River Thinking

Arawakan and Pano-Tacanan in the Upper Amazon Transition Area

Pieter Muysken
Department of Linguistics
CLS, Radboud University
Nijmegen, Netherlands
p.muysken@let.ru.nl

Rik van Gijn
Department of Comparative Linguistics
University of Zurich
Zurich, Switzerland
erik.vangijn@uzh.ch

Abstract—This paper studies whether being present in communities belonging to a particular river system influences the structural make up of languages spoken on the Upper Amazon (UA), combining qualitative and quantitative approaches. The UA has many rivers that spring in the Andes and further east join the Amazon. We have coded 76 languages for 23 features (phonology, syntax). Both phylogeny and river system can be taken into account, and phylogeny is overall a better predictor for the characters studied. However, a number of innovations within specific phylogenies in Arawakan and Pano-Tacanan can be accounted for as influence of the river system where a language is spoken.

Keywords—Amazon, river system, typology, Andes, phonology, syntax, Arawakan, Pano-Tacanan

I. Introduction

Ethnohistorical and anthropological work on diffusion of cultures, peoples, and linguistic features in the Amazon often assume that river systems are the crucial vehicle for spread. Indeed, some of the linguistic areas in South America are defined in terms of river systems, such as the Upper Rio Negro/Vaupés area [1] and the Guaporé-Mamoré [2]. Nonetheless, when we tried to test the influence of river systems for the distribution of noun phrase features for the whole continent [3], we found very few if any effects. Here we want to discuss the same issue with a smaller region, different features, and the combination of qualitative and quantitative approaches, focusing on the Upper Amazon (UA). The UA is characterized by the many rivers that spring in the Andes and which further northeast join to form the great Amazon River. Very little is known about pre-Columbian interethnic contacts in this region, and therefore we must rely on indirect evidence for plausibly reconstructing past interactive networks. Archaeological, anthropological, and linguistic evidence suggests that the languages in the region formed a complex and multi-layered network of interethnic interaction [4], [5], [6], [7].

[7] revealed a broad north versus south division among the UA languages, roughly coinciding with different river systems. In this paper, we further explore the relation between river systems and the distribution of individual linguistic features in the UA. We have sampled as densely as possible, to the extent that available material permitted us to include languages. We also included languages spoken in adjacent Amazonian and Andean area to gain a more complete picture. Based on an approach developed in [6], we have coded (or used existing codes for) 76 languages for 23 features commonly identified as areal features of the Andean and Amazonian areas (see accompanying materials), following the assumption that these are features that spread relatively easily and are therefore more likely to yield traces of contact.

Broadly speaking, the area can be divided into three subareas on the basis of drainage systems: northern UA (the Napo-Marañón drainage region), covering present-day Ecuador, and northern Peru, central UA (the Ucayali-Huallaga drainage region), covering present-day central and south Peru, and southern UA (the Madre de Dios-Mamoré drainage region). Whereas the northern and southern river systems are characterized by a general west-to-east flowing direction, the rivers of the central area mostly flow from south to north. Each major river system consists of smaller, more local river systems, allowing us to zoom in stepwise on different subareas.

II. Genealogical units

In this section we attempt to dig further into the materials by trying to establish per genealogical unit (language family or cluster of families) (a) which languages differ most from the others ones in the family; (b) which of the features coded show most diversity; (c) whether it is possible to attribute specific deviations to other languages or language families spoken within a specific river system. We focus here on two families in the sample: Arawakan (with 15 languages), Panoan (7 languages). We also consider Tacanan in relation to Panoan - a proposed deeper genealogical link [8].

A. Arawakan

Of the five first-order sub-branches of Arawakan mentioned in [9], two (western and southern Maipurean) are represented in our sample; southern Maipurean can be subdivided into Bolivian, Pre-Andine, and Purus. In Table I, the profiles of the Arawakan languages are shown, with for each feature the majority value and the percentage of the total of Arawakan languages in the sample that have that value (unknowns were disregarded). The numbers for the features refer to the features (see accompanying materials).

Note that the Arawakan languages mostly align in terms of their features, although a few features score relatively low (in particular the single liquid phoneme and AN order features). Chamicuro and Yanesha' (the western branch) often deviate from the norm (4 and 5 times, respectively), although in different ways: Chamicuro departs from the majority patterns exclusively in the morphosyntactic realm, whereas Yanesha' deviations are found both in phonology and morphosyntax.

In order to get an idea of the internal cohesion of the family and its sub-branches with respect to the features under investigation, we calculated the coherence coefficient by dividing of the sum of identical answers of pairwise comparisons by the sum of all pairwise comparisons (disregarding unknowns). Results are presented in Table II, first for Arawakan as a whole, then for each of the branches represented in our sample.

As can be seen, the overall cohesion is lower than would be expected on the basis of the different averages (E = .87, based on averages weighted by the number of languages), which suggests that at least some of the differences are between branches rather than within.

B. Pano-Tacanan

Although still undecided, there is a longstanding suspicion that the Panoan and Tacanan form two split-offs from a deep genealogical unit. [10] is the latest argument in the debate, favoring the Pano-Tacanan hypothesis. Most Upper-Amazonian Panoan languages belong to the Pano-Nawa branch of the larger Mainline branch, with Kashibo furthest removed from the others. A similar situation exists for Tacanan, where Cavineña is the most distant cousin of the other Tacanan languages of the sample.

Table III shows that Panoan is about as homogenous as Arawakan with respect to the features investigated (although there are more features with 100% agreement throughout the family). The least stable feature is the number of case markers.

TABLE I. FEATURE PROFILE OF ARAWAKAN SAMPLE LANGUAGES

Feature	Majority	Agr. (%)	Feature	Majority	Agr. (%)
F1	N	86.7	F13	Y	100.0
F2	Y	86.7	F14	Y	100.0
F3	N	93.3	F15	A	93.3
F4	Y	73.3	F16	N	100.0
F5	N	100.0	F17	N	86.7
F6	N	93.3	F18	N	86.7
F7	N	92.9	F19	Y	93.3
F8	N	60.0	F20	N	100.0
F9	A	91.7	F21	N	66.7
F10	N	71.4	F22	N	90.9
F11	N	100.0	F23	N	70.0
F12	N	86.7			

TABLE II. COHERENCE COEFFICIENTS FOR ARAWAKAN UNITS

Overall	Western	Bolivian	Pre-Andine	Purus
.81	.74	.93	.88	.84

Kashibo deviates 4 times (all phonology), Shipibo and Panobo also deviate 4 times. With Capanahua deviating twice, the Chama branch seems to be rather deviant among the Pano-Nawa branches. [11] (p. 11-12) classifies the Mainline languages on the basis of the extent to which they deviate from the Mainline prototype, and comes up with the following order (those branches that do not play a role in this research have been ignored):

Kashibo > Bolivian > Chama > Headwaters

The fact that Chácobo (Bolivian) conforms relatively well to our prototype of the family may be related to its poor documentation status. More research may reveal further discrepancies.

Tacanan is more closely-knit than Panoan (see Table IV), with most features completely homogeneous among all Tacanan languages, the presence of the palatal nasal phoneme and the presence of ideophones as a separate word class are the least homogeneous features. Cavineña and Reyesano deviate most often from the general pattern (each 3 times), followed by Ese Ejja (twice) and Araona (once).

If we look at internal cohesion scores for both families and the tentative higher genealogical unit, the picture as presented in Table V emerges.

TABLE III. FEATURE PROFILE FOR PANOAN SAMPLE LANGUAGES

Feature	Majority	Agr. (%)	Feature	Majority	Agr (%)
F1	Y	100.0	F13	Y	75.0
F2	N	62.5	F14	Y	87.5
F3	N	75.0	F15	C	57.1
F4	N	87.5	F16	Y	100.0
F5	N	100.0	F17	N	100.0
F6	N	75.0	F18	Y	75.0
F7	N	100.0	F19	N	100.0
F8	N	87.5	F20	N	100.0
F9	A	100.0	F21	N	100.0
F10	Y	75.0	F22	N	100.0
F11	N	100.0	F23	N	66.7
F12	N	100.0			

TABLE IV. FEATURE PROFILE FOR TACANAN SAMPLE LANGUAGES

Feature	Majority	Agr. (%)	Feature	Majority	Agr (%)
F1	N	100.0	F13	Y	100.0
F2	Y	100.0	F14	N	100.0
F3	N	100.0	F15	C	80.0
F4	N	60.0	F16	N	80.0
F5	N	100.0	F17	N	100.0
F6	N	80.0	F18	Y	100.0
F7	N	100.0	F19	N	100.0
F8	N	80.0	F20	N	80.0
F9	A	100.0	F21	N	100.0
F10	N	100.0	F22	N	100.0
F11	N	100.0	F23	Y	60.0
F12	N	100.0			

TABLE V. COHERENCE COEFFICIENTS FOR PANO-TACANAN UNITS

Overall	Panoan	Tacanan
.77	.80	.86

Again, the numbers suggest - this time even more clearly - that the variation within each unit is smaller than the variation between the units. In other words: genealogy accounts for a good part of the variation. Interestingly, the two somewhat larger Panoan subgroups of the sample, Chama and Headwaters, show a rather divergent internal cohesion: .71 versus .85, respectively.

Altogether, there seem to be rather strong genealogical signals in the data for the families we have reviewed. In the next section, we will explore the internal structure of the different river systems in more detail.

III. River systems

The area under investigation can roughly be divided into three main river systems. At Iquitos, The Napo and Marañon Rivers join to form the Amazon River. The Marañon river has many tributaries which can be roughly divided into two groups: those coming from the northeast, flowing in a general northeast to south-west direction (Napo, Aguarico, Pastaza, Tigre), and those to the south, which have a general south to north flowing direction (Ucayali, Huallaga). We have treated the systems to the north and south of the Marañon as different river systems, based on this general direction of flow. We term the northern system the Napo-Pastaza system, the southern system the Huallaga-Ucayali system. In southern Peru and Bolivia, the Madre de Dios and Beni Rivers spring in the south-central Andes and join in northern Bolivia to form the Madeira River. The Mamoré River originates in the Cochabamba mountain range and also joins the Madre de Dios and Beni Rivers after their confluence at the border of north Bolivia and Brazil.

Since the object of investigation of this paper is whether being part of one and the same river system facilitates language contact and consequently contact-induced change, it makes sense, on a par with the family descriptions above, to start determining the effects of river systems by calculating the internal cohesion of each river system (Table VI), and to determine the majority pattern for each river system and the degree of agreement for the values of this majority pattern (Tables VII, VII, and IX).

As can be seen in Table VI, the larger river systems show a much lower coherence than the genealogical units. In fact, given that most variables have two possible values, they are not much higher than chance level. One of the reasons for this may be that river systems play a role at a more local level, requiring us to look at smaller river systems. Before we break down the river systems into smaller units, however, we first look at the extent to which a majority pattern is discernible feature by feature. This will give us an idea of the diffusability of individual features within the large-scale river systems.

TABLE VI. Coherence coefficients for major river systems

Overall	Napo-Pastaza	Huallaga-Ucayali	Madre de Dios-Mamoré
59.3	62.7	62.5	58.3

As shown in Tables VII-IX, The highest coherence scores are often achieved for the absence of features. These high-scoring absences are found in all three river systems (features 3, 5, 6, 7, 11, 12, 20). Rather than a role for river systems, this points towards a more general, area wide dispreference for these features. Potentially interesting features for the role of river systems are 13, 15, 16, 17, 19, and 21 (see Table X). These are all morphosyntactic features, whereas the phonological features do not seem to correlate clearly with particular larger river systems.

As Table X shows, there are a few features for which the Napo-Pastaza area behaves differently from the other two, and where it actually behaves similar to the Andean profile. This is partly related to the fact that lowland Quechuan languages, maintaining their suffixing, accusative profile are spoken in that area. The Huallaga-Ucayali area is different in predominantly having languages without an elaborate case-marking system, as well as the presence of classifier/gender systems. These facts are largely attributable to the dominant presence of Arawakan languages in the area.

TABLE VII. Majority patterns in the Napo-Pastaza system

1	Y	.57	9	A	.77	17	Y	.78
2	Y	.57	10	N	.62	18	Y	.67
3	N	.68	11	N	1.0	19	N	.67
4	Y	.57	12	N	.93	20	N	.92
5	N	.96	13	N	.52	21	Y	.76
6	N	.89	14	N	.63	22	N	.61
7	N	.84	15	C	.77	23	N	.53
8	Y	.57	16	Y	.70			

TABLE VIII. Majority patterns in the Huallaga-Ucayali system

1	Y	.63	9	A	.80	17	N	.84
2	Y	.63	10	N	.63	18	N	.63
3	N	.89	11	N	.95	19	Y	.58
4	Y	.58	12	N	.84	20	N	1.0
5	N	.89	13	Y	.84	21	N	.64
6	N	.79	14	Y	.63	22	N	.75
7	N	.89	15	A	.62	23	N	.64
8	Y	.63	16	N	.58			

TABLE IX. Majority patterns in the M. de Dios-Mamoré system

1	N	.70	9	A	.67	17	Y	.50
2	Y	.80	10	N	.66	18	N	.53
3	N	.83	11	N	.70	19	N	.70
4	Y	.60	12	N	.70	20	N	.85
5	N	.77	13	Y	.77	21	N	.56
6	N	.93	14	N	.62	22	N	.68
7	N	.93	15	C	.55	23	N	.70
8	N	.53	16	N	.57			

TABLE X. Potential river-system features

Ft	Description	NP	HU	MM
13	Presence of prefixes	N	Y	Y
15	Elaborate case marking	C	A	C
16	Core case	Y	N	N
17	Accusative alignment	Y	N	N
19	Classifiers/gender	N	Y	N
21	AN order	Y	N	N

All in all, then, there seems to be very little evidence of contact-induced diffusion of the features under investigation along large river systems

The next step is to look at smaller river systems. Each of the three river systems mentioned above contains more than one great river with important tributaries, which can be called river systems in their own right. For the smaller river systems we focus on the northern and southern part of the Upper Amazon, which are genealogically more diverse than the central part. We divide both the northern and southern part into three smaller river systems: Napo-Aguarico, Pastaza-Tigre, and Marañon in the north, Madre de Dios, Beni, and Mamoré in the south. Table XI shows various characteristics of these 6 smaller river systems. The first row below the header row shows the number of sample languages spoken in each river system and between brackets the number of different families they represent. Below that the by now familiar coherence coefficient, followed by the number of features for which there was more than 80% agreement and finally the number of features where all languages behaved the same.

As can be seen in the table, in spite of the genealogical diversity, the coherence coefficients go up and, perhaps more importantly, so do the features for which there is considerable agreement between the languages (over 80%).

TABLE XI. POTENTIAL RIVER-SYSTEM FEATURES

	Napo-Aguarico	**Tigre-Pastaza**	**Marañon**	**M Dios**	**Beni**	**Mamoré**
Langs (fam)	10(5)	7(5)	6(5)	6(4)	10(7)	13(9)
Coherence	.67	.73	.62	.62	.62	.66
.8 agr	11	13	11	14	12	11
1.0 agr	5	10	6	3	0	1

TABLE XII. DIFFERENTIAL AREAL AND GENEALOGICAL CONSISTEENCY

	1	2	3	4	5	6	7	8	9	10	11	12
NA	N	Y	N	-	N	N	N	N	A	N	N	N
TP	Y	N	N	Y	N	N	N	Y	A	-	N	N
MN	Y	Y	N	Y	N	N	N	N	A	N	N	N
MD	N	Y	N	-	N	N	N	-	A	N	N	N
BE	N	Y	N	Y	-	N	N	N	A	N	Y	Y
MM	N	Y	N	Y	N	N	N	Y	A	Y	N	N
AR	N	Y	N	Y	N	N	N	N	A	N	N	N
PN	Y	N	N	N	N	N	N	Y	A	Y	N	N
TC	N	Y	N	N	N	N	N	N	A	N	N	N

	13	14	15	16	17	18	19	20	21	22	23
NA	N	N	C	Y	Y	N	-	N	Y	Y	Y
TP	Y	N	C	Y	Y	Y	N	N	N	N	Y
MN	Y	Y	C	Y	Y	-	N	N	Y	N	N
MD	Y	N	C	Y	N	Y	N	N	N	N	N
BE	Y	N	C	Y	Y	Y	N	N	Y	Y	N
MM	Y	Y	A	N	N	N	Y	N	N	N	N
AR	Y	Y	A	N	N	N	Y	N	N	N	N
PN	Y	N	C	Y	N	Y	N	N	N	N	N
TC	Y	N	C	Y	N	Y	N	N	N	N	Y

0.9-1.0	0.8-0.9	0.7-0.8	0.6-0.7	0.5-0.6

We now turn briefly to the role of individual features. Table XII shows the relative consistency per feature per areal and genealogical unit. As can be seen there are genealogically stable features (1,9,22), features that seem predominantly areal (2,4,23), features with "universal" preference (5,6,7,20). Also there seem to be some local areal odd-ones out (8,11,12,13,14,17,18.21).

IV. CONCLUSIONS

From this brief survey we can conclude first of all that genealogy is the stronger predictor for the features chosen, but genealogical subgrouping is often based on/correlated with river systems. Second, river systems seem to be more of a factor of importance on a smaller scale, although the effect is not a major one. Finally, not all features seem to be equally prone to diffuse over a river system. We have hoped to show that working on a smaller geographical scale than the whole continent does support the earlier insight that river systems help shape the language ecology in the Upper Amazon.

ACKNOWLEDGMENT

We thank Daniël Muysken for thinking along and the audience at the Leiden workshop on Capturing Phylogenetic Algorithms for Linguistics for comments. We also thank our earlier co-authors in the Radboud University Nijmegen Languages in Contact group for discussions and earlier data coding.

REFERENCES

[1] Aikhenvald, A.Y. 2002. *Language contact in Amazonia*. Oxford: OUP.

[2] Crevels, M. & H. van der Voort. 2008. The Guaporé-Mamoré region as a linguistic area. In P. Muysken (ed.) *From Linguistic Areas to Areal Linguistics*, 151-180. Amsterdam / Philadelphia: John Benjamins.

[3] Gijn, R. van, H. Hammarström, S. van de Kerke, O. Krasnoukhova, & P. Muysken (in press) Linguistic areas, linguistic convergence, and river systems in South America. In R. Hickey (ed.) *The Handbook of Areal Linguistics*. Cambridge: CUP.

[4] Renard-Casevitz, F-M, T. Saignes & A. Taylor. 1988. *Al este de los Andes: Relaciones entre las sociedades amazónicas entre los siglos XV y XVII*. Lima: Instituto Francés de Estudios Andinos.

[5] Eriksen, L. 2011. *Nature and Culture in Prehistoric Amazonia: Using G.I.S. to reconstruct ancient ethnogenetic processes from archaeology, linguistics, geography, and ethnohistory*. PhD. Thesis Lund University.

[6] Gijn, R. van. 2014. The Andean foothills and the adjacent Amazonian fringe. In. L. O'Connor & P. Muysken (eds.) *The native languages of South America: origins, development, typology*. Cambridge: CUP: 102-125.

[7] Gijn, R. van & P. Muysken (submitted) Highland-lowland language relations. In A. Pearce, D. Beresford-Jones, & P. Heggarty (eds.) Rethinking the Andes-Amazonia ´Divide´: A Cross-Disciplinary Exploration.

[8] Key, M. 1979. *The grouping of South American languages*. Tübingen: Gunter Narr Verlag.

[9] Hammarström, H., R. Forkel, M. Haspelmath, & S. Bank. 2016. Glottolog 2.7. Jena: Max Planck Institute for the Science of Human History. (http://glottolog.org, Accessed on 2016-02-28.)

[10] Valenzuela, P. & R. Zariquiey. 2014. Body-part nouns, prefixation and incorporation in Panoan and Takanan: some thoughts on a recent debate. Ms., Lima: PUCP.

[11] Fleck, D. 2013, *Panoan Languages and Linguistics*, Anthropological Papers of the American Museum of Natural History #99.