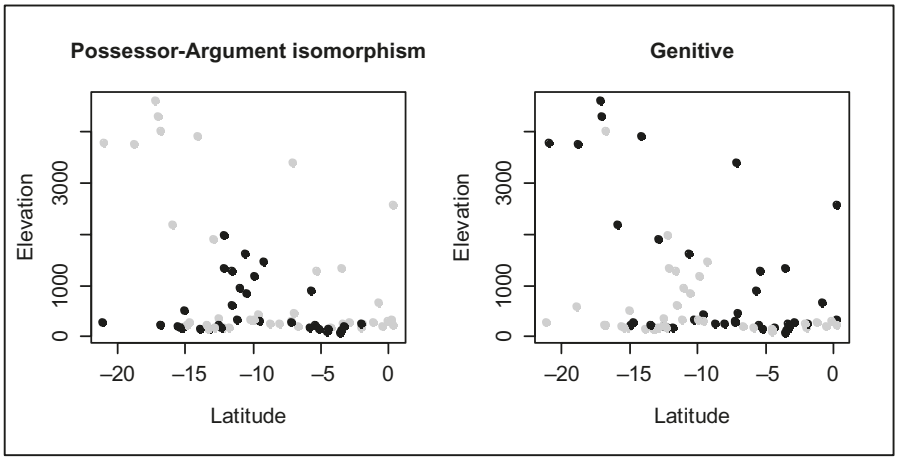

Figure 3.5.13 Distributions of possession-related features by latitude and elevation. © Rik van Gijn and Pieter Muysken.

common as one moves to higher elevations, the transition is not abrupt. Moreover, the far north of the upper Amazon lowlands seems to form a mini-area of exclusively suffixing languages.

The two languages spoken at higher elevations that do have prefixes are the related languages Uru and Chipaya, and both have only a marginal inventory of prefixes. The prefix system probably used to be more elaborate, involving referential (object) prefixes (Cerrón-Palomino 2006, 78–9; Hannss 2008, 133–4), so its current marginal status suggests that contact-induced influence from Quechuan and Aymaran languages has led to this decline in prefixes. The languages spoken at intermediate elevations and that do have prefixes are generally of the Aymaran and Jivaroan families, as well as a number of isolates (Cholón, Canichana, Leco). In the lowlands, the Panoan languages generally have very few or no prefixes, and there are also a few languages lacking prefixes in the northern Napo-Aguarico river system, including Quechuan languages (Imbabura, Napo, northern Pastaza), Tucanoan languages (Siona, Secoya), and isolates (Cofán, Waorani).

Figure 3.5.13 shows two aspects of how languages mark possession. The left-hand chart shows whether languages have bound possessive pronouns (like *my, your, his* in English, but attached to the verb) that are (nearly) identical (isomorphic) to the bound pronouns used for one of the verbal arguments (*very* roughly: is a possessor noun marked in the same way as either a subject or object noun?). This is an Amazonian characteristic, in that the black dots for languages that do show that isomorphism cluster mostly at lower elevations. The right-hand chart shows a more Andean feature, with black dots dominant at higher elevations: does the language have a genitive case marker?

In spite of some black dots at higher altitudes in the left-hand chart, possessor-subject/object isomorphism seems to be fundamentally a lowland rather than

highland feature,[12] since the black dots towards the top of the graph are mostly Arawakan languages (maintaining a feature typical of that family), as well as some of the higher-altitude (near) isolates and representatives of small families like Cholón, Leco and Chayahuita.

The genitive shows almost a mirror image, partly reflecting a more general contrast in languages' structural systems (head-marking versus dependent-marking). Quechuan and Aymaran languages do have a genitive marker (they actually use double marking). The grey dot conspicuous at high altitude is Uru: it does in fact have possessive dependent (case) marking, but only on pronouns (Hannss 2008, 186–7), whereas the diagnostic feature we study is focused on nouns. Nevertheless, possessive case marking may have been more widespread in Uru in the past, pronominal case marking being a remnant of that more encompassing system. Jivaroan languages also have genitive markers, as does Chayahuita. Arawakan languages generally do not, an exception being Yanesha'. Of the lowland languages, those of the Panoan family generally do have a genitive marker (this seems to be a genealogical predisposition) as do those of the Tacanan family. A number of other lowland languages (Mosetén, Chamicuro, Yagua, Iquito, Candoshi and others) also have genitives, so this cannot justifiably be called a highland feature per se. Genitive markers do seem to be relatively rare in the southern upper Amazon, though.

Figure 3.5.14 shows two further features reported in the literature as typically Andean: core case markers (that is, case markers for the obligatory arguments of a verb) on the left-hand side and accusative alignment (a system such as exists in English, where the subject of an intransitive clause – with a single obligatory argument – for example, 'I' in *I walk* – behaves in the same way as the subject of a transitive clause – with two obligatory arguments – for example, 'I' in *I hit him*), in simple main clauses. There do seem to be plenty of lowland languages,

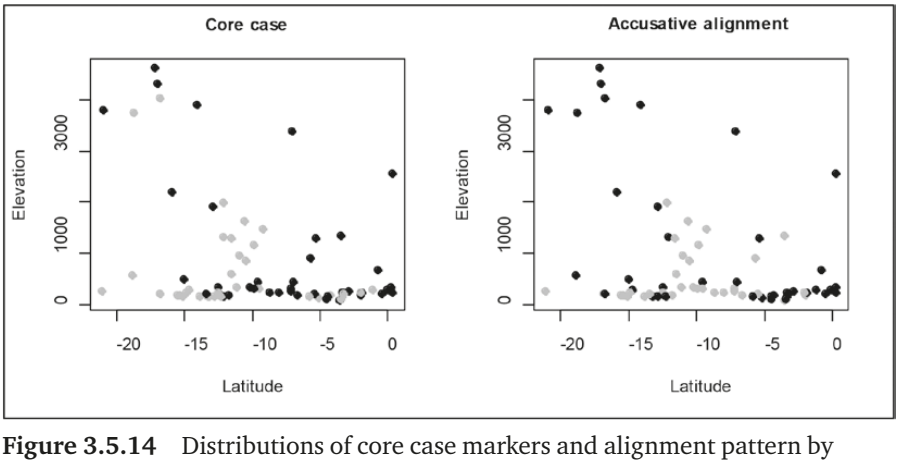

Figure 3.5.14 Distributions of core case markers and alignment pattern by latitude and elevation. © Rik van Gijn and Pieter Muysken.

however, that also have core case markers. Often these are ergative case markers (in the Tacanan and Panoan families), but accusative case markers certainly occur too (in Jivaroan and Tucanoan). Nonetheless, generally speaking, object case markers in lowland languages seem less 'structural', in that they are often subject to conditions, leading to differential object marking (for example, only animate objects are case-marked, see Van Gijn, 2019). The Uru-Chipaya languages do not have core case.

Accusative alignment is found throughout the higher Andes, as well as in some languages at lower altitudes (for example, Amarakaeri, Leco, Yurakaré, Canichana) and especially in the northern upper Amazon (for example, Aguaruna, Waorani, Cofán, Siona, Secoya, Candoshi). The Arawakan family has split S systems, while Tacanan and Panoan have ergative systems.

A final morphological parameter is the number of case markers in a language. Dixon and Aikhenvald (1999) claim that lowland languages generally have small case inventories, but for our sample this seems true mainly for Arawakan and most Tupian languages, as well as for a few (near-)isolates (for example, Movima, Iquito, Itonama, Muniche and Canichana). Otherwise, many lowland languages have extensive case inventories (see Figure 3.5.15). So although it is true that highland languages have extensive case inventories, so too do many lowland languages.

Syntactic features

In languages worldwide, there is an overwhelming universal preference for word orders in which A (transitive subject) comes before O (object) (Dryer and Haspelmath 2013a, 2013b). Exceptionally, however, deviant word orders with O before A have been claimed as areal patterns in parts of central Amazonia (Derbyshire and Pullum 1986). As can be seen in Figure 3.5.16, O before A orders are nonetheless decidedly rare in the upper Amazon. Only Urarina, Itene, Arabela, Sirionó, Yuki, and Reyesano were classified as having O before A.[13] Although it is true that these are all lowland languages, it seems a stretch to consider this an areal feature, given that these languages are so few and far apart.

The order adjective–noun is typical of Andean languages, as corroborated by Figure 3.5.16. However, Figure 3.5.16 also shows that this order is common in the lowlands of the upper Amazon, too. In fact, from a distributional point of view, a number of diffusion areas can be identified, as shown in Figure 3.5.17, where the northern Napo-Aguarico-Pastaza area in Ecuador, as well as the Marañón and the Madre de Dios, contain various languages spoken in contiguous areas that all have adjective–noun order.

To summarize, O before A order is uncommon in general and does not seem to follow any areal pattern. Adjective before noun order is found throughout the Andes, as well as in a number of adjacent areas in Ecuador, northern Peru, and Bolivia.

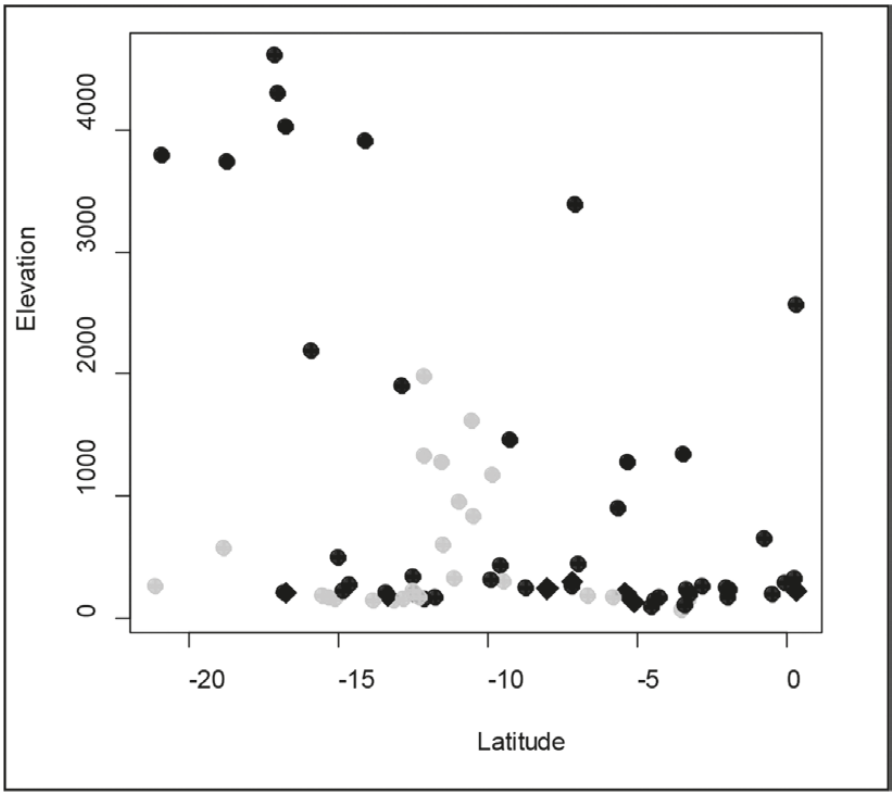

Figure 3.5.15 Distributions of elaborate case inventories by latitude and elevation. © Rik van Gijn and Pieter Muysken.

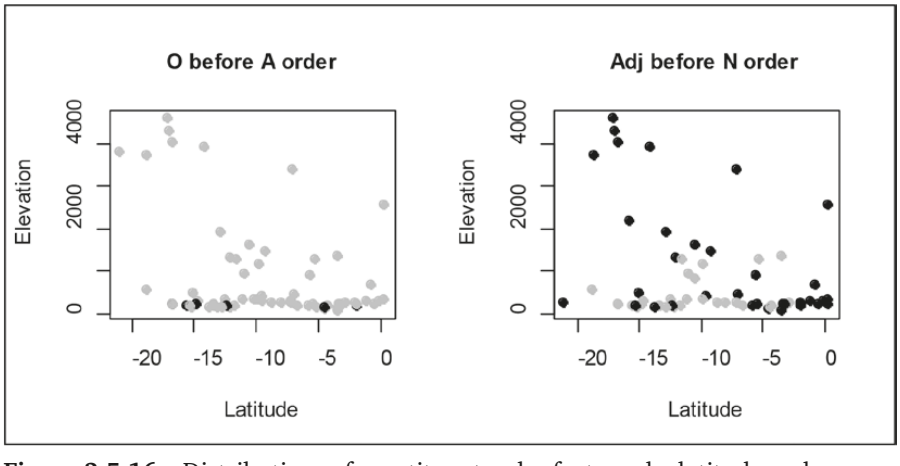

Figure 3.5.16 Distributions of constituent order features by latitude and elevation. © Rik van Gijn and Pieter Muysken.

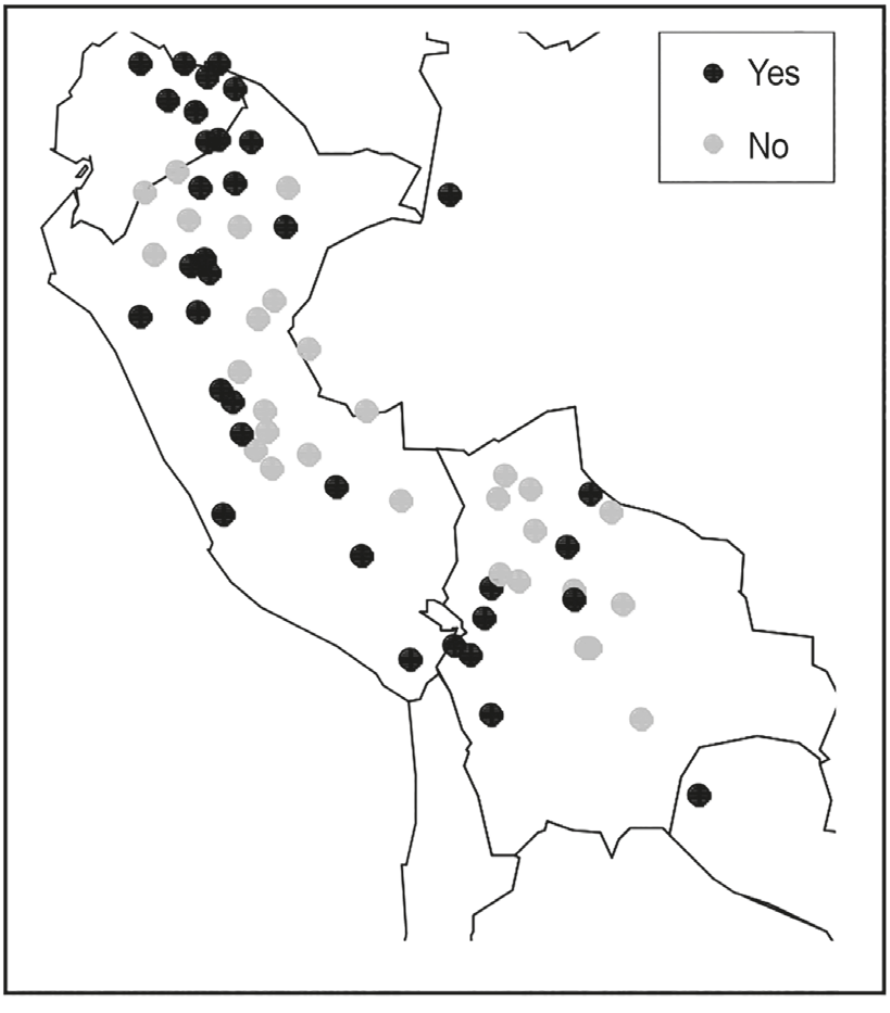

Figure 3.5.17 Map showing adjective-noun order. © Rik van Gijn and Pieter Muysken.

Lexical features

Figure 3.5.18 shows the distribution of three features related to the lexicon: whether languages categorise their nouns into classes or genders; whether they have (native) words for high numerals; and whether they have a clearly distinct word-class of ideophones that behave differently to other nouns.

Noun class or gender systems can be found in Arawakan languages in central Peru, but also in a number of Guaporé-Mamoré isolate languages (Cayubaba, Movima, Itonama, Mosetén), and in north-eastern Peru and Ecuador (Yagua, Muniche, Chayahuita, Omagua, Arabela, Záparo, Cofán, Secoya, Siona). None of the traditional Andean families (Quechuan, Aymaran, Uru-Chipaya) has a noun class system.

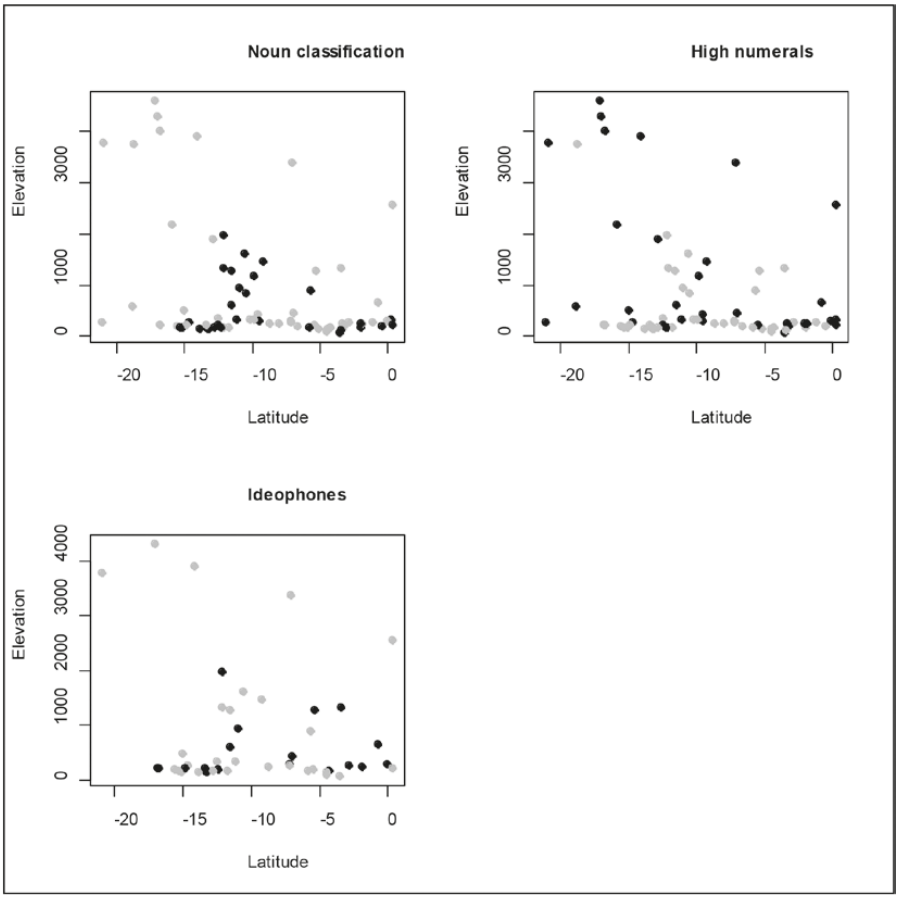

Figure 3.5.18 Distributions of lexical features by latitude and elevation.
© Rik van Gijn and Pieter Muysken.

While Andean languages generally have elaborate numeral systems, Amazonian languages have a reputation for having very small native numeral systems, often no more than just the first two or three numbers. The plot in Figure 3.5.18 counts only native numerals, to the extent that we can establish which words have been borrowed from other sources. That some larger indigenous numeral systems can be found in the lowlands is partly due to Quechuan languages that are intrusive here (southern and northern Pastaza, Napo, Huallaga, San Martín and Tena varieties of Quechua). Nonetheless, a few other (semi-)lowland languages do seem to have native conventionalized numeral systems that go beyond nine: for example, Itene, Taushiro, Mosetén, Cofán, Yine[14], Leco, Cholón.[15] Chipaya has replaced its native numerals above four with Aymaran numerals. Quechuan/Aymaran influence on numeral systems can be observed in several other upper Amazon languages, for example, Urarina, Kokama, Shipibo-Konibo, Yanesha', Cavineña and Chayahuita. Many other lowland languages use Spanish numerals for the higher numbers.

It is hard to say anything general about ideophones. There are many fewer dots than in other graphs, because from the sources available it is often unclear whether a language does not have ideophones. And even with the languages for which a coding decision was taken, there were different degrees of confidence. The general picture seems to be that highland languages lack ideophones, and that they are more common, though not ubiquitous, in the lowlands. Nuckolls has discovered extensive use of ideophones in (lowland) Pastaza Quechua (2001, 71), and has argued that this feature is common in the area where this is spoken, but we know of no systematic survey in this respect, and descriptions are not complete.

Results and discussion

We have surveyed the distribution of selected features in phonology, morphology, syntax and lexis in over 70 languages in the central Andes and adjacent parts of the Amazon. In this final section we return to Dixon and Aikhenvald's (1999) comment that the Amazonian and Andean areas fade into each other, and come to a rather more precise and detailed picture. Table 3.5.4 briefly evaluates the features studied here. Figure 3.5.19 organizes the features in terms of strongly highland (top left) to strongly lowland (bottom right), with the features in bold showing evidence of diffusion from the highlands toward the lowlands; the features between brackets are those that show less clear patterns as a result of low representation of a feature or feature value.

From Table 3.5.4 and Figure 3.5.19 we can conclude that a few features, notably phonological ones, pattern quite clearly along a highland-lowland divide: the vowel features are concentrated in the lowlands, whereas the stop features are predominantly restricted to the highlands. Other lowland features include the presence of prefixes (or rather, the lack of them seems to be a highland feature), isomorphism of markers for possessor and verbal argument, and gender/noun class systems. Other than in phonology, there seem to be few features clearly restricted to the highlands. Accusative alignment is found in the lowlands too, especially in the north, although accusative case-markers in the lowlands do generally seem subject to more conditions than in the highlands. Adjective–noun order is also found in many lowland languages, possibly due to contact in several sub-areas of the upper Amazon. Higher numerals are perhaps the most strongly Andean feature, and Aymaran and Quechuan languages have certainly influenced lowland languages in this respect, for a good many of them have adopted Quechuan or Aymaran numerals. Other reportedly typical highland or lowland features turned out to be either very rare in the sample in any case (retroflex affricates, more affricates than fricatives, O before S order), or common in both highlands and lowlands (palatal nasal, closed syllables, elaborate case inventories, core case marking, genitive marking).

The contact-induced diffusion of more abstract, grammatical features can be indicative of several different contact scenarios (Thomason and Kaufman 1988; Thomason 2001; Muysken 2010):

Table 3.5.4 Summary of linguistic features and their distributions by latitude and elevation. © Rik van Gijn and Pieter Muysken.

	Feature	Distribution pattern
1	Phonemic central high vowel	**Uniquely lowland feature**
2	Phonemic mid vowels	**Lowland feature**
3	Phonemic nasal vowels	**Uniquely lowland feature**, but fairly rare
4	Phonemic palatal nasal consonant	Widespread
5	Phonemic velar-uvular opposition for stops	**Highland feature**
6	Phonemic retroflex affricates	Rare in the sample
7	More phonemic affricates than fricatives	Rare in the sample
8	Single phonemic liquid phoneme	Mostly lowland
9	Permissibility of closed syllables	No clear pattern
10	Presence of morphophonemic nasal spread	**Uniquely lowland feature**
11	Presence of phonemic glottalized stops (Peru, Bolivia)	**Highland feature**
12	Presence of phonemic aspirated stops (Peru, Bolivia)	Mostly highland, some dispersal
13	Presence of prefixes	**Lowland feature**
14	Isomorphism of possessor and core verbal argument person markers	**Lowland feature**
15	Elaborate case marking system	No clear pattern
16	Presence of core case markers (erg, abs, nom, acc)	Widespread
17	Accusative alignment in simple clauses	Mainly highland and northern Upper Amazon lowlands
18	Dependent marking for possession	Fairly common throughout
19	Presence of classifier or gender systems	**Lowland feature**
20	O before S basic main clause constituent order	Rare in the sample
21	Basic adjective-noun order within NP	Highland, with potential diffusion into lowland areas
22	Indigenous numerals higher than nine	Highland and some lowland languages have complex numerals
23	Ideophone word class	Data limited, mostly lowland

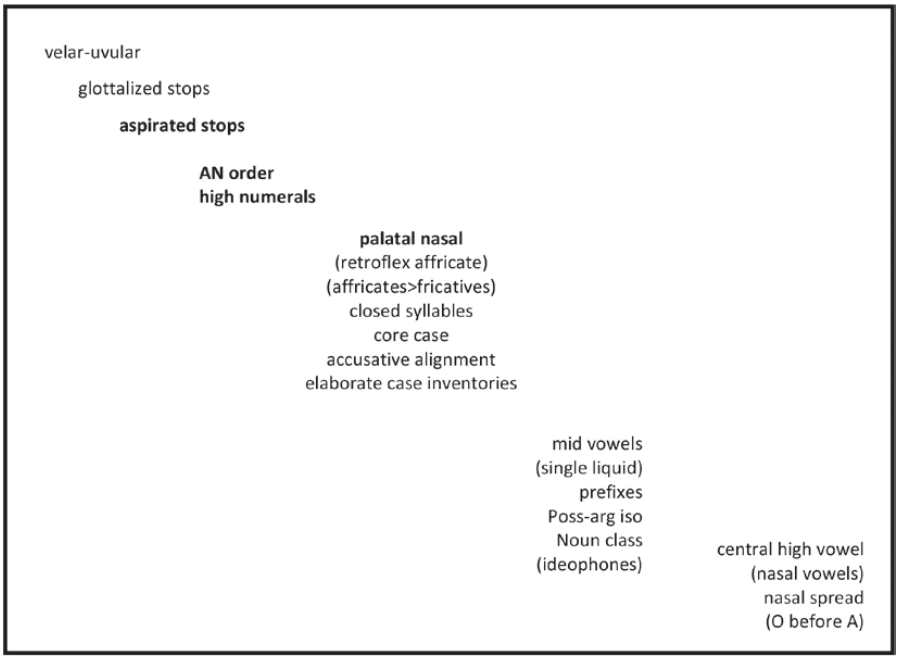

Figure 3.5.19 Classification of features as predominantly highland to predominantly lowland, and intermediate positions. © Rik van Gijn and Pieter Muysken.

1. Long-term and intensive contact with borrowing. In this case there should also be plenty of evidence of loanwords, which does not seem to be the case for either the lowland or highland languages.
2. Processes involving imperfect second language learning, for instance by (large) groups of immigrants who marry into a society. If this incoming group is numerous or prestigious enough, the variety they speak (which will include some of the abstract characteristics of the original language of the immigrants) can exert influence in the variety of the group as a whole.
3. Extensive multilingualism, where two (or more) linguistic systems stored in the brains of individuals may influence each other, becoming more alike, especially at an abstract level. If the situation of multilingualism is extensive enough and persists over time, this may lead to languages converging at the societal level (see for example, Matras 2011).

Scenario 1, above, seems unlikely because the amount of loanwords from highland languages in lowland languages and vice versa is limited (see also Bowern et al. 2011), although a definite answer to this matter requires a systematic investigation of lexica across the languages of the Andes and upper Amazon. Scenario 2 would require detailed and densely sampled genetic evidence to show great levels of admixture in upper Amazon groups, which, to our knowledge, is not available

at this point. Scenario 3 ideally requires the attestation of multilingual communicative practices. In the absence of such evidence, only indirect evidence, from archaeology, ethnology, and possibly geography can be brought to bear to make the case for scenario 3.

One striking conclusion that is suggested by the data discussed in this chapter, and visualized in Figure 3.5.19 is that if some of these distributions are indeed due to language contact, then the general picture suggests that such contact influences have operated mostly in one direction, from the highland languages into the lowland ones, rather than vice versa.

A non-contact-based account for the shared features between groups of languages is a deep-time genealogical link between them. Some linguists have claimed that grammatical features of languages tend to be highly stable (less changeable) through time (for example, Dunn et al. 2005). If particular grammatical characteristics tend to be very stable over time, they may be indicative of deep genetic links that cannot be recovered using more traditional methods. It is difficult to evaluate this claim, since linguists are still discussing the relative stability of individual linguistic features and the time depth they may represent, and no consensus seems as yet to be in sight (see Chapter 2.3).

From these considerations it becomes clear that a study such as this can only be preliminary, for several reasons. First of all, for many of the smaller languages, particularly in the northern part of our domain of research, the sources are fragmentary. Language data are coded on the basis of descriptions often written by missionary linguists with varying amounts of linguistic training, and the descriptions are far from systematic, making it difficult to be sure that one is coding reliably and consistently across all the different languages covered. In addition, only a limited set of features were included in our study.

Second, our study does not take a full historical perspective, as noted above, in at least two respects. We have not tried to establish, for each language family and its representatives, what the most likely original feature specifications may have been for that family as a whole. More historical research is certainly needed on the various families in this region. Furthermore, ethno-historical sources need to be taken into account in order to tell whether the current distribution of languages reflects their original distribution. It almost certainly does not. A good example of the type of study needed would be Wise (2014), who sketches the relationships between a number of languages on the eastern slopes of Peru, including Yanesha', Chamicuro, Cholón, Candoshi, and languages of the Jivaroan and Quechuan families. She establishes one cluster centred around the Jivaroan languages, but also including Candoshi, Shawi and Shiwilu, Chamicuro, Munichi, and Chachapoyas Quechua. The other cluster involves Campan languages, and Panao and Yaru Quechua. Wise notes that Yanesha' shares many features with languages in the northern cluster, which may point to population movements, possibly as late as the colonial period.

Third, many of the phenomena considered here will gain further significance from a geographically wider perspective, as in the work of González (2015) on the phonological features of the Chaco, within a wider South American context, and in papers by Lev Michael's group (Chang and Michael 2014; Michael et al. 2014).

As better language descriptions become increasingly available, along with modern techniques for analysing complex datasets, together these should allow for more sophisticated analyses of the complex patterns of interaction in the highland-lowland area. Such studies can also be backed up by historical-comparative work on individual families (which has so far lagged behind these structural comparisons), and by closer collaboration with ethno-historians, anthropologists, geneticists and archaeologists. All of this opens up promising perspectives for further research.

We hope that the data presented here will mark a step forward in the debate on the extensive linguistic areas of the Andes and Amazonia, and the interactions between them. In particular, we have tried to go beyond just presenting anecdotal evidence, by being as systematic as possible. Future work will hopefully flesh out the dynamics that lie behind the distributions of linguistic features that are found here.