

33

Linguistic Areas, Linguistic Convergence and River Systems in South America

Rik van Gijn, Harald Hammarström, Simon van de Kerke,
Olga Krasnoukhova and Pieter Muysken

33.1 Introduction

The linguistic diversity of the South American continent is quite impressive. There are 576 indigenous languages attested (well enough documented to be classified family-wise), of which 404 are not yet extinct. There are 56 poorly attested languages (which nevertheless were arguably not the same as any of the 576), of which three are known to be not yet extinct (Nordhoff et al. 2013). Furthermore, there are 71 uncontacted living tribes, mainly in Brazil, which may speak anywhere between zero and 71 languages different from those mentioned (Brackelaire and Azanha 2006).

Perhaps the most impressive aspect of the linguistic diversity of South American indigenous languages is their genealogical diversity. The languages fall into some 65 isolates and 44 families (Campbell 2012a; Nordhoff et al. 2013), based on a comparison of basic vocabulary. Nearly all of the families show little internal diversity, indicative of a shallow time depth. Alongside this picture of genealogical diversity, many areal specialists have noted the widespread occurrence of a number of linguistic features across language families, suggesting either sustained contacts between groups of people or remnants from times when the situation in South America may have been genealogically more homogeneous.

We only have historical records going back about 500 years, and the archaeological record is scanty and fragmentary (Eriksen 2011). It follows that there are large tracts of South American prehistory that are unknown. Areal-typological analysis aims to address this gap by finding out which languages have been in contact and how, as evidenced by their typological

features. Potentially, the location and intensity of language contact could answer questions about the (un)likelihood of deep genealogical relations and duration of interaction, and provide clues regarding the settlement of the continent. Intertwined with the goal of unravelling history is the modelling of different contact processes. We are not yet in a position to predict the linguistic outcome from a given contact scenario or vice versa (Muysken 2010). To achieve this, attested contact scenarios and attested typological feature distributions need to be confronted further.

However, there is another potentially significant set of variables that is so far largely unexplored in areal-typological linguistics, namely those related to physical geography. If physical geography mediates or even to some extent determines the location and intensity of contact in the region, as well as the dispersal of languages, specific historical events may play a less important role in accounting for the data. In the present chapter, we take up one variable from physical geography commonly thought to play an important role in South America, namely waterways.

We begin this chapter with a selective overview of previous studies in the field of areal linguistics in South America, discussing how the present study relates to them (Section 33.2). Section 33.4 describes the noun phrase structure questionnaire used for this study, and Sections 33.3 and 33.5 discuss the aspects of physical geography relevant to this study. In Section 33.6 we present the results of our survey, and in Section 33.7 we discuss the implications of our results and suggest some further lines of enquiry.

33.2 A Brief History of Areal-linguistic Studies in South America

33.2.1 Linguistic Areas

Traditional areal studies in South America have their roots in diffusionism in cultural anthropology, as reflected, for instance, in the writings of the Swedish archaeologist and anthropologist Nordenskiöld (1877–1932). A linguistic example of this approach is the work of Tessmann. Tessmann (1930: 617–627) attempted to group the languages of the Peruvian Amazon on the basis of lexical criteria (using a 33-word list). He assumes that the five largest accepted families represent ‘pure tribes’, and groups the remaining languages as ‘mixed’, depending how many lexical resemblances they show vis-à-vis words typical of the pure tribes. Areal influence is conceived here in terms of influence from large language families, as Tessmann (1930: 618) holds that ‘there is a considerable number of isolates’ but ‘nothing is achieved by positing a family of its own for each isolate’.

A recurring theme in later work in the areal linguistics of South America is often the Andean/Amazonian divide. The dominant perception (also

held in neighbouring disciplines such as anthropology, ethnohistory and archaeology) is that there is a basic two-way split in the indigenous languages of South America, between Andean (surveyed by Adelaar with Muysken 2004) and non-Andean or Amazonian (surveyed by Dixon and Aikhenvald 1999). This split is reflected in ideological constructs, with concomitant mental images, of the Andean highland societies and the Amazonian riverside village cultures, of the llama and the canoe. Andean societies are also commonly assumed to be 'more complex, with age-old sedentary habits and a highly diversified and technically well-developed architecture', to cite Adelaar (2008: 23–33).

The Peruvian linguist Alfredo Torero attempted in his final magnum opus (2002: 511–544), *Idiomas de los Andes*, to characterize the languages of the Andes typologically (see also van de Kerke and Muysken 2014). He contrasts the Andes with Mesoamerica, the Area Intermedia proposed by Constenla-Umaña (see below), and a possible Amazonian area (the latter with a question mark). The nine languages/small families taken into consideration are: Aru (Aymaran), Cunza, Cholon, Huarpe, Quechua, Uru(quilla), Mochica, Puquina and Mapuche. Torero uses 40 internal differentiation features to classify the nine languages. Out of the resulting 360 data points, only eleven are question marks. On the basis of his feature classification, Torero distinguishes a number of subareas (sometimes limited to a single language): *Nuclear* (Quechua, Aru), *Altiplano* (Uru, Puquina, Cunza, Huarpe), *Mochica*, *Cholon* and *Mapuche*.

The approach taken by Adelaar (2008) for the Andes is relatively hard to summarize, since it is not very systematic but at the same time very concise. Over 52 languages are taken into consideration, spanning the entire length of the continent and including the slopes on both sides of the Andes. It is clear that Adelaar conceives of the Andes in very wide terms. On the basis of his survey of structural features, Adelaar (2008: 31) concludes that 'there is still very little evidence that can be helpful for recognizing and delimiting linguistic typological areas, let alone, an Andean linguistic area that would encompass the entire region'. In fact, when contrasted with the rest of the continent, the area can be best defined in terms of negative features (after Adelaar 2008: 31):

- suffixing
- case marking
- no prosodic nasality
- no tone
- no complex vowel systems
- no nominal classifier systems
- no gender
- no stative/active systems
- no well-developed ergativity.

The Amazonian linguistic area is arguably more controversial than the Andean area. Nevertheless, specialists have observed several linguistic traits that are spread over large areas and across language families. We summarize the different characterizations of the Amazonian features in Table 33.1. Work has been done in this area by, among others, Derbyshire and Pullum (1986: 16–20), Doris Payne (1990) and Dixon and Aikhenvald (1999: 8–10). Those features that are mentioned by several authors are underlined; those features that are also found in the Andes are italicized.

The Costa Rican Chibcha specialist Constenla-Umaña (1991) has argued for a linguistic area in between the Andes and the well-known Mesoamerican area (Campbell et al. 1986), which we may call the Colombian–Central America area. He identifies almost twenty features as characteristic of the southern part of his sample, including parts of Ecuador and Peru. This part of his sample is assumed to be characterized by the features in Table 33.2.

Apart from these three major proposed areas of diffusion, proposals for smaller areas have also been made. Perhaps the best-known example of this is the Vaupés area (Aikhenvald 2002) in the border area between Colombia and Brazil, where Arawakan, Tucanoan and – to a lesser extent – Nadahup languages share a complex contact network resulting in convergence between the languages. The Guaporé-Mamoré area between the Bolivian and Brazilian border (Crevels and van der Voort 2008) is composed of two to three cultural areas. It is characterized by extreme genealogical diversity, but the languages share a number of structural traits; as argued by Crevels and van der Voort, these shared features are the result of contact. Seki (1999) argues that the upper Xingú area in central Brazil should be regarded as an incipient linguistic area on the basis of lexical borrowing and a few shared structural traits. Further cases of linguistic areas are summarized in Campbell (2012b: 301–309). It is interesting from the perspective of this chapter that these smaller areas in particular are associated with river systems.

33.2.2 Areal Typology

Until the present century, all proposals for linguistic areas, although drawn from decades of expertise in certain areas, were argued on the basis of impressions and selectively chosen feature lists. Recent studies have effectively closed the methodological gap by defining objective procedures that find and test prototypical linguistic areas, given language features and language locations as input. Muysken et al. (2015) provide a method for testing whether a proposed linguistic area actually has more homogeneity than expected by chance, and a method for finding the area delimitation that has the strongest areal signal.

Table 33.1 *Features used to classify the languages of the Amazon*

	Derbyshire and Pullum	Payne	Dixon and Aikhenvald
Word order	OS languages, noun-adjective, <i>possessor-possessed</i> , <i>noun-postposition</i>	Verb-initial orders	<i>Possessor-possessed</i>
Co- and subordination	No coordinating conjunctions, juxtaposition		<i>Subordination often through nominalization</i>
Alignment	(Remnants of) <u>ergative</u> case marking		Few oblique cases, many signs of <u>ergative</u> , purely nominative/accusative rare
Morphology		High degree of <u>polysynthesis</u>	<u>Polysynthesis</u> , <i>head-marking</i> , <i>agglutination</i> , many <i>prefixes</i> , but more <i>suffixes</i> , <i>person marking towards the periphery of the verb</i> , many TAME categories optional
Categories		Directional suffixes in the verb, <u>nominal classification</u>	Extensive semantic gender or <u>noun class</u> marking, <i>person markers in the nominal paradigm identical to the verbal markers</i> , small number systems
Phonology			Five vowel systems, nasalization, tone in some areas
Syntax			<i>Possession marked on the possessed element</i> , only one argument marked on the predicate, only obligatorily possessed nouns can be incorporated

Table 33.2 *Features of the Colombian–Central America area in Constenla-Umaña (1991)*

Segmental phonology	Absence of mid/high contrast in back vowels, absence of voiced/voiceless distinction in affricates, a voiceless alveolar affricate, palatal consonant subsystem, retroflex fricatives and affricates (not Quechua), mid/high contrast in front vowels (not Quechua), no glottalized sounds, no uvulars, voiceless labial fricative, roundedness opposition in back vowels (not Quechua)
Syntax	Adjective/noun order, numeral/noun order, question word in initial position, accusative case, genitive case, no nominal person marking, tense and aspect with prefixes (not Quechua)

Similarly, Michael et al. (2014) provide a slightly different method for testing whether a proposed area is distinguishable from its complement, and Chang and Michael (2014) define a method for inferring borrowing (as opposed to inheritance) and show that a series of borrowings between neighbouring languages is tantamount to finding a linguistic area.

Another set of recent studies aims to objectively test various proposals concerning feature distributions, typically involving (and calling into question) the Andes/Amazon divide, using standard statistical techniques to explore systematically completed databases. Van Gijn's survey (2014a) of the distribution of Andean and Amazonian features in the upper Amazon area shows that the transition from the Andean to the Amazonian area is gradual and complex, consistent with the intricate history of contact between the different ethnic groups of the area, and casting doubt on the reality of the Andean/Amazonian divide.

In the domain of verbal argument marking, Birchall (2014a, 2014b) examined the diverse array of verbal argument marking patterns encountered across the continent, and tested for regional distributions of certain often-discussed features. Statistical tests showed that many proposed areal distinctions in the literature are in fact not significant, and that an east/west division was often more significant than the classic Andean/Amazonian division. For instance, an inclusive/exclusive distinction in verbal argument marking does have an east/west distinction. Contrary to the suggestion by Adelaar (2008: 31), Birchall (2014a: 205–206) finds that 'the languages of the Andes show a lower distribution of clusivity in indexation than the rest of the continent' and that eastern South America shows 'a statistically significant areal distribution' of clusivity in indexation.

In the nominal domain, Krasnoukhova (2012, 2014) has also shown that in noun phrase structure there is a split between languages spoken in the western part and the eastern part of the continent, and not between the Andes and the Amazon as has been traditionally assumed. While the

western part corresponds to the Andean sphere, the eastern part includes languages spoken far beyond the Amazon region. Furthermore, in a case study of semantic features encoded by demonstratives, Krasnoukhova (2014) has shown that the Chaco and the southwest Amazon region stand out on the continent for encoding verbal categories with demonstratives (a case of the ‘trait-sprawl areas’ suggested by Campbell, Chapter 2, this volume).

Other areal-typological studies suggest that linguistic features that are statistically over-represented in South America, as opposed to other parts of the world, do not necessarily show clear systematic patterns of diffusion within the continent. Van Gijn (2014b) showed that nominalization as a subordination strategy is significantly more pervasive in South America than would be predicted on the basis of global patterns. The patterns found within South America are most consistent with a scenario of several smaller spreads, possibly promoted by a few language families with major extensions (e.g. Quechuan, Tupían, Cariban). In the domain of tense/aspect/mood/evidentiality (TAME) systems, Müller (2014) presented evidence that there is little systematicity in either the genealogical or the areal distribution of specific features, although there are several continent-wide general patterns, including highly frequent desiderative marking.

33.3 Exploring the Role of Geography: River Systems

So far, physical geography has remained an understudied factor, both in the study of linguistic areas and of areal typology. However, several authors (e.g. Dahl et al. 2012; Evans and Levinson 2009; Hammarström and Güldemann 2014; Nettle 1999; Nichols 1990, 1992) have surmised that physical geography plays an important role in shaping patterns of diversity around the globe. This is also the picture that emerges from the overview of the linguistic area studies given above.

The present study is intended as a first step towards integrating factors of physical geography into areal-linguistic studies. We are fully aware that a direct connection between geography and linguistic diversity would grossly oversimplify the complex relation between humans, their environment, and their personal, social and cultural answers to this ecology. It is therefore not our goal to present a deterministic picture of human behaviour. Rather, we want to try and isolate the possible contribution of geography to the patterns of linguistic diversity, in order to assess the facts that are unaccounted for, and which should most probably be explained in terms of human choice and socio-cultural organization and innovation. Having thus acknowledged the relative simplicity of the model presented here, we must try to

assess which aspects of geography are important to the rise and maintenance of ethnolinguistic diversity.

We conceive of the role of geography in shaping patterns of diversity as a rather indirect one. In broad terms, the physical surroundings potentially influence the migratory and contact-seeking or contact-avoiding behaviour of groups of people, which in turn influences diversification patterns and creates new contact situations. The role of geography in these processes has two sides: there is a set of factors to do with the *incentive* to spread or move, or get into contact with one's neighbours, and on the other hand a set of factors that either facilitate, or hamper or impede *mobility*, in terms of pathways and barriers.

33.3.1 Incentive Factors

Incentive factors can be described by what Nettle (1999) calls *ecological risk*, defined as 'the probability of a household facing a temporary shortfall, at whatever timescale, in food production' (1999: 79). In a constant and favourable climate and a rich environment (such as a seashore with plenty of molluscs), people are less forced to look for strategies to cope with ecological risk. If the climate is less dependable and the environment less rich, they may resort to a variety of strategies. The most important one in terms of language diversity is 'exchange', since that creates ties between peoples that become so strong that they start to identify or at least interact with each other, thus facilitating the spread of linguistic features.

What are the ecological factors that play a role in the degree of ecological risk? Nettle (1996, 1999) uses climatic data (rainfall and temperature per month) to calculate the length of the growing season. Based on these figures, each month is either included or excluded as part of the growing season. Months for which the average daily temperature is above 6 °C are included in the growing season if the average precipitation in millimetres exceeds twice the average temperature per month. The rationale behind this measure is: the longer the growing season, the lower the ecological risk and therefore the lower the incentive to seek contacts with other groups or to expand or move to other territories. Therefore linguistic diversity is expected to be especially high in areas of low ecological risk.

The growing season is a proxy for the degree of ecological risk, and as such it could be refined further on the basis of soil fertility, for example, or risk factors (especially the risk of inundation), or factors that facilitate the abundance of high-protein food (e.g. larger animals and fish), such as the presence of water and of dense forest. The notion of a growing season is an important one, but it should also be interpreted in terms of the availability of specific crops or gatherable plants.

33.3.2 Mobility Factors

Another set of factors of potential influence on linguistic diversity is concerned with mobility. It is probably true that a people with an incentive to migrate will do so, no matter what the difficulties they encounter on the way; see e.g. Nettle (1996) for a critique of explanations for linguistic diversity on the basis of topographical isolation hypotheses. It is nonetheless useful to consider the topographical context in terms of barriers or pathways of mobility, because it may still predict likely directions of spread, or perhaps highlight the fact that topographical barriers may raise the threshold for incentives to leave. Nichols (1990, 1992) observes that mountainous areas are in some cases linguistically diverse (Himalayas, Caucasus) but in others (e.g. the Andes versus the foothills and adjacent lowlands in South America) the diversity patterns are reversed. We will show that the Andes region is much more homogeneous than neighbouring areas.

In the absence of a fully fledged theory of the interaction between topographical factors and human mobility, our approach is sketchy and tentative, and focuses on only one part of the programmatic sketch outlined above: the mobility factor of waterways. Especially in the Amazon area, the many rivers provide possibilities for faster travel over relatively long distances, and increased carrying capacity. To the present day, boats are the main means of transportation for many of the indigenous peoples of South America.

In order to get a firmer grip on the role of physical geography in shaping patterns of diversity, we aim to answer the following three questions:

- (i) Are river system networks congruent with land distances as the crow flies? If not, do riverine distances mirror topological distances better than distances as the crow flies?
- (ii) Do languages on the same river system converge on the same structural profile?
- (iii) Are there specific features that spread easily along rivers?

To answer these questions, in the next two sections we give a more detailed introduction to the two major components of our study, the noun phrase database (Section 33.4) and the classification of geographical areas in South America (Section 33.5).

33.4 The Noun Phrase (NP) Database and Earlier Results

As mentioned, the structural features that we take into account in this chapter relate to NP structure. Technically speaking, the basis for the analysis is a questionnaire database, developed by Krasnoukhova (2012)

in order to outline the structure of the NP in South American languages. The NP questionnaire consists of six general questions (e.g. constituent order at the clause level) and 50 questions which specifically relate to the NP. The latter group of questions is divided into 31 main questions and 19 dependent questions. The areas of the NP which are explored in the questionnaire include those listed below.

NP structure

- Constituent order within the NP. Modifier categories such as demonstratives, numerals and adjectives were all considered as semantic categories.
- Presence and realization of agreement within the NP.

Modifiers within the NP

- Articles, demonstratives.
- Adjectives, grammatical status of adjectives.
- Numerals, grammatical status of numerals.

NP related issues

- Grammatical expression and conditions on the realization of number within the NP.
- Noun categorization devices, such as classifiers, and gender and noun class systems.
- Attributive possessive constructions. The parameters under investigation include: head versus dependent marking of possession, and the presence and formal realization of (in)alienability.
- Spatial deixis, with a focus on semantic features that can be encoded by adnominal demonstratives.
- Grammatical marking of temporal distinctions within the NP.

This questionnaire was filled in by Krasnoukhova and van de Kerke for 97 languages using descriptive materials, and whenever possible using information provided by specialists working on a specific language. The language sample used for this chapter includes representatives of 30 language families and 16 isolates. When constructing the sample we paid attention to the areal distribution of languages in order to ensure an adequate geographical coverage. Table 33A.1 in the appendix lists the languages included in our sample, along with their genealogical affiliation.

In what follows, we give a summary of observations based on Krasnoukhova (2012, 2014) with respect to the areal distribution of NP features. Krasnoukhova observed that there is evidence in the NP domain for a split between languages spoken in the *western* versus the *eastern* part of the continent, and not so much the split between the Andes and the Amazon. The western group consists of languages spoken along the western part of the continent and roughly corresponds with the Andean sphere, while the eastern group includes languages spoken in the rest of the continent, and thus is not limited to the Amazon region. Namely, the

eastern group includes languages spoken far beyond the Amazon region and includes, for instance, the eastern and southern part of Brazil, the Chaco and the Southern Cone. What are the linguistic features that suggest the west/east split in the NP domain?

The most robust features of the languages spoken in the western part are the following:

- word order within the NP: pre-head position for all modifiers (demonstratives, lexical possessors, numerals, property words);
- property words are morphologically nominal;
- absence of gender in NP;
- absence of gender distinction in personal pronouns;
- absence of classifiers;
- absence of temporal distinction within the NP;
- absence of a class of inalienable nouns.

The most robust features of the languages spoken in the eastern group can be condensed to the following list:

- word order within the NP, i.e. pre-head position for demonstratives, lexical possessors and numerals, and post-head position for property words;
- small or no distinct adjective class;
- property words are verbal;
- presence of gender in the NP;
- presence of classifiers, often of the multifunctional type;
- presence of temporal distinction within the NP (a less robust feature);
- presence of a class of inalienable nouns (a very robust feature).

Just to give a few concrete illustrations, for instance, hardly any language of the Andean sphere encodes property words by means of verbs, whereas languages in which property concepts are encoded by verbs are predominant in the northwest Amazon and the southwest Amazon regions; they are also found in the Chaco (e.g. Tapiete and Wichí), the eastern and southern part of Brazil (e.g. Timbira and Bororo) and in the Southern Cone (Tehuelche).

As for NP constituents, templates in which all modifiers tend to occur on one side of the noun are all found along the western edge of the continent, e.g. in Aymara, Huallaga Quechua, Imbabura Quechua, Leko, Mapuche, Tsafiki and Yanesha'. In Miraña, a language of the northwest Amazon region, all modifiers also tend to occur pre-head, therefore it constitutes an exception in this observation. Conversely, a template in which some modifiers (mainly demonstratives, possessors and numerals) always precede the head noun and some modifiers (mainly property words) always follow it, is found predominantly in languages outside the Andean sphere: e.g. Warao, Ninam, Dâw, Hup, Puinave, Urarina, Matsés, Yaminahua, Jarawara, Baure, Movima, Itonama, Mekens, Gavião, Wari', Karo, Kanoê,

Mamaindê, Sabanê, Wichí, Pilagá, Chamacoco, Bororo, Kamaiurá, Trumai and Timbira. There are a few exceptions here: the latter template is also found in three languages spoken in the northwestern part of Colombia, i.e. Ika, Nasa Yuwe and Northern Emberá.

To give one more example, a great majority of languages spoken along the western edge of the continent lack a class of inalienable nouns, whereas languages in the rest of the continent predominantly have such a class. Specifically, the following languages do not have a class of inalienable nouns: Quechua and Aymara, spoken in the Andes; Mapuche, spoken in the Southern Andes; Tsafiki, Awa Pit, Nasa Yuwe and Northern Emberá, spoken in the western part of Ecuador and Colombia. Inalienable nouns are also absent in Aguaruna, in the northern Peruvian foothills, and Shipibo-Conibo and Urarina, spoken in the Western Amazon (Peru). Exceptions from the observations are Warao (spoken in western Guiana and northeastern Venezuela) and Kwaza and Sabanê (spoken in the southwest Amazon region). All other languages in the sample have a class of inalienable nouns.

The data on the NP also show that the following features *cannot* be treated as characteristic of any particular larger area (western versus eastern part, specifically Andes versus Amazon), as they are found in the languages across the continent (see also the discussion in Campbell 2012b: 301–304):

- number distinction/markings in personal pronouns;
- number distinction/markings in the NP;
- inclusive/exclusive distinction in free personal pronouns.

Features such as ‘locus of possession marking’ were carefully examined for areality too, since the head-marking pattern has been proposed in the literature among the features of the ‘Amazonian linguistic area’ and the double-marking pattern was proposed among the features of the ‘Andean linguistic area’ (Dixon and Aikhenvald 1999: 8, 10). Our NP data showed no evidence for this areal division. It is correct to say that the double-marking pattern is found predominantly in the languages spoken in the Andes (specifically, among the Aymaran and many of the Quechuan variants, and in Aguaruna), but this is not the only possession strategy found in the languages spoken in the Andes. For instance, Chipaya is dependent-marking, as well as Imbabura Quechua. Likewise, there is no good evidence to generalize that Amazonian languages are predominantly head-marking for possession: both a dependent-marking and a head-marking pattern are equally common in the Amazonian languages in the sample. What was found was a much smaller areal clustering: namely, dependent-marking languages are more concentrated in the Western Amazon region, while head-marking languages are more present in the Bolivian lowlands and in the Chaco. In addition,

head-marking languages are found along the western coast of South America (e.g. Tsafiki, Ika) and in the Southern Cone (e.g. Tehuelche) (see Krasnoukhova 2012 for more details).

33.5 The Major Drainage Basins of South America

The importance of rivers and river systems for the population dynamics in South America cannot be overestimated (Dunne and Mertes 2007). It is no coincidence that many proposals for linguistic areas of the continent are associated with river systems: the Vaupés-Içana (Aikhenvald 2002), the Guaporé-Mamoré (Crevels and van der Voort 2008) and the Xingú (Seki 1999). River systems, and the ecologies associated with them, to a considerable extent determine the possibilities for gathering and cultivating food and cultivating plants, and they facilitate mobility and with it contact between peoples. The effect of the ecological circumstances can be expected to be even greater for locally organized, rural economies (Nettle 1999), as is the case for most indigenous cultures of South America.

The connection between linguistic areas and river systems, mentioned above, raises the question of whether there is a more general interaction between river systems/ecologies and the degree of linguistic convergence. What we set out to do in this chapter, therefore, is to define a number of ecological zones mainly based on river systems, and look for patterns of structural diversity and convergence within each of these areas.

33.5.1 The South American Physical Space

The elements responsible for creating the ecological 'hardware', that is, the physical landscape and the basic ecological conditions of an area, are tectonism (the shaping of the Earth's crust) and climate (Orme 2007). In turn, these two constrain flora and fauna, and to a certain extent human activity. The South American continent is delimited in the north-west by the border of Colombia with Panama, and on all other sides by seas and oceans – the Caribbean Sea to the north, the South Pacific Ocean to the west and the Atlantic Ocean to the east. Broadly speaking, the physical geography of South America is dominated by three highland areas: the ancient Precambrian (prior to 541 million years ago) Guyanan and Brazilian highlands, and the much younger Andean mountain range, which only assumed its present shape in the Cretaceous period (145–166 million years ago). The rivers that spring from these highland areas flow through the vast lowland areas, forming some of the largest river systems in the world. These elements form the basis that underlies the different ecological systems of the continent.

In order to achieve a more local perspective (assuming that contact in local ecosystems or river systems becomes even more manifest), we divide the continent into several smaller regions. First, we distinguish between the very different ecosystems of the Pacific, the Andes and the Amazon. Second, we make further distinctions within the Amazon and Andean area, based on more local river systems and other factors.

33.5.2 Ecological Zones

The Andean mountain range is usually divided into a southern, central and northern part. However, there are different ways to determine the borders of these areas (Torero 2002: 13): to the south the Bolivian Altiplano may or may not be counted as part of the Southern Andes, and to the north, the Ecuadorian Andes are either included in or excluded from the Northern Andes. In line with our continental north/south divide, and with the secondary north/south divide in Western Amazonia, we put the border between the Central and Northern Andes in northern Peru and southern Colombia, where the Andean mountain range takes the shape of three parallel cordilleras. To the south, the border is determined by the very broad Bolivian Andean range, where the rivers flowing north join the Amazon but those flowing south join the Paraguay River, and eventually the Río de la Plata. Because of low language diversity to the south of the Gran Chaco, the Southern Andes and Southern Cone are considered together as a single area. The Andean areas are strictly speaking not river-defined areas, but they do form ecologically coherent areas, and they cannot easily be grouped with any of the river-based areas.

We also regard the Pacific coast as a separate area. Few Pacific coastal languages have survived, causing the focus of our Pacific area for the purposes of this chapter to be on the northern Peruvian/Ecuadorian coast, where a relatively low passage of the Andes between northern Peru and southern Ecuador made trade between the coast and the Amazon possible (Adelaar with Muysken 2004: 6).

In the northeast, we separate the Orinoco basin in Venezuela from the adjacent Guyanas and northern Brazil, and south of this area we recognize smaller river systems, which are defined in Table 33.3. The Río de la Plata basin (in particular the western part, the Gran Chaco in northern Argentina, southern Bolivia and Paraguay) and the Southern Cone plus Southern Andes are also separated. In the list that follows, we briefly outline the smaller areas and indicate which of the sample languages are spoken in each region. The numbers in the first column of the table refer to the numbers in the accompanying map (Map 33.1).

Mapping the NP questionnaire database discussed in Section 33.4 onto the geographical zones discussed in this section, we can now move to a discussion of the results.

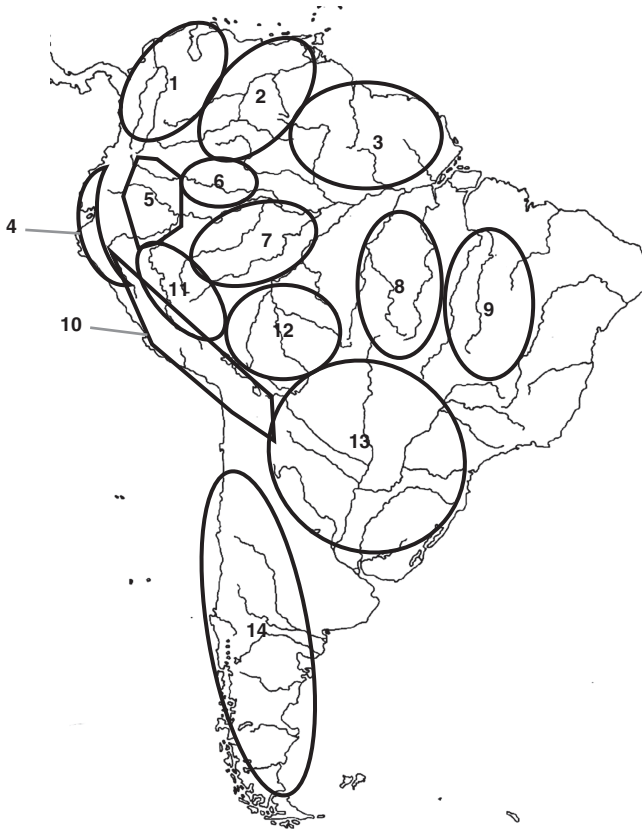
Table 33.3 *Ecological zones of South America: regions, locations and sample languages*

	Region	Delimitation/description	Sample languages
1	Northern Andes	Colombian Andes	Arhuaco, Embera, Páez
2	Orinoco	Venezuelan Orinoco basin	Nhengatú, Ninam, Panare, Puinave
3	Guyanas	French and British Guyana, Surinam, northern Brazil	Emérillon, Hixkaryána, Papiamento, Trió, Warao
4	Pacific	Pacific coast: practically the Ecuadorian Pacific	Awa-Cuaiquer, Colorado, Mochica
5	Napo-Marañon	Ecuadorian and north Peruvian eastern slopes and lowlands	Bora, Aguaruna, Kokama, Muniche, Shuar, Urarina, Cofán, Iquito, Arabela
6	Vaupés	The Colombian–Brazilian border	Hup, Cubeo, Dâw, Desano, Tariana
7	Solimões	Western Brazil between the Juruá and Purus Rivers, and up to the Ucayali in the west	Jarawara, Matsés, Yaminahua
8	Tapajos-Madeira	River basins in southern Amazonia directly east of the Madeira-Guaporé line	Apurinã, Bororo
9	Xingú-Tocantins	Easternmost river basin in southern Amazonia	Kamaiurá, Krikati-Timbira, Trumai
10	Central Andes	Ecuadorian, Peruvian and Bolivian Andes	Aymara, Cusco Quechua, Huallaga Quechua, Uru, Chipaya
11	Huallaga-Ucayali	Central-south Peruvian lowlands	Shipibo-Conibo, Yanesha', Cholon
12	Beni-Guaporé	Area between the Beni and Guaporé Rivers in Bolivia and Rondônia	Baure, Cavineña, Gavião, Itonama, Kanoê, Karitiana, Karo, Kwazá, Lakondê, Leko, Mamainde, Moseten, Movima, Sabanê, Mekens, Warí, Yurakaré
13	Río de la Plata basin	Northern Argentina, Paraguay	Wichí, Chamacoco, Mocoví, Pilagá, Tapiete
14	Southern Cone	Central and southern Argentina and adjacent areas in Chile	Tehuelche, Mapudungun, Selknam-Ona, Gününa Küne-Puelche

33.6 Results

Given the geographical information about the various river systems, and the datasets on NP features for the languages in our sample, we are now in a position to try and answer a number of questions. As mentioned above, and repeated here for convenience, we address the following issues.

- (i) Are river system networks congruent with land distances as the crow flies? If not, do riverine distances mirror typological distances better than distances as the crow flies?
- (ii) Do languages on the same river system converge on the same structural profile?
- (iii) Are there specific features that spread easily along rivers?



Map 33.1 Ecological zones of South America: smaller regions considered in the present study. See Table 33.3 for legend

Several methodological remarks are in order here. First, the added geographical information by itself does not help decide whether a certain feature distribution could reflect deep time geographical relationships or later contact. A river may be a conduit for exchange relations between unrelated groups as well as a path along which a population spread. Second, not all measures can be applied globally. It does not make sense to calculate river distance between two points which are not directly connected by rivers.

Our data on rivers in South America come from the freely accessible country-by-country river information from DIVA-GIS.¹ The granularity of these data is too fine for the purposes of this chapter, as it includes what appears to be every small creek on the continent. We selected the 600 largest shapes, resulting in what is impressionistically a granularity suitable for the purposes in the present chapter. This yields the river system shown in

¹ www.diva-gis.org/gdata



Map 33.2 River data and language locations in South America. Lines from every language location to their nearest point on a river

Map 33.2, shown at a resolution of $1/8$ of a degree per point. Languages are represented with their centrepoint coordinates from Nordhoff et al. (2013).

We define river distances only for pairs of languages which are in the same drainage area, that is, languages that are actually connected by a river system. The river distance between two languages A and B is defined by the distance of A to the nearest point on a river plus the distance of B to the nearest point on a river plus the shortest river path between those two points. Map 33.2 shows a line from every language location to its nearest point on a river (based on the river data explained above). As an example,



Map 33.3 The river path between the languages Hixkaryána and Nhengatu

in Map 33.3 the river path between the languages Hixkaryána and Nhengatu is indicated. The distance between Hixkaryána as the crow flies is 917 kilometres whereas the river distance, following the river path, is 1,642 kilometres.

33.6.1 River versus Bird Distances

Before relating any linguistic data to water distances, we must ask if there are any essential differences between riverine distances and distances as the crow flies. There are some logical possibilities. One possibility is that river distance for all language pairs merely prolongs the bird distance by

a certain percentage. In this case, the landscape of contact effects would be the same as if there were no rivers, and the only reason to compare linguistic data and river distances would be to estimate the size of the effect, not the locus. On the other hand, it may be that some language pairs differ greatly from other pairs in terms of water distance. That is, some pairs of languages connected by a straight river path have a river distance which is very close in kilometres to their bird distance, while other pairs, that require a lengthy detour by river, have a river distance far higher than their bird distance. In this case it would be possible to test for riverine contact as a factor different from contact by proximity in general. If so, the linguistic distances should correspond better to the riverine distances than the bird distances.

Figure 33.1 shows all 510 pairs of languages for which water distances are defined, plotted against their bird distance. The water distance for most language pairs can indeed be found by adding a percentage to their bird distance, though not for all. The river distances between Cusco Quechua and Quechuan languages further north requires a very long downward route via the Madre de Dios, up the Madeira, and back up the Amazon and the Ucayali, totalling around 5,000 kilometres. Similarly, there are other pairs which require river travel down the Napo in

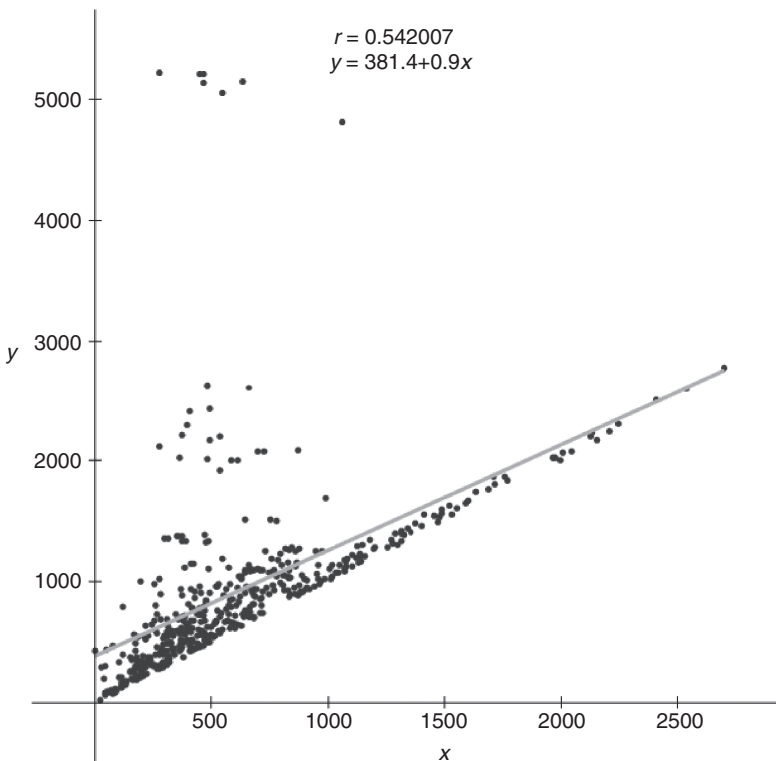


Figure 33.1 Bird versus water distance

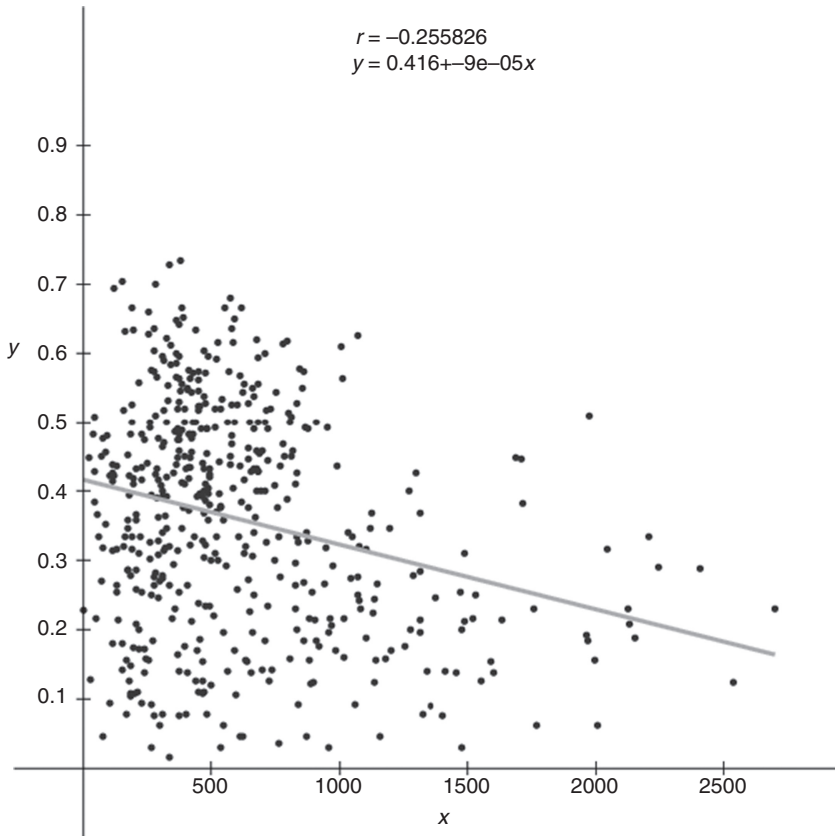


Figure 33.2 Bird versus NP distance

Ecuador only to turn around at the Japurá and go as far back up, totalling around 2,000 kilometres.

For the linguistic distance between a pair of languages, we use the Hamming distance, which is simply the percentage of features which do not have the same value in the two languages (see Hammarström and O'Connor 2013). For example, if 80 features are defined for the pair of languages and they agree in value for 20 of them, then the distance is $60/80 = 0.75$. The correlation between bird distance and linguistic (NP) distance is shown in Figure 33.2. The correlation between water distance and linguistic (NP) distance is shown in Figure 33.3.

The size of the correlation for bird and NP distance is $r = -0.26$, while for water and NP distance it is slightly less strong, $r = -0.23$. So there is little overall improvement in using water distances to predict NP distance. We also did a more specific test, looking only at the regions where the river systems were thought to be more significant, that is, Huallaga-Ucayali, Orinoco, Guaporé-Mamoré, Vaupés, Napo-Marañon, Tapajos-Madeira, Xingú-Tocantins and Solimões, thereby excluding pairs such as those involving the highly unnatural river route between Cusco Quechua

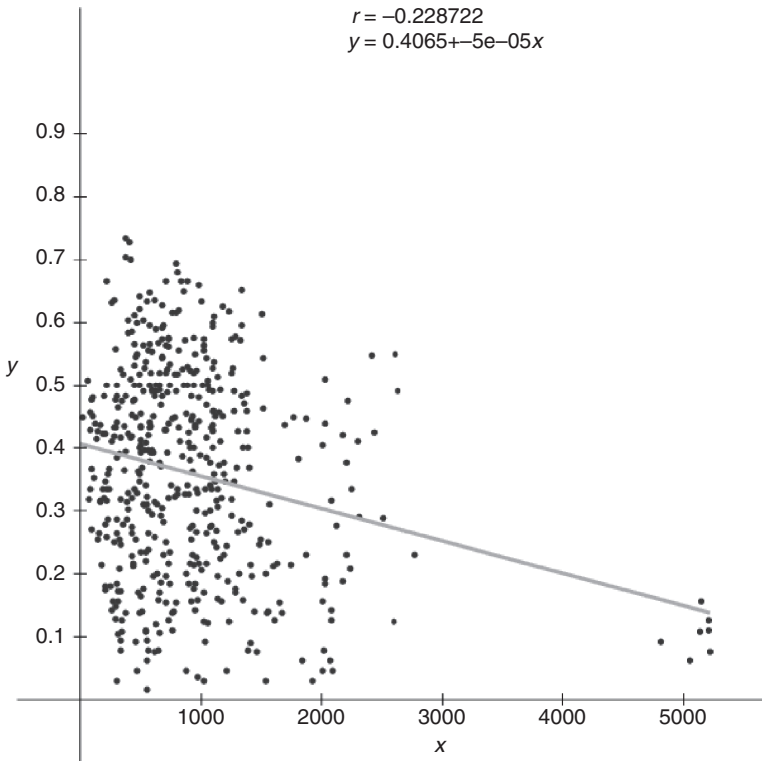


Figure 33.3 Water versus NP distance

and other Quechuan languages. In those cases, there were plausible land routes, as indeed were created during the Inca Empire.

Doing this drastically changes both bird/water distance correlations with NP distance to the values $r = 0.11$ and $r = 0.12$ respectively. So again there was no appreciable difference of water vis-à-vis bird distance.² We can thus answer the first question. Riverine distances are very different from distances as the crow flies for many language pairs, but the linguistic differences are not better predicted by the riverine distances.

33.6.2 Riverine Areas

While there is no simple relationship involving water distance and NP distance in our data, there may be a more indirect effect of river

² While we were concerned with differences between water and bird distance, the reader may have noted that the correlation in the first case was negative, i.e. that longer geographical distance predicts shorter linguistic distance. This unintuitive result is, however, due to the many pairs of languages belonging to the shallow but geographically widespread Quechua family. Quechuan languages are much more similar to other Quechuan languages far away than to many neighbouring non-Quechuan languages and, in general, much more similar to each other than a random pair of languages from the rest of the sample. In the second case the Quechuan languages were not included, and the geographic versus linguistic correlation then showed the intuitively expected positive polarity.

systems. Languages on the same river system may be hypothesized to interact both along the main waterways and cut across headwaters. Given a sufficiently long time, these languages may then influence each other structurally. The hypothesis is that they influence each other more than any set of geographically adjacent languages which do not share a river system. We thus test if the languages belonging to the same riverine area are more similar to each other than (1) random collections of the same size (Sig Random), (2) geographically coherent sets of languages of the same size (Sig Geo), and (3) random collections of languages of the same size *from different families* (Sig FamI). To generate a random geographically coherent set of languages, we start from a random language and then continue picking the geographically nearest language until the desired size has been reached. The difference between riverine areas and geographically coherent areas is merely that the riverine areas are geographically coherent sets of languages necessarily associated with a river system. To generate a random set of languages from different families, we first select a random set of families of the desired number and then select a random member from each. Table 33.4 shows the average distance of the languages within each area and the significance vis-à-vis random sets of languages, random geographically coherent sets of languages and random sets of languages from different families, of the same size as the area in question. The significance number counts the number of random sets with a higher average distance. A number of 950 or higher can be equated with a conventional significance level of 0.05.

Table 33.4 *Convergence of the NP profile in different areas of South America*

	Avg Dist ^a	Sig Random	Sig Geo	Sig FamI
Northern Andes	0.30	924	965	975
La Plata	0.47	265	114	461
Guyanas	0.43	541	612	628
Southern Cone	0.43	564	637	701
Huallaga-Ucayali	0.38	859	892	924
Central Andes	0.18	1,000	1,000	1,000
Orinoco	0.47	379	357	444
Vaupés	0.43	611	605	703
Napo-Marañon	0.40	879	907	971
Solimões	0.45	521	496	507
Xingú-Tocantins	0.33	864	910	935
Guaporé-Mamoré	0.47	157	82	304
Pacific	0.34	883	971	

^a Avg dist is the average distance for languages in the same area. Sig Random indicates how many (out of 1,000) random sets of languages of the corresponding size have a lower average similarity. Sig Geo indicates how many (out of 1,000) random sets of geographically coherent languages of the corresponding size have a lower average similarity. Sig FamI indicates how many (out of 1,000) random sets (of the corresponding size) of languages from different families have a lower average similarity.

Few, if any, riverine areas show significant convergence. The only one to clearly do so is the well-known Central Andes area, where the similarity is known to be due to genealogical inheritance and intense language contact, neither involving rivers in any significant way. Also the Northern Andes and Napo-Marañon areas owe their high significance numbers to the fact that they encompass subsets of very close Quechua varieties. Interestingly, the set of riverine areas tested includes prospective and established linguistic areas like the Guaporé-Mamoré and the Vaupés, whose NP convergence is well within the bounds of similarity expected for random sets of languages. Thus we can answer the second question: merely being in the same river system is not sufficient for convergence of the entire NP profile.

33.6.3 Riverine Features

While entire typological profiles do not appear to converge in languages of the same river system, a weaker effect of rivers may be operational. Certain unstable features, not entire profiles, may spread easily along rivers. The hypothesis is that this concerns only the most unstable features, because if it concerned every feature the entire profiles would have converged. To get an idea of which features are likely to be inherently stable versus unstable, we take over the stability measure calculated in Muysken et al. (2014). This essentially measures the probability that a feature value remains the same in the transition from a proto-language to a modern language within a language family.

Suppose we define a ‘riverine’ feature as one which tends to have the same value in each riverine area, i.e. homogeneous within a riverine area. For example, if two-thirds of languages in an area have the same value in a certain riverine area, then the homogeneity in that area is 0.67. The overall homogeneity is the average it has across all riverine areas. (For this test, we have used only the strictly river-defined areas, excluding the Central Andes and Southern Cone areas.) We may also test for significance of homogeneity by comparing simulated random and geographically coherent sets of the same size (similar to the above). The riverine-ness of each feature is shown below (with the number of languages in the full sample for which the feature is defined, the overall homogeneity and significance of this).

Table 33.5 shows figures for how riverine a feature is, along with its stability value. Although there are a number of NP features which achieve high significance individually, notably many that concern possession, the numbers have not been controlled for multiple testing. There are about equally many that achieve a correspondingly low significance, suggesting that there are random fluctuations in both directions. Also, more importantly, there is no correlation between low stability and riverine-ness ($r \approx 0.01$).

Table 33.5 *How 'riverine' are the NP features?*

Feature ^a	No.	Hom.	Sig Random	Sig Geo	Stability
NP406: Do possessive constructions with optionally possessed nouns differ from those with obligatorily possessed nouns? (= Is alienable and inalienable possession formally distinguished?)	61	0.683	0.999	1.000	0.872
NP405: Are there nouns denoting obligatorily possessed items?	94	0.809	0.999	1.000	0.901
NP408: Does the language have a separate form for third person possessive reflexive?	34	0.824	0.996	0.998	1.000
NP160: What is the most frequent order of adjective and degree word?	31	0.807	0.994	0.998	1.000
NP13: In what way are indefinite pronouns realized?	30	0.807	0.994	0.998	1.000
NP409: Can pronominal possessors (pronoun, affix, clitic) be related to main tense and verbal person markers (pronoun, affix, clitic)?	25	0.794	0.993	0.996	0.825
NP150: What is the most frequent order of relative clause and noun?	82	0.709	0.989	0.997	0.789
NP404: In possessive constructions with a pronominal possessor, is the POSSESSED noun usually marked?	95	0.632	0.962	0.921	0.837
NP740: Is there a morphologically indicated collective marker?	81	0.691	0.947	0.962	0.869
NP305: How many distance contrasts do adnominal demonstratives encode?	91	0.668	0.900	0.951	0.757
NP304: What system of demonstratives is present in the language? (Distance-oriented system vs person-oriented system.)	80	0.725	0.891	0.920	0.820
NP307: Can adnominal demonstratives encode altitude?	92	0.974	0.889	0.872	0.968
NP508: Which word class do (the majority of) words denoting SPEED belong to?	66	0.550	0.882	0.958	0.792
NP507: Which word class do (the majority of) words denoting HUMAN PROPENSITY belong to?	67	0.587	0.866	0.882	0.861
NP402: In possessive constructions with a nominal possessor, is the POSSESSED noun usually marked?	96	0.686	0.855	0.819	0.806
NP401: In possessive constructions with a nominal possessor, is the POSSESSOR usually marked?	96	0.700	0.851	0.964	0.896

^a The columns of the table consist, from left to right, of (i) every NP feature, (ii) the number of languages for which it is defined, (iii) its average homogeneity in riverine areas, (iv) the significance of this homogeneity with respect to randomized mirrors of the riverine areas, (v) the significance of this homogeneity with respect to geographically coherent randomized mirrors of the riverine areas, and (vi) the feature's stability value.

Table 33.5 (cont.)

Feature	No.	Hom.	Sig Random	Sig Geo	Stability
NP311: Can adnominal demonstratives encode gender?	93	0.837	0.830	0.826	0.975
NP230: Is there agreement between demonstrative/ determiner and the noun in the NP?	88	0.727	0.823	0.897	0.896
NP810: How many gender distinctions are realized within the NP?	94	0.853	0.822	0.926	0.922
NP312: Can adnominal demonstratives encode number?	92	0.737	0.809	0.561	0.850
NP140: What is the most frequent order of property word (adjective) and noun?	95	0.626	0.797	0.626	0.771
NP309: Can adnominal demonstratives encode movement?	93	0.928	0.782	0.854	0.963
NP12: Is nominalization the main strategy for complement clause formation?	60	0.605	0.774	0.772	0.900
NP316: What is the morphological composition of adverbial demonstratives (as compared to adnominal demonstratives)?	86	0.645	0.713	0.748	0.785
NP317: How many distance contrasts do adverbial demonstratives encode?	77	0.603	0.688	0.815	0.849
NP506: Which word class do (the majority of) words denoting PHYSICAL PROPERTY belong to?	69	0.574	0.685	0.677	0.882
NP308: Can adnominal demonstratives encode position of the referent in space or posture?	93	0.896	0.684	0.673	0.963
NP901: What kind of adposition type is most dominant?	87	0.684	0.666	0.721	0.861
NP505: Which word class do (the majority of) words denoting COLOUR belong to?	70	0.574	0.647	0.686	0.864
NP11: Is nominalization the main strategy for adverbial clause formation?	61	0.562	0.622	0.676	0.925
NP130: What is the most frequent order of possessor and possessed?	97	0.897	0.622	0.525	0.941
NP315: Do demonstratives receive different inflectional features when used pronominally or adnominally?	80	0.642	0.616	0.858	0.828
NP313: Can adnominal demonstratives encode physical properties of the referent?	92	0.879	0.609	0.585	0.942
NP720: Do nouns have a morphologically marked dual?	87	0.834	0.601	0.473	0.952
NP820: Are there any (grammaticalized) sex-markers (realized on the noun itself)?	90	0.726	0.573	0.505	0.851
NP730: Do nouns have a morphologically marked paucal form?	89	0.899	0.570	0.443	1.000
NP2: What is the pragmatically unmarked word order for a transitive clause?	94	0.632	0.544	0.561	0.798
NP314: Do adnominal demonstrative roots require further derivation?	90	0.868	0.531	0.377	0.953
NP403: In possessive constructions with a pronominal possessor, is the POSSESSOR usually marked?	96	0.635	0.523	0.726	0.831
NP310: Can adnominal demonstratives encode animacy?	93	0.836	0.481	0.275	0.950
NP9: Is there a number distinction in third person pronouns?	96	0.800	0.479	0.311	0.861
NP301: Are there indefinite articles in use?	86	0.655	0.466	0.585	0.763
NP7: Is there an inclusive/exclusive distinction in personal pronouns?	96	0.685	0.436	0.182	0.924
NP407: Is a word/morpheme meaning 'pet' (or similar) required in possessive constructions involving nouns which denote (domesticated) animals and/or food?	83	0.717	0.368	0.384	0.858
NP303: Are third person pronouns and demonstratives related?	91	0.597	0.334	0.373	0.773
NP830: Are there classifiers?	93	0.690	0.329	0.172	0.906
NP503: Which word class do (the majority of) words denoting AGE belong to?	68	0.524	0.328	0.224	0.774
NP4: Is there a copula for predicate nouns? (Only present tense clauses are taken into account.)	79	0.590	0.327	0.548	0.857
NP1020: Can temporal or aspectual distinctions with a propositional scope be marked on the head noun?	84	0.818	0.308	0.412	0.944
NP501: Is there a class of synchronically underived adjectival elements (lexemes)?	96	0.650	0.302	0.297	0.825
NP10: Is nominalization the main strategy for attributive (relative) clause formation?	70	0.579	0.259	0.338	0.913
NP1: What is the pragmatically unmarked word order for an intransitive clause?	95	0.711	0.197	0.286	0.866
NP5: Is there a copula for predicate locative phrases? (Only present tense clauses are taken into account.)	61	0.526	0.173	0.313	0.823
NP6: Is there a copula for predicate adjectives?	76	0.560	0.161	0.356	0.861
NP504: Which word class do (the majority of) words denoting VALUE belong to?	70	0.533	0.143	0.165	0.855
NP210: Is there agreement between adjective (property word) and the noun in the NP?	86	0.654	0.139	0.242	0.890
NP306: Can adnominal demonstratives encode visibility?	93	0.804	0.139	0.216	0.925
NP610: Which class do native numerals belong to?	65	0.490	0.128	0.261	0.824
NP110: What is the most frequent order of demonstrative and noun?	96	0.819	0.123	0.101	0.935
NP8: Is there a gender distinction in independent personal pronouns?	96	0.685	0.090	0.054	0.900
NP120: What is the most frequent order of numeral and noun?	93	0.663	0.089	0.071	0.906
NP710: Do nouns have a morphologically marked singular versus plural distinction?	96	0.619	0.087	0.058	0.815
NP502: Which word class do (the majority of) words denoting DIMENSION belong to?	69	0.513	0.086	0.065	0.858
NP220: Is there agreement between numeral and the noun in the NP?	83	0.655	0.079	0.103	0.932
NP302: Are there definite articles in use?	89	0.583	0.053	0.103	0.818
NP1010: Can any temporal or aspectual distinctions, restricted to NPs, be marked on the head noun?	86	0.625	0.024	0.015	0.910

We are thus inclined to answer the third question in the negative too: we find little evidence for specific features whose spread would be facilitated by rivers.

33.7 Discussion and Conclusion

The results of our study suggest that river systems do not contribute significantly to patterns of diffusion of structural NP features in South America. This is a surprising result given the fact that waterways have played such an important role in the mobility of the indigenous peoples of the continent. A few reasons for this discrepancy can be put forward, leading to a number of suggestions for further research.

First, it may be the case that structural features of the noun phrase are not very prone to diffusion through contact, perhaps due to the fact that overt NPs in South American languages with their highly synthetic verbal morphology are not altogether common. This would for instance also explain why some of the established or proposed linguistic areas associated with particular river systems do not show any significant degree of convergence either. On the other hand, Krasnoukhova (2012) does show that there are some large-scale areal patterns in South America for some of the NP-related features in her study. One suggestion for further research may nevertheless be to look at other domains of grammar.

The picture regarding word borrowing, for instance, is also quite complex. Bowerman et al. (2011) report that word borrowing is not frequent among the Amazonian languages. Altogether, there is no evidence that suggests the contrary. Greenberg (1987) mentions some lexical items that have a surprisingly wide distribution, but without a systematic pattern. Muysken (2012: 252) draws attention to specific culture items that have a very wide distribution, such as the words for 'chicken' or 'hundred'. It would be good to test the distribution of specific lexical items in terms of river systems. Similarly, the spread of phonological features should be studied keeping river systems in mind. The same holds, of course, for other morphosyntactic features.

Second, it may be the case that river systems are simply less relevant to human migratory and contact patterns than is often assumed. It is probably too early to say whether or not this conclusion holds up, especially given the fact that many of the linguistic areas discussed in Campbell (2012b) for South America are connected to some river system. Furthermore, it may be a matter of finding the right scale. Even the smaller areas in our study are still rather large, and perhaps a more fine-grained analysis is required to find areal patterns of a more local nature. Finally, as mentioned in Section 33.3.2 above, waterways are only one element of

a potentially much more encompassing geographical approach to linguistic diversity. Embedding the role of river systems into such a larger programmatic approach may also yield more refined results.

Third, the relation between river systems and specific structural features may simply be too indirect to yield significant results. After all, although physical geography may influence whether people decide to expand, or to contact other people, and also the directions they are most likely to travel for these purposes, there is still a large field of unknown variables that will influence the eventual patterns of structural linguistic diversity.

Fourth, rivers can unite peoples on both sides if they are easily navigable and have good communication systems. On the other hand, very large rivers or those difficult to navigate can separate peoples, in which case we have the opposite effect. So size and navigability play an important role.

Fifth, these unknowns lie within the field of human agency. Rivers exist 'in the mind' as well.³ People may have different strategies to reduce their ecological risk, which do not necessarily involve either migration or trade (such as diversification or storage, see Nettle 1999: 80) and therefore not necessarily language contact. And even if they respond to disadvantageous ecological circumstances by mobility, this may lead to different types of encounters between peoples: they might for instance be trade-oriented or bellicose, intensive or shallow, short-term or long-term, and policies with respect to contact may vary considerably from one contact situation to another (the Vaupés area as described by Aikhenvald 2002 is a case in point).

Finally, as mentioned in Section 33.1, there is no obvious connection between contact scenarios and their linguistic outcomes to the extent that we can predict one on the basis of the other. Further research should ideally take into account as much as possible of the ethnohistorical and sociological information available.

Appendix

Table 33A.1 *Areal linguistics of South America: the language sample*

Language	ISO code	Genealogical affiliation	Region
Arhuaco	arh	Chibchan	Northern Andes
Inga	inb	Quechuan	Northern Andes
Páez	pbb	Paez	Northern Andes
Chamacoco	ceg	Zamucoan	La Plata
Ayoreo	ayo	Zamucoan	La Plata
Kadiwéu	kbc	Guaicuruan	La Plata

³ We are grateful to Raymond Hickey for drawing our attention to this psychological component and its possible consequences for language development.

Table 33A.1 (cont.)

Language	ISO code	Genealogical affiliation	Region
Wichí Lhamtés Nocten	mtp	Matacoan	La Plata
Santiago del Estero Quichua	qus	Quechuan	La Plata
Maca	mca	Matacoan	La Plata
Tapiete	tpj	Tupían	La Plata
Pilagá	plg	Guaicuruan	La Plata
Toba	tob	Guaicuruan	La Plata
Mocoví	moc	Guaicuruan	La Plata
Emérillon	eme	Tupían	Guyanas
Hixkaryána	hix	Cariban	Guyanas
Trió	tri	Cariban	Guyanas
Warao	wba	Warao	Guyanas
Tehuelche	teh	Chonan	Southern Cone
Puelche	pue	Puelche	Southern Cone
Ona	ona	Chonan	Southern Cone
Mapudungun	arn	Araucanian	Southern Cone
Shipibo-Conibo	shp	Panoan	Huallaga-Ucayali
Matsés	mcf	Panoan	Huallaga-Ucayali
Yanesha'	ame	Arawakan	Huallaga-Ucayali
Cholón	cht	Hibito_Cholon	Huallaga-Ucayali
Yaminahua	yaa	Panoan	Huallaga-Ucayali
San Martín Quechua	qvs	Quechuan	Huallaga-Ucayali
Uru	ure	Uru-Chipaya	Central Andes
North Junín Quechua	qvn	Quechuan	Central Andes
Cusco Quechua	quz	Quechuan	Central Andes
Imbabura Highland Quichua	qvi	Quechuan	Central Andes
Huaylas Ancash Quechua	qwh	Quechuan	Central Andes
Ayacucho Quechua	quy	Quechuan	Central Andes
Huangascar-Topara-Yauyos Quechua	qux	Quechuan	Central Andes
Chipaya	cap	Uru-Chipaya	Central Andes
Callawalla	caw	Mixed Language	Central Andes
Pacaraos Quechua	qvp	Quechuan	Central Andes
Cajamarca Quechua	qvc	Quechuan	Central Andes
Southern Aymara	ayc	Aymaran	Central Andes
Huallaga Huánuco Quechua	qub	Quechuan	Central Andes
Jauja Wanca Quechua	qxw	Quechuan	Central Andes
South Bolivian Quechua	quh	Quechuan	Central Andes
Jaqaru	jqr	Aymaran	Central Andes
Salasaca Highland Quichua	qxl	Quechuan	Central Andes
Central Aymara	ayr	Aymaran	Central Andes
Awa-Cuaiquer	kwi	Barbacoan	Pacific
Mochica	omc	Mochica	Pacific
Northern Emberá	emp	Chocoan	Pacific
Colorado	cof	Barbacoan	Pacific
Ninam	shb	Yanomamic	Orinoco
Puinave	pui	Puinave	Orinoco
Panare	pbh	Cariban	Orinoco
Itonama	ito	Itonama	Guaporé-Mamoré
Karo (Brazil)	arr	Tupían	Guaporé-Mamoré
Leco	lec	Leko	Guaporé-Mamoré
Movima	mzp	Movima	Guaporé-Mamoré
Gavião Do Jiparaná	gvo	Tupían	Guaporé-Mamoré
Wari'	pav	Chapacuran	Guaporé-Mamoré

Table 33A.1 (cont.)

Language	ISO code	Genealogical affiliation	Region
Kanoê	kxo	Kanoë	Guaporé-Mamoré
Mosetén-Chimané	cas	Moseten-Chimane	Guaporé-Mamoré
Lakondê	lkd	Nambiquaran	Guaporé-Mamoré
Cavineña	cav	Tacanan	Guaporé-Mamoré
Sabanê	sae	Nambiquaran	Guaporé-Mamoré
Sakirabiá	skf	Tupian	Guaporé-Mamoré
Ese Ejja	ese	Tacanan	Guaporé-Mamoré
Reyesano	rey	Tacanan	Guaporé-Mamoré
Yuracaré	yuz	Yurakare	Guaporé-Mamoré
Baure	brg	Arawakan	Guaporé-Mamoré
Mamaíndé	wmd	Nambiquaran	Guaporé-Mamoré
Yuqui	yuq	Tupian	Guaporé-Mamoré
Karitiána	ktn	Tupian	Guaporé-Mamoré
Kwaza	xwa	Kwaza	Guaporé-Mamoré
Dâw	kwa	Nadahup	Vaupés
Nhengatu	ysl	Tupian	Vaupés
Desano	des	Tucanoan	Vaupés
Tariana	tae	Arawakan	Vaupés
Hupdê	jup	Nadahup	Vaupés
Cubeo	cub	Tucanoan	Vaupés
Bora	boa	Boran	Vaupés
Southern Pastaza Quechua	qup	Quechuan	Napo-Marañon
Urarina	ura	Urarina	Napo-Marañon
Tena Lowland Quichua	quw	Quechuan	Napo-Marañon
Iquito	iqu	Zaparoan	Napo-Marañon
Muniche	myr	Muniche	Napo-Marañon
Shuar	jiv	Jivaroan	Napo-Marañon
Siona-Tetete	snn	Tucanoan	Napo-Marañon
Cocama-Cocamilla	cod	Tupian	Napo-Marañon
Arabela	arl	Zaparoan	Napo-Marañon
Aguaruna	agr	Jivaroan	Napo-Marañon
Cofán	con	Cofan	Napo-Marañon
Bororo	bor	Bororoan	Tapajos-Madeira
Kamaiurá	kay	Tupian	Xingú-Tocantins
Canela	ram	Nuclear Macro-Je	Xingú-Tocantins
Trumai	tpy	Trumai	Xingú-Tocantins
Jamamadí	jaa	Arawan	Solimões
Apurinã	apu	Arawakan	Solimões

References

- Adelaar, Willem (with Pieter Muysken), 2004. *The Languages of the Andes*. Cambridge: Cambridge University Press.
- Adelaar, Willem, 2008. Towards a typological profile of the Andean languages. In Alexander Lubotsky, Jos Schaeken and Jeroen Wiedenhof (eds), *Evidence and Counter-Evidence: Essays in Honour of Frederik Kortlandt*, vol. 2: *General Linguistics*, pp. 23–33. Amsterdam: Rodopi.

- Aikhenvald, Alexandra Y., 2002. *Language Contact in Amazonia*. Oxford: Oxford University Press.
- Birchall, Joshua, 2014a. Verbal argument marking patterns in South American languages. In O'Connor and Muysken (eds), pp. 223–249.
- Birchall, Joshua, 2014b. *Argument Realization in the Languages of South America*. Doctoral dissertation, Radboud University, Nijmegen.
- Bowern, Claire, Patience Epps, Russell Gray et al., 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS ONE* 6 (9): e25195.
- Brackelaire, Vincent and Gilberto Azanha, 2006. Últimos pueblos indígenas aislados en América Latina: Reto a la supervivencia [Last isolated indigenous peoples in Latin America: The challenge of their Survival]. In *Lenguas y Tradiciones Orales de la Amazonía: ¿Diversidad en Peligro? [Languages and Oral Traditions of Amazonia: Diversity in Peril?]*, pp. 313–367. La Habana: Casa de las Américas.
- Campbell, Lyle, Terrence Kaufman and Thomas C. Smith-Stark, 1986. Meso-America as a linguistic area. *Language* 62 (3): 530–570.
- Campbell, Lyle, 2012a. Classification of the indigenous languages of South America. In Lyle Campbell and Verónica Grondona (eds), *The Indigenous Languages of South America: A Comprehensive Guide*, pp. 59–166. Berlin: Mouton de Gruyter.
- Campbell, Lyle, 2012b. Typological characteristics of South American indigenous languages. In Lyle Campbell and Verónica Grondona (eds), *The Indigenous Languages of South America: A Comprehensive Guide*, pp. 259–330. Berlin: Mouton de Gruyter.
- Campbell, Lyle, this volume, Chapter 2. Why is it so hard to define a linguistic area?
- Chang, Will and Lev Michael, 2014. A relaxed admixture model of contact. *Language Dynamics and Change* 4 (1): 1–26.
- Constenla-Umaña, Adolfo, 1991. *Las Lenguas del Area Intermedia: Introducción a su Estudio Areal [The Languages of the Intermediate Area: Introduction to Their Areal Study]*. San José: Editorial de la Universidad de Costa Rica.
- Crevels, Mily and Hein van der Voort, 2008. The Guaporé-Mamoré region as a linguistic area. In Pieter Muysken (ed.), *From Linguistic Areas to Areal Linguistics*, pp. 151–180. Amsterdam: John Benjamins.
- Dahl, Östen, J. Christopher Gillam, David G. Anderson, José Iriarte and Silvia M. Copé, 2012. Linguistic diversity zones and cartographic modeling: GIS as a method for understanding the prehistory of lowland South America. In Alf Hornborg and Jonathan David Hill (eds), *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archaeology, Linguistics, and Ethnohistory*, pp. 211–224. Boulder, CO: University Press of Colorado.
- Derbyshire, Desmond C. and Geoffrey K. Pullum, 1986. Introduction. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds), *Handbook of Amazonian Languages*, vol. 1, pp. 1–28. Berlin: Mouton de Gruyter.

- Dixon, R. M. W. and Alexandra Y. Aikhenvald, 1999. Other small families and isolates. In Robert M. W. Dixon and Alexandra Y. Aikhenvald (eds), *The Amazonian Languages*, pp. 341–381. Cambridge: Cambridge University Press.
- Dunne, Thomas and Leal Anne Kerry Mertes, 2007. Rivers. In Thomas T. Veblen, Kenneth R. Young and Anthony R. Orme (eds), *The Physical Geography of South America*, pp. 76–90. Oxford: Oxford University Press.
- Eriksen, Love, 2011. *Nature and Culture in Prehistoric Amazonia*. PhD Dissertation, Lund University.
- Evans, Nicholas and Stephen C. Levinson, 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32 (5): 429–492.
- van Gijn, Rik, 2014a. The Andean foothills and adjacent Amazonian fringe. In O'Connor and Muysken (eds), pp. 102–125.
- van Gijn, Rik, 2014b. Subordination strategies in South America: Nominalization. In O'Connor and Muysken (eds), pp. 274–297.
- Greenberg, Joseph H., 1987. *Language in the Americas*. Stanford, CA: Stanford University Press.
- Hammarström, Harald and Loretta O'Connor, 2013. Dependency-sensitive typological distance. In Lars Borin and Anju Saxena (eds), *Approaches to Measuring Linguistic Differences*, pp. 337–360. Berlin: de Gruyter Mouton.
- Hammarström, Harald and Tom Güldemann, 2014. Quantifying geographical determinants of large-scale distributions of linguistic features. *Language Dynamics and Change* 4 (1): 87–115.
- van de Kerke, Simon and Pieter Muysken, 2014. The Andean matrix. In O'Connor and Muysken (eds), pp. 126–151.
- Krasnoukhova, Olga, 2012. *The Noun Phrase in the Languages of South America*. Doctoral dissertation, Radboud University, Nijmegen.
- Krasnoukhova, Olga, 2014. The noun phrase: Focus on demonstratives, redrawing the semantic map. In O'Connor and Muysken (eds), pp. 250–273.
- Michael, Lev, Will Chang and Tammy Stark, 2014. Exploring phonological areality in the circum-Andean region using a Naive Bayes Classifier. *Language Dynamics and Change* 4 (1): 27–86.
- Müller, Neele, 2014. Language internal and external factors in the development of the desiderative in South American indigenous languages. In O'Connor and Muysken (eds), pp. 203–222.
- Muysken, Pieter, 2010. Scenarios for language contact. In Raymond Hickey (ed.), *The Handbook of Language Contact*, pp. 265–281. Malden, MA: Wiley-Blackwell.
- Muysken, Pieter, 2012. Contacts between indigenous languages in South America. In Veronica Grondona and Lyle Campbell (eds), *Handbook of South American Historical Linguistics*, pp. 235–258. Berlin: de Gruyter Mouton.
- Muysken, Pieter, Harald Hammarström, Joshua Birchall et al., 2014. The languages of South America: Deep families, areal relationships, and language contact. In O'Connor and Muysken (eds), pp. 299–322.

- Muysken, Pieter, Harald Hammarström, Joshua Birchall, Rik van Gijn, Olga Krasnoukhova and Neele Müller, 2015. Linguistic areas, bottom up or top down? The case of the Guaporé-Mamoré. In Bernard Comrie and Lucia Golluscio (eds), *Language Contact and Documentation*, pp. 205–238. Berlin: de Gruyter.
- Nettle, Daniel, 1996. Language diversity in West Africa: An ecological approach. *Journal of Anthropological Archaeology* 15 (4): 403–438.
- Nettle, Daniel, 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Nichols, Johanna, 1990. Linguistic diversity and the first settlement of the New World. *Language* 66 (3): 475–521.
- Nichols, Johanna, 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nordhoff, Sebastian, Harald Hammarström, Robert Forkel and Martin Haspelmath, 2013. Glottolog 2.1. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://glottolog.org>
- O'Connor, Loretta and Pieter Muysken (eds), 2014. *The Native Languages of South America: Origins, Development, Typology*. Cambridge: Cambridge University Press.
- Orme, Anthony R., 2007. The tectonic framework of South America. In Thomas T. Veblen, Kenneth R. Young and Anthony R. Orme (eds), *The Physical Geography of South America*, pp. 3–22. Oxford: Oxford University Press.
- Payne, Doris, 1990. Morphological characteristics of Amazonian languages. In Doris Payne (ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*, pp. 213–241. Austin, TX: University of Texas Press.
- Seki, Lucy, 1999. The Upper Xingú as an incipient linguistic area. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds), *The Amazonian Languages*. Cambridge: Cambridge University Press, pp. 417–430.
- Tessmann, Günter, 1930. *Die Indianer Nordost-Perus: Grundlegende Forschungen für eine systematische Kulturkunde*. Hamburg: Friedrichsen, de Gruyter and Co.
- Torero, Alfredo, 2002. *Idiomas de los Andes: Lingüística e Historia [Languages of the Andes: Linguistics and History]*. Lima: IFEA and Editorial Horizonte.