
Comparing morphosyntactic patterns in Katukinan, Harakmbut and Arawan: Genealogical relations or language contact?

— An Van linden & Fabrício Ferraz Gerardi —

The languages

Katukinan family:

Katukina (of Biá)
Kanamari
Katawixí

Harakmbut:

Arasaeri
Amarakaeri/Arakmbut
Sapiteri
Toyeri

Arawan:

Arawa
Banawa
Deni
Jamamadí
Jarawara
Kulina
Paumari
Sorowahá
+2



Introduction

Based on **lexical** similarities:

Adelaar (2000) proposed a genetic link between Harakambut and Katukina.
(53 cognates, 20 of which we deem not convincing)

Adelaar (2007) added Katawixi.

Jolkelsky (2011) added the Arawan family to Adelaar's (2000) proposal.

Morphosyntactic evidence for these hypotheses has been slim and unsystematic.

Introduction & outline

This paper: wide-scope approach comparing morphosyntactic & lexical information

- Compare grammatical descriptions of the three families
- Compare inflectional and derivational formatives from the three families (own database; no conclusions yet)
- Katukinan-Arawan-Harakmbut Database (KAHD)* (Ferraz Gerardi et al. 2022b, In prep) seeking to identify lexical cognates and possible sound-change patterns (see e.g. Ferraz Gerardi et al. 2022a for Tupían)
- Grambank (Skirgård et al. 2023), but corrections needed

→ draw conclusions on genetic affiliation and/or the borrowing of matter or pattern

*Next version adds Apurinã (Arawakan)

Morphosyntax: comparison of grammatical descriptions

Noun phrase:

- Noun classes
- Adnominal possession
- Adjectives

Verb phrase:

- Verbal template
- TAM coding distinction
- Evidentiality
- Noun incorporation

Simple sentence:

- Clause types
- Grammatical relations: head marking
- Grammatical relations: case marking

Complex sentence:

- Clause linking
- Relative clauses

Morphosyntax: comparison of grammatical descriptions

Noun phrase:

- Noun classes
- Adnominal possession
- Adjectives

Verb phrase:

- Verbal template
- TAM coding distinction
- Evidentiality
- Noun incorporation

Simple sentence:

- Clause types
- Grammatical relations: head marking
- Grammatical relations: case marking

Complex sentence:

- Clause linking
- Relative clauses

NP - adnominal possession (human possessors)

	Inalienable possession	Alienable possession
KatKan	Pron. Psor: Psor-Psed Nom. Psor: Psor-CASE=Psed	Pron. Psor: Psor=GenericRelNoun Psed Nom. Psor: Psor-CASE=GenericRelNoun Psed
Har	Pron. Psor: Psor-GEN-Psed OR Psor-GEN NPF-Psed Nom. Psor: Psor-GEN-Psed OR Psor-GEN NPF-Psed	Pron. Psor: Psor-GEN Psed Nom. Psor: Psor-GEN Psed
Ara	Pron. Psor: Psor-Psed Nom. Psor: Psor Psed	Pron. Psor: Psor-Linker Psed Nom. Psor: Psor Linker Psed

NP - adnominal possession (human possessors)

	Inalienable possession		Alienable possession	
	Pron. Psor	Nom. Psor	Pron. Psor	Nom. Psor
KatKan (Queixalós Forthc.)	<i>Yuk-ukpü</i> 1SG-child 'my child' (Katukina)	<i>[[Hanani-na=] okpu]</i> Hanani-CASE= child 'Hanani's child' (Katukina)	<i>nu= wa puaku</i> 2SG= GRN paddle 'your paddle' (Katukina)	<i>[[[Kuntan-na=]NP wa]NP puaku]NP</i> Kuntan-CASE= GRN paddle 'Kontan's paddle' (Katukina)
Har (Van linden Forthc.)	<i>ndoʔ-edn wa-ʔi</i> 1SG-GEN NPF-foot 'my foot'	<i>Lupeʔ-edn wa-ʔi</i> Lupe-GEN NPF-foot 'Lupe's foot'	<i>ndoʔ-edn hak</i> 1SG-GEN house 'my house'	<i>Lupeʔ-edn hak</i> Lupe-GEN house 'Lupe's house'
Ara	<i>u-umahiri</i> 1SG-odor 'my odor' (Suruahá)	<i>Kawakani ahady</i> Kawakani father 'Kawakani's father' (Suruahá)	<i>iri uda</i> <i>his house</i> 'his house' (Suruahá)	<i>jara iri uda</i> jara his house Jara's house (Suruahá)

NP - adnominal possession (human possessors)

	noun/pronoun split	Alienability split
KatKan	✓ (case for nominal possessors)	✓
Har	✗	✗
Ara	✗ (but pronominal possessors are bound)	✓

NP - adnominal possession: possessor markers

	Katukina	Kanamari	Harakmbut	Jarawara	Kulina	Paumari	Suruahá
1.SG	<i>yu(k)-</i>	<i>(h)i-</i>	<i>ndoʔ-edn</i> ¥	<i>o-</i>	<i>o-</i>	<i>o- -na</i>	<i>(aru) na</i>
2.SG	<i>nu-</i>	<i>no-</i>	<i>on-en</i> ¥	<i>ti-</i>	<i>ti-</i>	<i>i- -ni</i>	<i>ara</i>
3.SG	<i>(h)a-</i>	<i>(h)a-</i>	<i>ken-en</i> ¥	<i>hinanka*</i>	<i>po-</i>	<i>-ni</i>	<i>iri</i>
1.PL	<i>tyu-</i>	<i>tso-</i>	<i>oroʔ-edn</i> ¥	<i>ota:</i> (EXCL) <i>e:</i> (INCL)	<i>i-</i>	<i>a- -na</i>	<i>(ari) ja</i>
2.PL	<i>na-</i>	<i>Idiki na-</i>	<i>opudn-en</i> ¥	<i>te:</i>	<i>ti-^{**}</i>	<i>ava- -ni</i>	<i>timia ini</i>
3.PL	<i>ma-</i>	<i>ma-</i>	<i>ken-en</i> ¥	<i>me:</i>	<i>po-^{**}</i>	<i>va- -na</i>	<i>madi iri</i>

*Animate.

**Plural marker omitted.

¥Genitive marker.

NP - adjectives

KatKan: no separate word class → property-denoting elements are verbs or nouns (Queixalós Forthc.)

Har: separate word class (Van linden 2023: 457-458)

Ara: relatively small closed class (Jarawara 14, Suruwahá 3), otherwise verbs are used

Verb Phrase - Morphological template

KatKan:

PERS | APPL | DISLC | (INCORP.N) **STEM** | DIR | REFL | CAUS | NEG | NMLZ/DEP/DUR

“Verbal words usually allow the concatenation of only a small number of affixes (Up to two in naturally occurring discourse).” (Queixalós Forthc.)

Har:

MOOD+AGR | APPL | CLF/INCORP | CAUS.SOC | **STEM** | ASP | TRNS | ASP/AM | AVRT | ASP | TNS | MOOD+GR/MOD/EVID
(Van linden 2023: 461)

Ara (Jar):

O | S/A/CS | PRON/APPL/CAUS | **STEM** | AUX | SUF₅₆ | TM | PRON | SECV | IF | TM/NEG/MOOD (Dixon 2004)

Verb Phrase - Noun incorporation

KatKan (Queixalós Forthc.):

- incorporated noun: placed before the simple verb stem to form a complex verb stem (no intervening element is allowed)
- Only NI Type I (Mithun 1984)
- (also postpositions incorporate into verb → applicatives)

Har:

- IN need not be adjacent to the verb stem (sociative causative marker and/or spatial prefixes may intervene) (Van linden 2023: 461)
- All four types of NI (Mithun 1984) (cf. Rose & Van linden 2022; Van linden 2023)
- only bound nouns get incorporated (with one exception: *hak* 'house') (Rose & Van linden 2022)

Ara: no noun incorporation

Simple sentence - Grammatical relations: head marking

KatKan: ergative alignment (recent pattern) (Queixalós Forthc.)

- intransitive verbs go uninflected for person
- transitive verbs go uninflected for person if the subject is in external NP (Queixalós Forthc)

Har: hierarchical indexation (Van linden 2022)

- all verbs go inflected for person
- transitive verbs showing a hierarchical indexation system in which speech act participant O-arguments trigger relational person prefixes

Ara: different types of alignment depending on grammatical category indexed

- **Person** agreement on verbs is nominative-accusative in all the languages
- **Number** agreement is ergative-absolutive when coded by suppletive verbs in Kulina and Jarawara, with different stems for singular, dual and plural (number category of S/O)
- **Gender** agreement on verbs differs across languages, e.g. determined by discourse factors in Jarawara and Paumarí (A vs. O-cxns), but NOM-ACC in Sorowahá

Simple sentence - Grammatical relations: case marking

KatKan: ergative alignment (recent pattern) (Queixalós Forthc.)

- A-participants realized by external NPs carry ergative marker (pronominal A's: person index on verb)
- S- and O-participants go unmarked

Har: tripartite system (Van linden 2023: 472-473)

- A-participants carry nominative marker: differential A-marking is governed by both animacy and focus
- S-participants go unmarked (but exceptionally nominative-marked)
- O-participants carry accusative marker: animacy-based differential marking

Ara: unary accusative system (except in Paumari)

- O-participants carry accusative marker in Eastern Jamamadi, Kulina-Deni, Paumari, and optionally so in Suruwahá (archaic in Jarawara)
- A-participants carry ergative marker in Paumari (unique!); marking determined by gender agreement on verb

Complex sentence - Relative clause

KatKan (Queixalós Forthc.):

- verb in relative clauses takes suffix *-nin*
- Kanamari: relative clauses are internally headed and introduced by a deictic *nyan*; head typically immediately after introductory deictic (a), but may also remain post-verbally (b)
- Katukina: “post-nominal” relatives; no introductory deictic; obligatory fronting of head (c)

a. *hi-tuhi:k* [***nyan*** *tüküna* *Kadai-na=* *dahü-dyi-nin*]
1SG-see **DEIC** Indian Kadai-CASE= carry-VEN-**DEP**
'I saw the Indian that Kadai brought.' (Kanamari; Queixalós Forthc.)

b. *hi-tuhi:k* [***nyan*** *kadai-na=* *dahü-dyi-nin* *tüküna*]
1SG-see **DEIC** Kadai-CASE= carry-VEN-**DEP** Indian
'I saw the Indian that Kadai brought.' (Kanamari; Queixalós Forthc.)

c. *yu-tiuk* [*dyara* *yu-tuman-nin*]
1SG-know white.man 1SG-kill-**DEP**
'I knew the white man that I killed.' (Katukina; Queixalós Forthc.; his bracketing)

Complex sentence - Relative clause

Har (Van linden 2023: 473-474):

- relative clauses are externally headed (a), (c) or adjoined (b), and verb takes suffix *-niŋ*
- Only introductory deictic when relativized noun phrase (NP-REL) is oblique (c)
- *-niŋ* marked verb takes **ACC** marking when NP-MATRIX ('antecedent') has O-role in matrix (b)

(a) *iʔ-uk-i* *siro* [*opudn-a* *on-yok-me-niŋ*] [NP-REL = object in S-REL]
1_{SG}-search-1._{IND} machete 2_{PL}-NOM 1<>2_{PL}-give-REC._{PST}-REL
'I am looking for the machete that you (PL) gave me' (Van linden 2023: 474)

(b) *wambo-ta* *i-mba-uk-i* [*tareʔ* *men-ok-mbere-me-niŋ-ta*] [NP-REL = subject in S-REL]
boy-ACC 1_{SG}-V._{PL}-search-1._{IND} manioc 3_{PL}>1_{SG}-SPAT:separation-steal-REC._{PST}-REL-ACC
'I am looking for the boys that stole manioc from me.' (Van linden 2023: 470)

(c) *mboerek* *õ-ĩrĩŋ-ã-mě* *widn* [*ken* *toto-ta* *o-arak-me-niŋ*]
man 3_{SG}._{IND}-hide-TRNS-REC._{PST} stone 3/_{DIST} evil.spirit-ACC 3_{SG}._{IND}-kill-REC._{PST}-REL
'The man hid the stone with which he killed the evil spirit.' (Van linden 2023: 474) [NP-REL = instr in S-REL]

Complex sentence - Relative clauses

Ara:

- relative clauses are externally headed (a)-(b) or adjoined (c), and are marked by suffixes on the verb; suffixes differ depending on whether NP-MATRIX is the subject (c) or object (b) in the relative clause (e.g. in Suruwahá):

Kulina:

- (a) awi [shite o-na-e] Ø-watia -pa
tapir shoot.with.arrow 1SG-AUX-REL.F 3-die-HPST
'The tapir I've shot has died.' (Dienst 2014: 247)

Suruwahá:

- (b) mahuny [idia-hady-ka] h<i>awixa-nawaky
tapir shoot.with.arrow-REL_OBJ.M-FOC <1PL>eat.SEQ-1PL.PST.FH
'Then we ate the tapir that I had killed.' (Dienst and Huber: forthcoming)

- (c) índio ahidi jara bus-a ady, [saro-ri] hia-zy]
native son jara steal? grow-REL_SUBJ.M thus-DEIC
'Jara stole the native's son who grew like this.' (Porta and Frank 1996)

The Database - lexical information

Kahd (Katukinan-Arawan-Harakmbut Database)

Ferraz Gerardi et al. (2022) / (2023)

ID	DOCULECT	CONCEPT	VALUE	FORM	TOKENS	COGID	COGIDS	MORPHEMES	BORROWING	NOTE
13527	Amarakaeri	MAIZE	toket	toket	t o k e t	1982	2221	MAIZE		
12475	Apurinã	MAIZE	kemi	kemi	k e m i	105 ⁷	105 ⁷	MAIZE	YES	Loan of Arawan origin
12337	Arasairi	MAIZE	sinke	sinke	s i n k e	1624	1607	MAIZE		
222	Banawa	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		
557	Deni	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		
830	Jamamadi	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		
1183	Jarawara	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		
13406	Katawishi	MAIZE	iniasi	iniasi	i n i a s i	2106	2116	MAIZE		
12111	Kulina	MAIZE	tapa	tapa	t a p a	1502 ²	1483 ²	MAIZE		
12108	Paumari	MAIZE	d3aroa	d3aroa	d3 a r o a	1501 ²	1482 ²	MAIZE		
12107	Paumari_1833	MAIZE	farua	farua	f a r u a	266 ¹⁰	266 ¹⁰	MAIZE		
10677	Proto-Arawan	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		
1954	Sorowaha	MAIZE	kimi	kimi	k i m i	105 ⁷	105 ⁷	MAIZE		

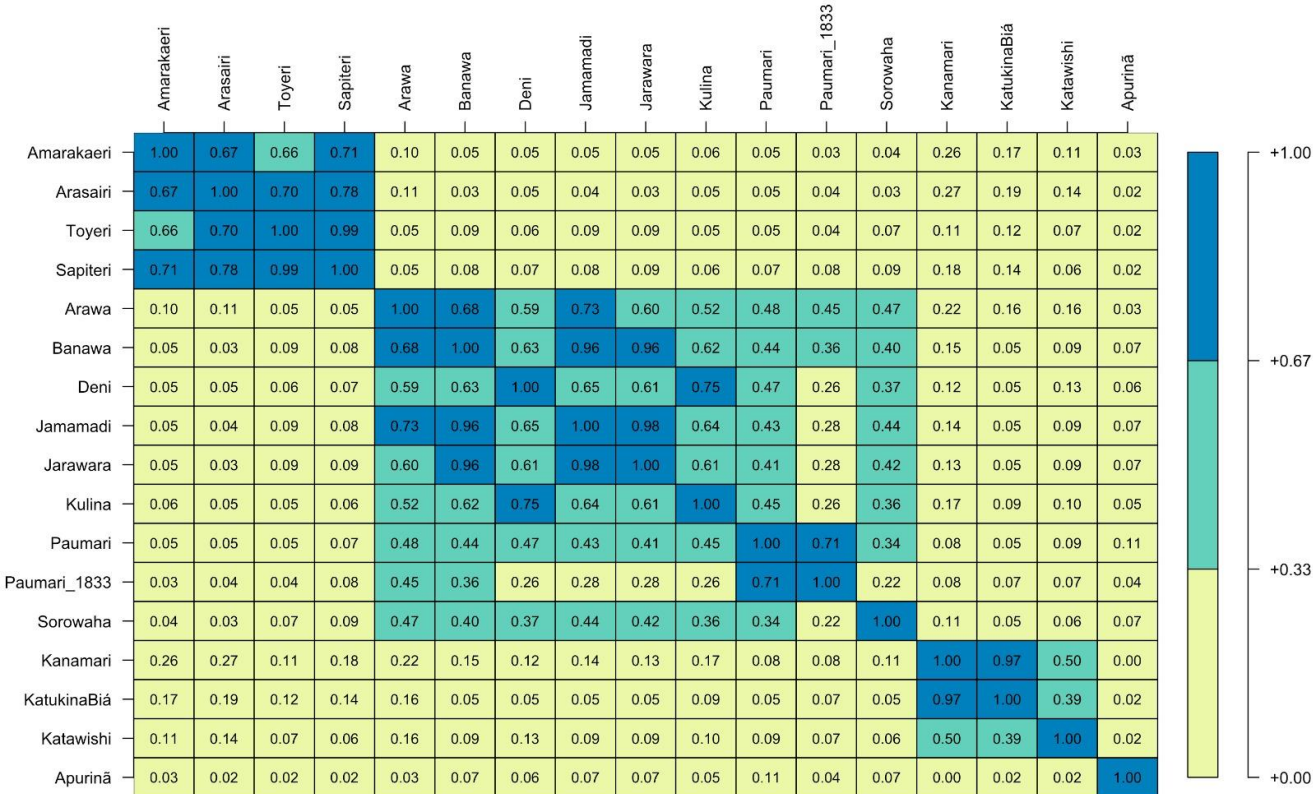
Next release:

Languages: 18

Concepts: 680

Cognacy: simple (root) and partial (semantic), manually annotated

Shared cognacy



Amount of shared cognacy from KAHD (Ferraz Gerardi et al. 2023).

Note how the doculects of well established families/groups are clearly captured.

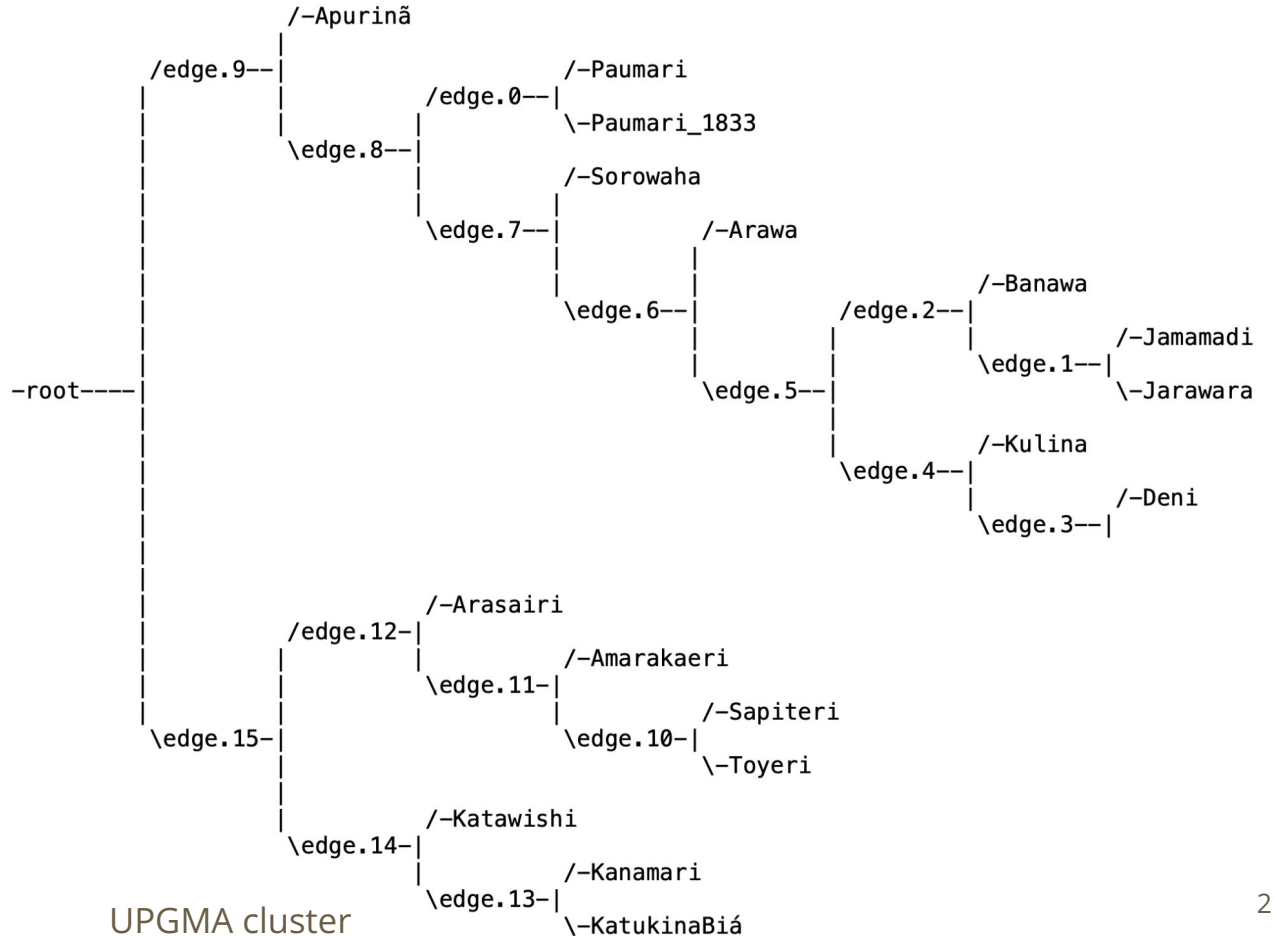
The yellow values, low values, suggest that there is little evidence for a genetic relation among all the doculects. The shared cognacy indicated by these values (yellow) are better interpreted as borrowing, *Wanderwörter*, common Amazonian terms, etc.

Some results

B-Cubed-Scores	
Precision	0.972
Recall	0.877
F-scores	0.922

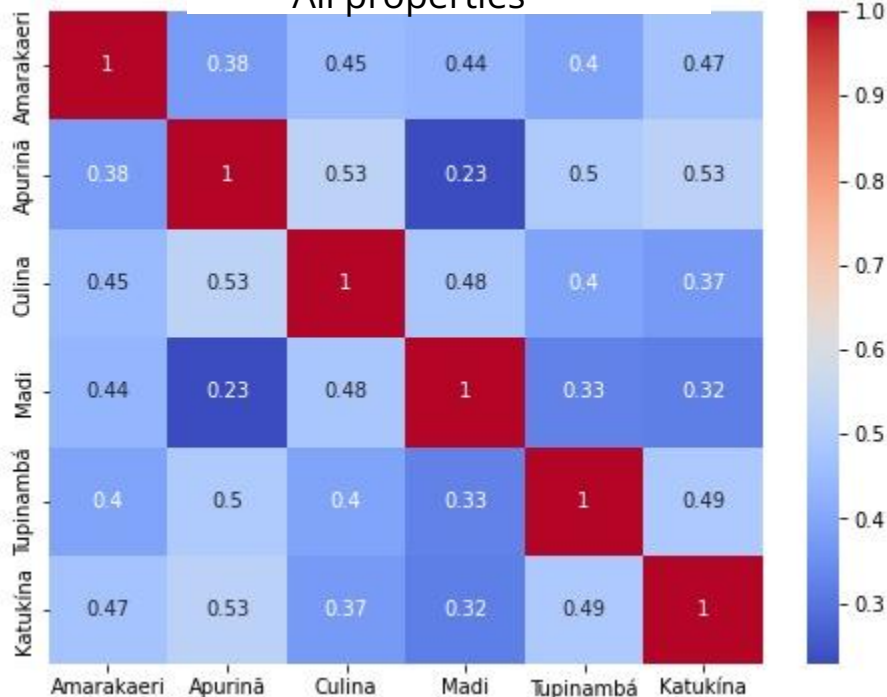
cogid/lexstat (List et al. 2017)

Cognacy diversity: 0.34

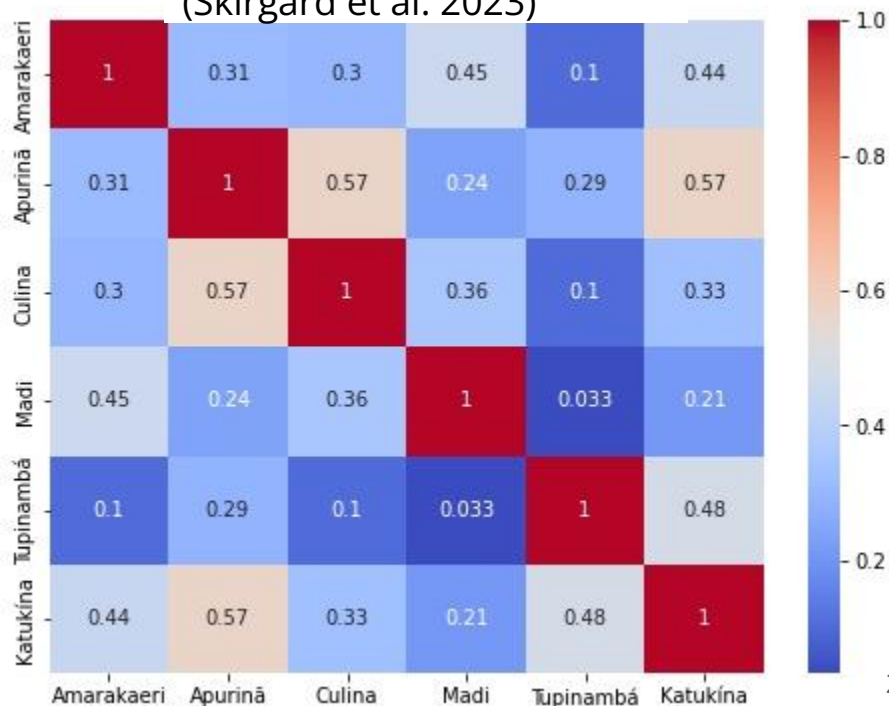


Grambank

All properties

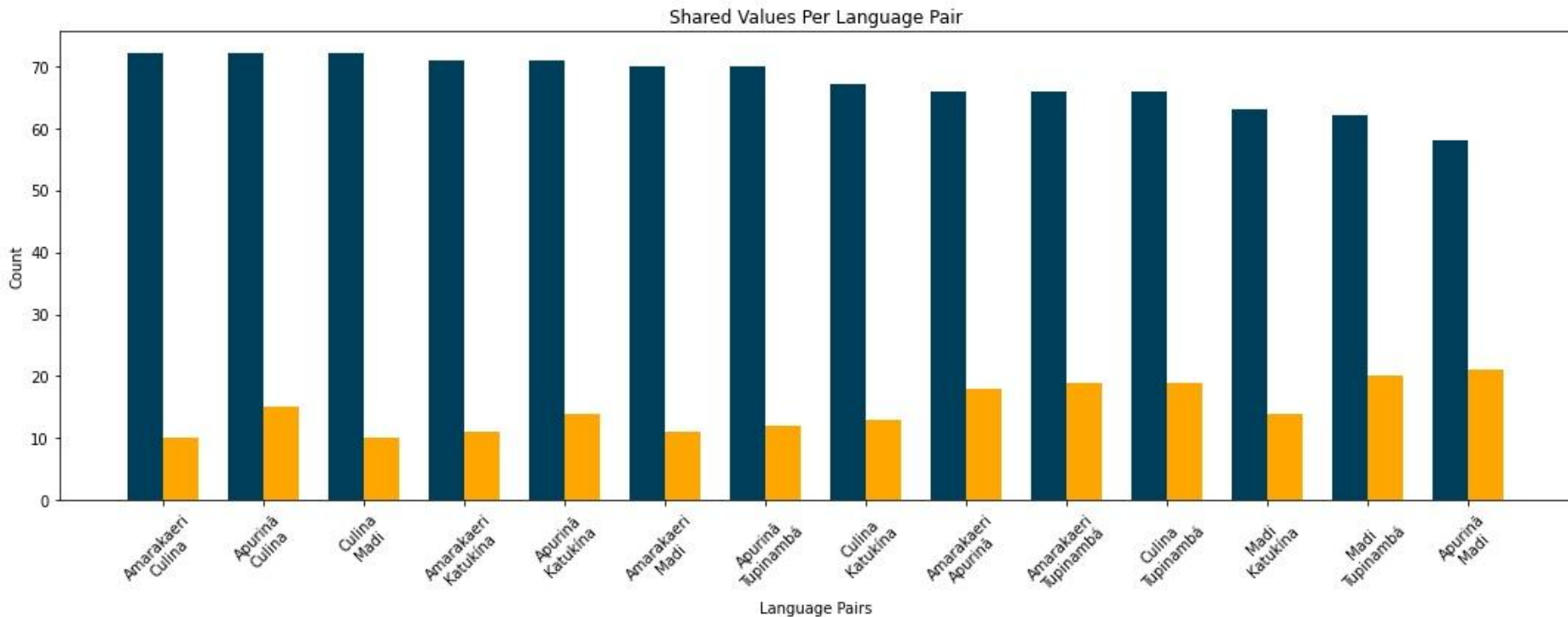


>70% phylogenetic signal
(Skirgård et al. 2023)



Grambank: language pairs

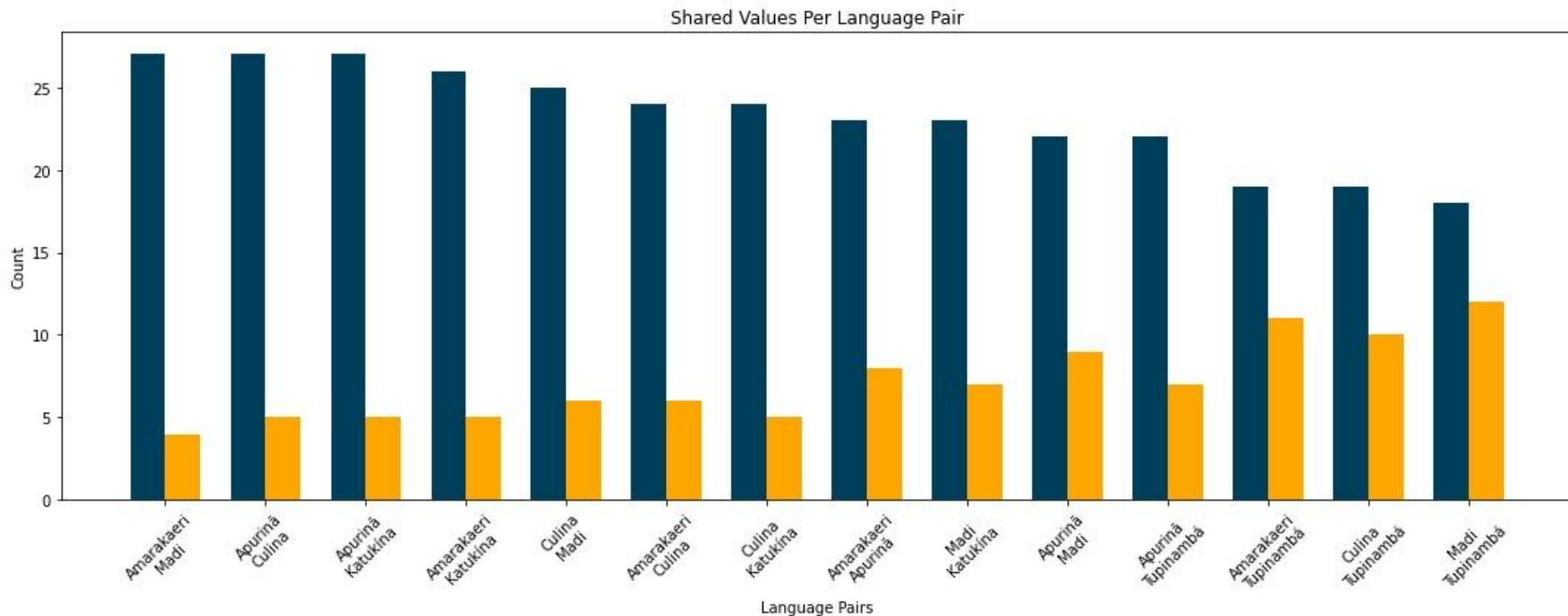
Blue bars: shared values
Yellow bars: distinct values



Grambank: language pairs

Blue bars: shared values
Yellow bars: distinct values

Only properties with phylogenetic signal > 70% (Skirgård et al. 2023)



Grambank

Not useful in its present version due to error in property coding.

Discussion

Preliminary conclusions for **linguistic** evidence:

- Evidence for genetic affiliation is scarce for morphosyntactic properties, scarce for lexical cognacy.
- Similarities in the lexicon are probably due to contact.

Extralinguistic evidence:

- Anthropology: Katukina and Suruwahá are culturally indistinguishable.
- Archeology:
 - Iriarte et al. (2020) suggest strong shift in population in the Brazilian area where the languages examined are located (from Geoglyphs (400 BC–AD 950) to Mound Villages (~AD 1000–1650))
 - Prümers et al.'s (2022) lidar survey reveals pre-Hispanic low-density urbanism in the Bolivian Amazon

References

- Adelaar, Willem F. (2000), Propuesta de un nuevo vínculo genético entre dos grupos lingüísticos indígenas de la Amazonía occidental: Harakmbut y Katukina, in L. Miranda Esquerre (ed.), *Actas del i congreso de lenguas indígenas de sudamérica*, vol. 2, Lima: U. Ricardo Palma, 219–236.
- Adelaar, Willem F. (2007), Ensayo de clasificación del Katawixí dentro del conjunto Harakmbut-Katukina, in A. Romero-Figueroa, A. Fernández Garay and A. Corbera Mori (eds.), *Lenguas indígenas de américa del sur: Estudios descriptivo-tipológicos y sus contribuciones para la lingüística teórica*, Caracas: Universidad Católica Andrés Bello, 159–169.
- Diest, S. and Huber, A. (forthcoming) Arawan, in P. Epps and L. Michael (eds.), *Amazonian Languages, An International Handbook*. Berlin: de Gruyter Mouton.
- Dos Anjos, Z. (2011). *Fonología e gramática katukina-kanamari*. Netherlands Graduate School of Linguistics.
- Hedvig Skirgård, et al. (2023). Grambank v1.0 (Version v1.0.3) [Data set]. Zenodo. doi: 10.5281/zenodo.7844558
- Gerardi, Fabrício. F., Carolina C. Aragon, and Stanislav Reichert (2022), KAHD: Katukinan-Arawan-Harakmbut Database (Pre-release). *Journal of Open Humanities Data*, 8, 18. DOI: <http://doi.org/10.5334/johd.80>
- Gerardi, Fabrício. F., Carolina C. Aragon, Stanislav Reichert, Alan Vogel, and An Van linden (in preparation), KAHD: Katukinan-Arawan-Harakmbut Database (v.1). Previous release available at: [DOI: 10.5281/zenodo.7536378](https://doi.org/10.5281/zenodo.7536378)
- Iriarte, J., Robinson, M., de Souza, J. G., Damasceno, A., da Silva, F., Nakahara, F., ... & Aragao, L. (2020). Geometry by design: Contribution of lidar to the understanding of settlement patterns of the mound villages in SW Amazonia. *Journal of Computer Applications in Archaeology*. 2020; 3 (1): 151-69.
- Jolkesky, Marcelo (2016), Estudo arqueo-ecolinguístico das terras tropicais Sul-Americanas. Unpublished doctoral dissertation. Universidade de Brasília.

References

- Queixalós, Francesc (Forthcoming), Katukinan. In P. Epps and L. Michael (eds.), *Amazonian Languages, An International Handbook*, Berlin: de Gruyter Mouton.
- Porta, E. L. and Frank, J. (1996) *Mitos e histórias dos Sorowaha*. OPAN/CIMI Cuiabá.
- Prümers, H., Betancourt, C. J., Iriarte, J., Robinson, M., & Schaich, M. (2022). Lidar reveals pre-Hispanic low-density urbanism in the Bolivian Amazon. *Nature*, 606 (7913), 325-328.
- Rose, Françoise & An Van linden. 2022. The derivational use of classifiers in Western Amazonia. In Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes*, 237-276. Berlin: De Gruyter Mouton. [<https://doi.org/10.1515/9783110673494-008>].
- Tastevin, Clément (ms.) 'Dialecto Catauishi (Katawaxy).' Archives Générales de la Congrégation du Saint Esprit, Chevilly-la-Rue.
- Van linden, An (2022). Spatial prefixes as applicatives in Harakmbut. In S. Pacchiarotti and F. Zúñiga (eds.), *Applicative morphology: Neglected syntactic and non-syntactic functions*, Berlin: De Gruyter, 129-159. <https://doi.org/10.1515/9783110777949-006>
- Van linden, An (2023), Harakmbut, in P. Epps and L. Michael (eds.), *Amazonian Languages, An International Handbook. Language Isolates, Volume 1: Aikanã to Kandozi-Shapra*. Berlin: de Gruyter Mouton, 441-481. <https://doi.org/10.1515/9783110419405-010>
- Van linden, An. Forthc. When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut. *Linguistics*.
- Vogel, Alan (2006) Jarawara-English dictionary. SIL.

Thank you!

an.vanlinden@uliege.be

fabricio.gerardi@uni-tuebingen.de