

# Data sharing in linguistics: Why and how?

An Van linden

ULiège Open Science Day

7 November 2024

# Outline

Different types of data are shared for different reasons

1. Corpus examples of English and Dutch *as if/alsof* – clauses
2. Sound recordings and annotation files of Harakmbut
3. Questionnaire on binominal lexemes in Harakmbut

Overall impression of the field:

- No habit of sharing data underlying language-specific studies of well-described languages
- Funding agency of ENDANGERED LANGUAGES DOCUMENTATION PROGRAMME (ELDP): obligatory data deposit at ELAR for grantees
- Examples of shared data: either comparative work (often in large databases) or data collected in the field (audio/video recordings in language archives or corpus data from self-compiled corpus)

# 1. Corpus examples of English and Dutch *as if/alsof*-clauses

- FNRS-funded inter-university project “How do grammar and discourse interact? Answers from subordination, coordination and insubordination” (PDR T.0065.20), awarded to Liesbeth Degand (UCLouvain) and An Van linden
- Van Praet, Wout, Liesbeth Degand & An Van linden. Accepted for publication. As if grammar, discourse and prosody don’t interact: A comparative study of [hypothetical manner clauses](#) in English and Dutch. *Folia Linguistica*.

## SUBORDINATE use

(1) *They look at me as if I’m stupid.* (British National Corpus)

(2) *Lilan zwaaide met Jeans ansichtkaart alsof het een waaier was.* ‘Lilan waved Jean’s postcard as if it was a fan.’ (Corpus Gesproken Nederlands)

## INSUBORDINATE use

(3) *As if we haven’t got enough on our plate!* (BNC)

- Syntactic dependence
- Semantic-pragmatic value
- Prosody

# 1. Corpus examples of English and Dutch *as if/alsof*-clauses

- Data: pre-existing corpora with transcripts of spoken data and audio files
  - English: British National Corpus → exhaustive sample (n=225 → after cleaning n=156)
  - Dutch: Corpus Gesproken Nederlands ('Spoken Dutch Corpus') → 250-hit random sample (→ after cleaning n=247)
- Analysis of dataset in spreadsheet: coding data for analytical parameters, e.g. BNC

- **Syntactic dependence**

SyntacticDependency

Omissability

Clefting

Proportionality

Preposability

- **Semantic-pragmatic value**

SemanticPragmaticFunction

- **Prosody**

ProsodicIntegration

PriorBreak

PriorPause

ToneAsIf

ToBIContourAsIf

F0\_Onset

F0\_previousTU

F0\_difference

# 1. Corpus examples of English and Dutch *as if/alsof*-clauses

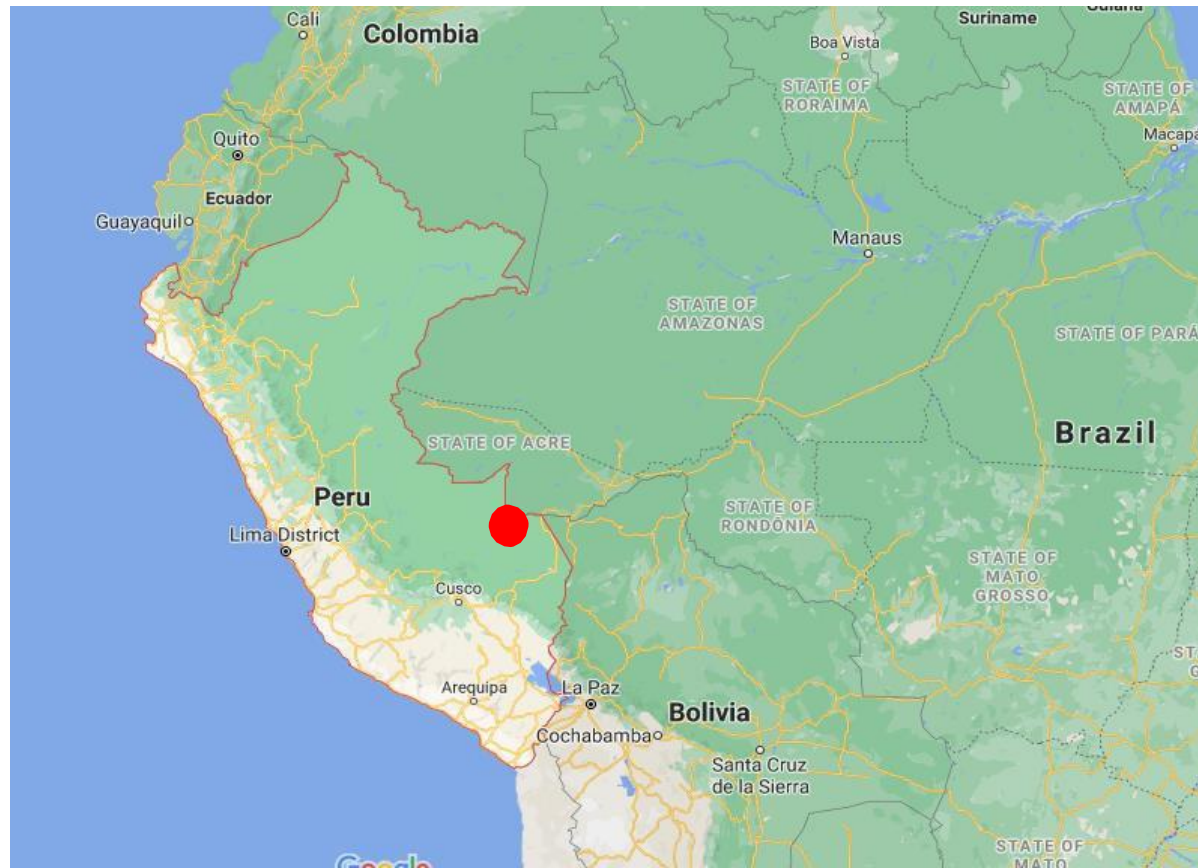
|    | A   | B        | C          | D      | E       | F      | G      | H      | I         | J        | K         | L         | M        | N        | O        |
|----|---|----------|------------|--------|---------|--------|--------|--------|-----------|----------|-----------|-----------|----------|----------|----------|
| 1  | Example   | Syntac   | Seman      | Omissa | Cleftin | Propor | Prepos | Prosod | PriorBr   | PriorPa  | ToneA     | ToBICo    | F0_On    | F0_pre   | F0_diff  |
| 2  | and (pause) but you can't. In nineteen eighty six (unclear) treat peo     | strong   | manner     | no     | yes     | yes    | no     | yes    | 1         | 0        | rise-fall | H*-!H H-H | NA       | NA       | NA       |
| 3  | one (pause) and we have the picture of the cock in the background         | independ | denial     | yes    | no      | no     | no     | no     | 4         | 0,417453 | fall      | H* H-L%   | 169,5814 | 113,9743 | 6,879242 |
| 4  | end of September (unclear)? (SP:J8DPS004) Yeah but they're the we         | independ | denial     | yes    | no      | no     | no     | no     | startturn | 1,157467 | fall      | H* L-L%   | 281,0199 | NA       | NA       |
| 5  | That's close! (pause) Here (pause) Only me about me getting drunk,        | independ | denial     | yes    | no      | no     | no     | no     | 4         | 0        | fall      | H* L-L%   | 227,867  | 170,3243 | 5,038899 |
| 6  | ! (SP:PS02G) (cough) And you wouldn't be? (SP:PS03M) No no. (SP:P         | independ | denial     | yes    | no      | no     | no     | no     | 4         | 0        | fall      | H* L-L%   | 178,0756 | 148,5566 | 3,137727 |
| 7  | in the process of doing. (SP:PS05H) Yet another flower. (SP:PS05B) M      | independ | denial     | yes    | no      | no     | no     | no     | startturn | 1,14189  | fall      | H* L-L%   | 280,0843 | NA       | NA       |
| 8  | I'm never sure if they eat this weed cos it doesn't look as if it's going | strong   | qualificat | no     | yes     | yes    | no     | yes    | 1         | 0        | rise-fall | L*+H L    | NA       | NA       | NA       |
| 9  | bit of both. (SP:PS087) I am chickening out cos I don't feel as if I've ( | strong   | qualificat | no     | yes     | yes    | no     | yes    | 1         | 0        | fall      | H* L-L%   | NA       | NA       | NA       |
| 10 | and tell them that it's a no-smoking area and they look at me as if I'r   | weak     | manner     | yes    | yes     | yes    | no     | yes    | 1         | 0        | fall      | H* L-L%   | NA       | NA       | NA       |
| 11 | mean I've never known problem before. (SP:PS6ST) No. (SP:PS0E8) s         | strong   | qualificat | no     | yes     | yes    | no     | yes    | 1         | 0        | rise-fall | L*+H L-   | NA       | NA       | NA       |
| 12 | 's nothing to do with that! And she goes, oh yeah, as if you'd know!      | independ | denial     | yes    | no      | no     | no     | no     | 4         | 0        | fall-rise | H* L-H%   | 253,0733 | 188,8034 | 5,072047 |
| 13 | a good bit on there and soaked it (pause) let it soak in. (pause) as if   | independ | denial     | yes    | no      | no     | no     | no     | 4         | 4,090651 | fall-rise | H* L-H%   | 196,7049 | 251,4427 | -4,25036 |
| 14 | is stupid this is, it really is daft (SP:PS13K) why? (pause) (SP:PS0KN)   | independ | denial     | yes    | no      | no     | no     | no     | startturn | 2,900023 | fall-rise | H* L-H%   | 255,5814 | NA       | NA       |
| 15 | just a number on a damn computer but they crack on (SP:PS50U) Yes         | weak     | manner     | yes    | yes     | yes    | yes    | no     | 4         | 0,58511  | fall      | H* L-L%   | 162,1667 | 189,4184 | -2,68918 |
| 16 | in trouble. So every pitchfork used to come out er one after anothe       | independ | qualificat | yes    | no      | no     | no     | no     | 4         | 0,366984 | fall      | H* L-L%   | 137,3976 | 139,5731 | -0,27197 |
| 17 | put his own clothes on. (pause) And er I b-- I remember him as if it      | weak     | manner     | yes    | yes     | yes    | no     | yes    | 1         | 0        | rise-fall | L*+H L-L% | NA       | NA       | NA       |
| 18 | cemeteries then (pause) and carrying a wreath of er flowers, I reme       | weak     | manner     | yes    | yes     | yes    | no     | yes    | 1         | 0        | fall      | H* L-L%   | NA       | NA       | NA       |
| 19 | 've got in some instances. You make commitments for next year, as         | weak     | manner     | yes    | yes     | no     | yes    | no     | 4         | 0,707013 | fall      | H* H-L%   | 172,0945 | 190,8096 | -1,7872  |
| 20 | going through investigators rather than (SP:PS421) Well (unclear) th      | weak     | manner     | yes    | yes     | no     | yes    | no     | 4         | 0        | rise      | L* L-H%   | 83,19698 | 114,8551 | -5,58255 |
| 21 | , please, give me a literal translation and then an idiomatic one, as i   | independ | manner     | yes    | no      | no     | no     | no     | 4         | 0        | fall-rise | H* L-H%   | 160,8565 | 147,0762 | 1,550517 |

# 1. Corpus examples of English and Dutch *as if/alsof*-clauses

- Why share?
  - Not required by journal
  - But replicability concerns + inviting further research
- What?
  - Csv-files with datasheets (transcripts of corpus data + values for distinct analytical parameters)
  - ~~Audio files~~
  - README file in PDF
  - Note: we were the **first** to ask compilers of Spoken Dutch Corpus to share data
  - BNC: BNC User Licence provides for sharing of corpus data → we rely on the fair dealings provision of copyright law
- Where?
  - ULiège Dataverse (currently under review; checking copyright issues)

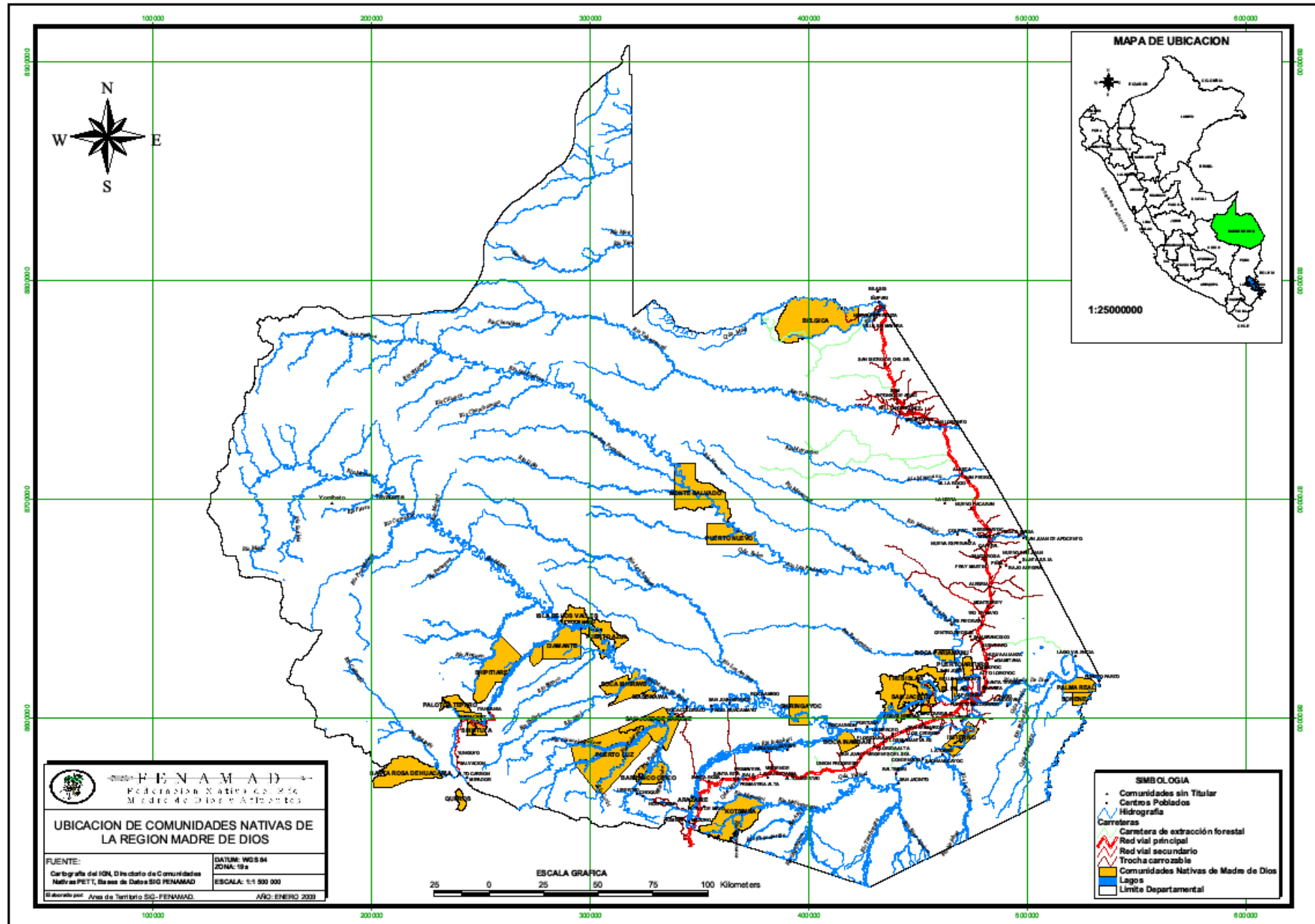
## 2. Sound recordings and annotation files of Harakmbut

- As of 2010: documenting and describing Harakmbut, isolate language spoken in the Peruvian Amazon → linguistic fieldwork of critically endangered indigenous language

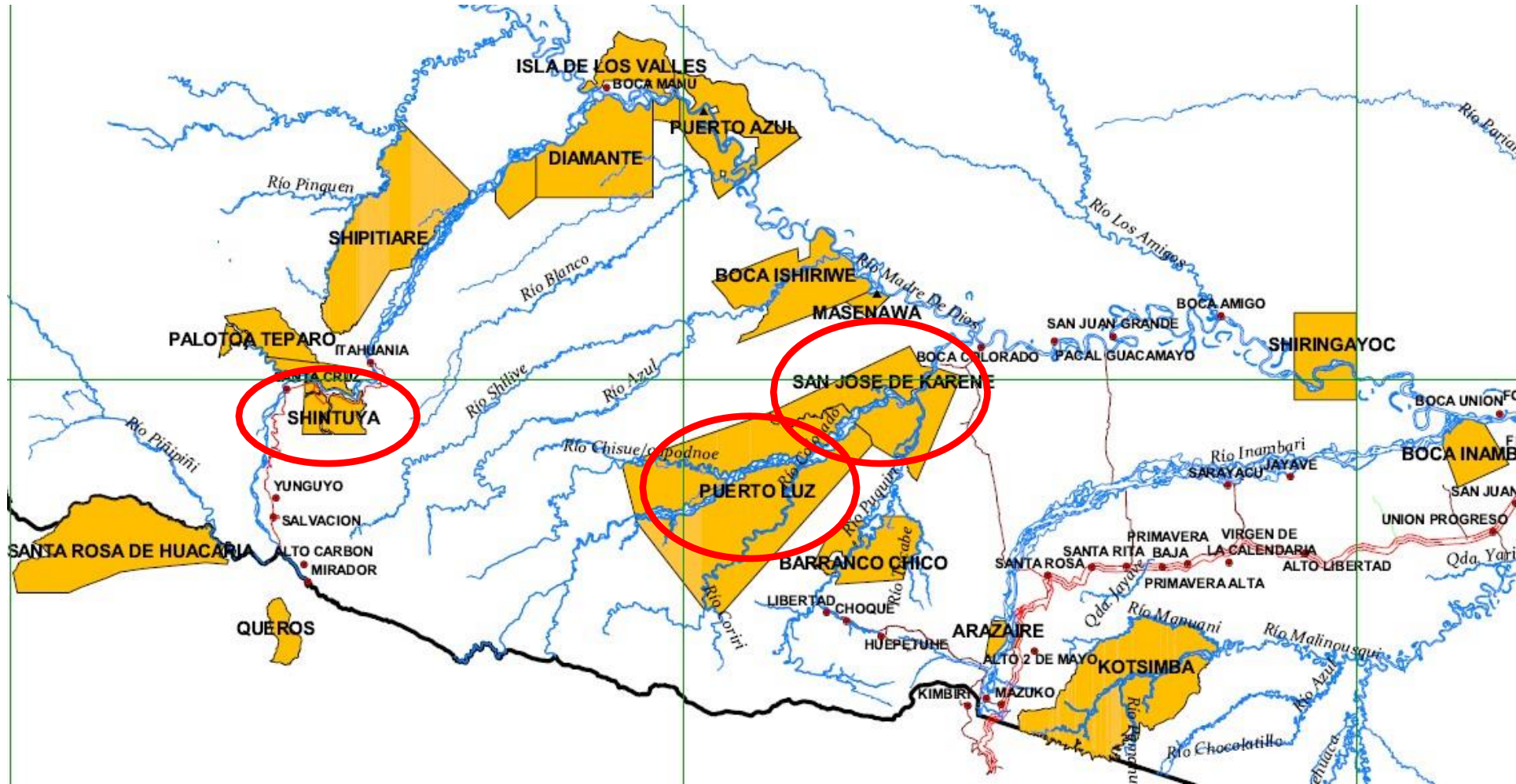




- Harakmbut live in 'native communities': patches of land entitled to them by the government
- subtropical climate
- around tributaries of the Madre de Dios River, which eventually flows into the Amazon River;








- About 1000 speakers left; distinct dialects
- Previous linguistic work: focus on Arakmbut/Amarakaeri dialect (Hart 1963; Helberg 1984, 1990; Tripp 1976ab, 1995); currently I am the only linguist working on the language
- Fieldwork in Puerto Luz, San Jose de Karene and Shintuya → Arakmbut/Amarakaeri variety



A wide river with a rocky bed and a sandy bank. Several people are wading in the water, some standing on mats or nets. In the foreground, a person in a blue shirt and a striped head covering sits on the rocky shore, looking towards the water. The background shows a dense line of trees under a blue sky with light clouds.

Communal fishing activity in the  
Harakmbut native community of San José  
del Karene,  
Madre de Dios, Peru, 2011









Translating Harakmbut recordings with bilingual speaker, community of Puerto Luz, Madre de Dios, Peru, 2016

## 2. Sound recordings and annotation files of Harakmbut

- Primary data collected in the field: mainly recordings, a few videos
- Secondary data: annotation files including the segmentation, transcription and analysis of primary data
- By and large two types of data in terms of content:
  - Oral literature (myths, fables, stories about jaguars, history of communities, biographies), representing the cultural heritage of the Harakmbut  
→ useful mainly for Harakmbut communities themselves (preserving culture)
  - Elicitation data on vocabulary and grammar: Harakmbut translation of Spanish lexemes or ‘invented’ sentences (paradigms of pronouns and verb forms, system of subordination, etc)  
→ useful mainly for scientific community (linguists, anthropologists, biologists, archaeologists)
- Currently in the process of archiving my data with the [California Language Archive](#) (primary data: 346 wav files; 11 AVI files, but also secondary data in txt files, ELAN files)
- 2023: I obtained permission from all speakers (n=41) to make the recordings publicly available (now still pre-archiving phase)




# Archiving recordings and annotation of my language documentation and description project at <https://cla.berkeley.edu/>


The screenshot shows the California Language Archive website. The browser address bar displays <https://cla.berkeley.edu>. The website header includes the title "California Language Archive" and navigation links: HOME, MAP SEARCH, BROWSE CLA, RESOURCES, ABOUT US, and DEPOSITORS. Below the header is a grid of portrait photos of individuals. A search filter overlay is active, titled "Arak", with the following options:

- Filter by language:
  - Arakbut
  - Arakmbut
- Filter by item:
  - Arak

A search button labeled "Search CLA" is located at the bottom right of the filter overlay. At the bottom of the page, a footer text reads: "The California Language Archive is a physical and digital archive for materials related to the Indigenous languages of the Americas, housed".

## Primary contributions? File bundles?


 An Van linden (donor)

 An Van linden (researcher)


[2024-11.001](#) El tigre que mató al abuelo de Maribel / The jaguar who killed Maribel's grandfather (07 Aug 2010 to 12 Aug 2011)


## Controlled vocabulary


Contributions you can add


 Alfonso Mekeliwa (consultant)


 Ameriko Aika (consultant)


 Aquilino Burapa Subia (consultant)


 Armando Carase Mabirey  
(consultant)


 Artemio Sikeng (consultant)


 Bernardo Sonke Dávila (consultant)


 Delia Mackavi (consultant)

 Dilcia Arique Moque (consultant)


 Dominga Maca Aniri (consultant)


 Efraín Manuqueue Coreche  
(consultant)

 Elías Arbildo Ojeda (consultant)


 Emilio Ayma Anahija (consultant)

 Eries Kodnaro Elimuna (consultant)

 Fermín Chimatani Tayori  
(consultant)

 Guadalupe Tayori Kendero  
(consultant)

 Jonas Arique Moque (consultant)

 Jonas Arique Moque (translator)

[2024-11.002](#) La esposa que engañó al cazador / The wife who cheated on the hunter (07 Aug 2010 to 10 Aug 2011)

[2024-11.003](#) La guerra entre los Tacas y los Arakmbut / The war between the Taca and the Arakmbut (07 Aug 2010)

[2024-11.004](#) La hija que se perdió en el agua / The daughter who got lost in the river (07 Aug 2010 to 25 Aug 2011)

[2024-11.005](#) Los Tacas atacaron a los Keroseris / How the Taca attacked the people of Queros (07 Aug 2010)

[2024-11.006](#) Los Tacas atacaron a los niños y mujeres Arakmbut / How the Taca attacked the children and women of the Arakmbut (07 Aug 2010)

[2024-11.007](#) El tigre que orinó al fuego / The jaguar that urinated on the campfire (07 Aug 2010)

[2024-11.008](#) El tigre que atacó una choza / The jaguar who attacked a hut (14 Aug 2010 to 12 Aug 2011)

[2024-11.009](#) Kumamin en la choza de caza / Kumamin in the hunting hut (14 Aug 2010 to 30 Aug 2011)

[2024-11.010](#) La tortuga que mordió a la sachavaca / The turtle that bit the tapir (14 Aug 2010 to 09 Aug 2011)

[2024-11.011](#) Nombre y edad de Yoma / Asking Yoma for her name and age (14 Aug 2010)

[2024-11.012](#) El hombre de la lombriz cuyo hijo sabía secar el agua / The worm man whose son could dry a pond (14 Aug 2010 to 13 Sep 2011)

[2024-11.013](#) El tigre que mató a las mujeres / The jaguar who killed the women (14 Aug 2010 to 09 Aug 2011)



## 2. Sound recordings and annotation files of Harakmbut

- Two collections: Arakmbut Texts & Elicitation
- Interface of CAL might pose problems for Harakmbut speakers to find the recordings; suggestions for ‘shortcuts’?

### 3. Questionnaire on binominal lexemes in Harakmbut

- Initial purpose of these data: providing data for Pepper's (2020) cross-linguistic study on binominal lexemes (= lexemes that consist of two nouns, e.g. *railway*, *spoorweg*, *chemin de fer*)
  - Pepper's list of 100 complex concepts → 78 translation equivalents in Harakmbut for 76 entries & glosses
  - Clear instructions from Pepper for datasheet
- Subsequent use: (partial) dataset underlying my own publications:
  - Rose & Van linden (2022) on the derivational use of classifiers in Mojeño Trinitario and Harakmbut (e.g. *aymõrõ-po* [honey-CLF:sphere] 'bee')
  - Van linden (2023) on adnominal possession in Harakmbut (e.g. 'my hand', 'my pot'), with focus on bound nouns

→ Single data deposit

Initial purpose: data for Pepper (2020)

|    | A                  | B             | C                               | D  | E                             | F           |
|----|--------------------|---------------|---------------------------------|--|-------------------------------|-------------|
| 1  | CASTELLANO         | INGLES        | HARAKBUT translation equivalent | Gloss (complex words only)   | Notes                         | AVL         |
| 2  | tobillo            | ankle         | wa'ikope                        | wa.'i.ko.pe [NMLZ.foot.shape_half_moon.CLF:disc]                                     |                               |             |
| 3  | aurora boreal      | arctic lights | wandak kurudn, u'ruwatayokmb    | wa.ndak kurudn [NMLZ.good heaven], u'ruwatayokmba                                    |                               |             |
| 4  | mochila            | backpack      | amikoen kutamah                 | amiko.en kutamah [foreigner.GEN bag]   | amiko: from                   | Kutama: bo  |
| 5  | abeja              | bee           | sērāpo, aymorōpo                | sērāpo, aymorō.po [honey.CLF:round]  |                               | sērōā'pō: e |
| 6  | colmena            | beehive       | aymorō'onpe                     | aymorō.'on.pe [honey.?.CLF:disc]   | doublecheck                   | colmena: a  |
| 7  | cera de abejas     | beeswax       | korön                           | -  | this is the w                 | cera de abe |
| 8  | bicicleta          | bicycle       | ....                            | ....   |                               |             |
| 9  | bomba de bicicleta | bicycle pump  | ....                            | ....   |                               |             |
| 10 | herrero            | blacksmith    | sirowakeri                      | siro.wa.ka.eri [iron;metal.NMLZ.make.AN]   |                               |             |
| 11 | chico              | boy           | wambosipo                       | wambo.si.po [youngster.peel.CLF:round]   | the bound noun -sipo [p       |             |
| 12 | pulsera            | bracelet      | sakmbi wambatamehte wamed       | sakmbi wa.mba.ta.meh.te wa.mbedn.nda [beaded_necklace NMLZ.hand.base.hump.LOC NMLZ.r | lit. 'a red be; sakbij: colla |             |
| 13 | desayuno           | breakfast     | sikyoedn aypo embape            | sik.yo.edn aypo e.mba.pe [black.LOC.GEN food NMLZ.VPL.eat]                           | lit. 'food eaten in the m     |             |
| 14 | carpintero         | carpenter     | wamationkohkoh'eri              | wa.ma.tionkohkoh'.eri [NMLZ.VPL.?.AN]  |                               |             |
| 15 | cacique            | chieftain     | o'po                            | -  |                               |             |
| 16 | gallo              | cock/rooster  | wāka'ra                         | -  |                               | [nothing in |
| 17 | clavícula          | collarbone    | wambagnkutope'ai                | wa.mbagn.ku.to.pe'.ai [NMLZ.shoulder.head.down.CLF:disc.bone]                        |                               |             |
| 18 | cocina             | cookhouse     | wamanwadna hak                  | wa.manwadna hak [NMLZ.cook_in_pots house]  | lit. 'a house i e.man.wad     |             |
| 19 | vaca lechera       | dairy cow     | mbaka önwe'et                   | mbaka önwe'et [cow ?]  | mbaka: from Spanish va        |             |
| 20 | cena               | dinner        | sikyo aypo embape               | sik.yo aypo e.mba.pe [black.LOC food NMLZ.VPL.eat]                                   | lit. 'food eaten in the da    |             |
| 21 | iamba de puerta    | doorpost      | akwikaimbozne                   | akwikai.mbozn.pe [door.lip:edge.CLF:disc]  |                               |             |

# Pepper (2020: 479)

(Cariban) Guianan

## Galibi Carib (12): 65W / 20B / 7C

Base.DEV (1)

*kamisa.npo* [piece\_of\_tissue.DEV] HANDKERCHIEF OR RAG

Base.DIM (3)

*kawale.menpo* [horse.DIM] FOAL OR COLT

*kapilita.menpo* [goat.DIM] KID

*kapala.menpo* [sheep.DIM] LAMB

Mod Head (6)

*wano.weti* [bee excrement] BEESWAX

*wato.aponi* [fire place] FIREPLACE

*ime.noti* [son grandmother] MOTHER-IN-LAW (OF A MAN)

*moyowai.etaweini* [spider web] SPIDER WEB

*aina.yumi* [hand father] THUMB

*wewe.unti* [tree base] TREE TRUNK

Mod Head.POSS (6)

*pana.sepele.li* [ear lobe.POSS] EARLOBE

*manati.poti.li* [breast tip.POSS] NIPPLE OR TEAT

*aina.la.li* [hand flat.POSS] PALM OF HAND

*upupo.kuwai.yi* [head calabash.POSS] SKULL

*pupu.sikili.li* [foot little\_end.POSS] TOE

*i'mo.kami.li* [egg yellow.POSS] YOLK

Mod.Head (2)

*amek.unti* [wrist.base] BRACELET

*emu(lu).pipo* [eye.skin] EYELID

Mod.POSS Head (1)

*-pana.li.weti* [ear.POSS dirtiness] EARWAX

Mod.POSS Head.POSS (1)

*emo.li.sakila.li* [nose.POSS aperture.POSS] NOSTRIL

*wa.kal.i* [grape tree.SPEC] VINE

Harakmbut

## Harakmbut (3): 76W / 28B / 27C

Base.CLF (2)

*aymörö.po* [honey.CLF(round)] BEE

*sērā.po* [honey.CLF(round)] BEE

Base.DIM (2)

*wambo.sipo* [youngster.DIM] BOY

*wettonē.sipo* [woman.DIM] GIRL

Base.SPAT.CLF (2)

*wa.kmbere.ku.to.pa*

[NPOS.forehead.CLF(edge?).SPAT(down).CLF(ROD)]

EYEBROW

*wa.kpo.ku.ti.mba*

[NPOS.eye.CLF(edge?).SPAT(on).CLF(hand)] EYELID

Head.LOC.Mod (1)

*nö.pö.te.menoe* [vital\_centre.CLF(round).LOC.day] MIDDAY

Mod.GEN Head (1)

*amiko.en.kutamah* [foreigner.GEN bag] BACKPACK

Mod.Head (18)

*wa.'i.ko.pe* [NPOS.foot.shape\_half\_moon.CLF(disc)] ANKLE

*aymörö.'on.pe* [honey.##.CLF(disc)] BEEHIVE

*akwikai.mbogn.pe* [door.lip/edge.CLF(disc)] DOORPOST

*wa.pe.sodn* [NPOS.ear.flappy\_object] EARLOBE

*wa.pe.sodn'.oro* [NPOS.ear.flappy\_object.gold] EARRING

*wa.pe.oktasörö* [NPOS.ear.viscose\_substance] EARWAX

*wa.kpo.ku.ti.mba.wih*

[NPOS.eye.CLF(edge?).SPAT(on).CLF(hand).hair] EYE-

LASH

Subsequent use: dataset underlying own publications (taking issue with Pepper's (2020) analysis)

|    | A                  | B             | C                                | D   | E                                       | F                     | G                    | H                  | I      | J    | K           |
|----|--------------------|---------------|----------------------------------|---|---|-----------------------|----------------------|--------------------|--------|------|-------------|
| 1  | Meaning_CASTELLAN  | Meaning_INGLE | HARAKMBUT translation equivalent | Gloss (complex words only)  | Notes                                   | AVL                   | analysis into classe | type c             | type c | coun | Suptype o   |
| 2  | tobillo            | ankle         | wa'ikope                         | wa.'i.ko.pe [NMLZ.foot.shape_half_moon.CLF:disc]                            |   |                       | NMLZ-BN-BN-CLF       | clf                |        | yes  | Mod.Head    |
| 3  | aurora boreal      | arctic lights | wandak kurudn, u'ruwatayokmba    | wa.ndak kurudn [NMLZ.good heaven], u'ruwatayokmba                           |   |                       | ADJ_N                | descriptive phrase |        | no   |             |
| 4  | mochila            | backpack      | amikoen kutamah                  | amiko.en kutamah [foreigner.GEN bag]  | amiko: fr                               | Kutama: N-GEN_N       |                      | gen                |        | yes  | Mod.GEN H   |
| 5  | abeja              | bee           | sëräpo, aymoröpo                 | aymorö.po [honey.CLF:round]   |   | sëröä'pö N-CLF        |                      | clf                |        | yes  | Base.CLF    |
| 6  | abeja              | bee           | sëräpo, aymoröpo                 | sërä.po [honey.CLF:round]   |   | sëröä'pö N-CLF        |                      | clf                |        | yes  | Base.CLF    |
| 7  | colmena            | beehive       | aymorö'onpe                      | aymorö.'on.pe [honey.?.CLF:disc]  |   | colmena N-SPAT?-CLF   |                      | clf                |        | yes  | Mod.Head    |
| 8  | cera de abejas     | beeswax       | korön                            | [T1995: 146a: körön]  | this is th                              | cera de æ N           |                      | simple             |        | no   |             |
| 9  | bicicleta          | bicycle       |                                  |   |   |                       |                      |                    |        |      |             |
| 10 | bomba de bicicleta | bicycle pump  |                                  |   |   |                       |                      |                    |        |      |             |
| 11 | herrero            | blacksmith    | sirowakeri                       | siro.wa.ka.eri [iron;metal.NMLZ.make.AN]                                    |   |                       | N-NMLZ-V-AN          | V-based            |        | no   |             |
| 12 | chico              | boy           | wambosipo                        | wambo.si.po [youngster.peel.CLF:round]                                      | the bound noun -s                       | N-DIM                 |                      | der                |        | yes  | Base.DIM    |
| 13 | pulsera            | bracelet      | sakmbi wambatamehte wambednd     | sakmbi wa.mba.ta.meh.te wa.mbedn.nda [beaded_necklace NMLZ.hand.base.hump.L | lit. 'a rec sakbij: c                   | N_NMLZ-BN-SPAT-BN-LO  |                      | descriptive phrase |        | no   |             |
| 14 | desayuno           | breakfast     | sikyoedn aypo embape             | sik.yo.edn aypo e.mba.pe [black.LOC.GEN food NMLZ.VPL.eat]                  | lit. 'food eaten in                     | ADV_N_PAPA            |                      | V-based; descript  |        | no   |             |
| 15 | carpintero         | carpenter     | wamationkohkoh'eri               | wa.ma.tionkohkoh'.eri [NMLZ.VPL.?.AN]                                       |   |                       | NMLZ-V-AN            | V-based            |        | no   |             |
| 16 | cacique            | chieftain     | o'po                             | -   |   |                       | N                    | simple             |        | no   |             |
| 17 | gallo              | cock/rooster  | wäka'ra                          | -   |   |                       | [nothing N           | simple             |        | no   |             |
| 18 | clavicula          | collarbone    | wambagnkutope'ai                 | wa.mbagn.ku.to.pe'.ai [NMLZ.shoulder.head.SPAT:down.CLF:disc.bone]          | wakutopa: 'algo q                       | NMLZ-BN-BN-SPAT-CLF-B |                      | gen                | b-b    | no   |             |
| 19 | cocina             | cookhouse     | wamanwadna hak                   | wa.manwadna hak [NMLZ.cook_in_pots house]                                   | lit. 'a hoi e.man.w                     | NMLZ-V_N              |                      | V-based; descript  |        | no   |             |
| 20 | vaca lechera       | dairy cow     | mbaka önwe'et                    | mbaka önwe'et [cow ?]   | mbaka: from Spanish vaca 'cow'; önwe'et |                       |                      | V-based            |        | no   |             |
| 21 | cena               | dinner        | sikyo aypo embape                | sik.yo aypo e.mba.pe [black.LOC food NMLZ.VPL.eat]                          | lit. 'food eaten in                     | ADV_N_PAPA            |                      | V-based; descript  |        | no   |             |
| 22 | jamba de puerta    | doorpost      | akwikaimbognpe                   | akwikai.mbogn.pe [door.lip;edge.CLF:disc]                                   |   |                       | N-BN-CLF             | cmp                | i-b    | yes  | Mod.Head    |
| 23 | lóbulo de oreja    | earlobe       | wapesodn                         | wa.pe.sodn [NMLZ.(interior_)ear.flappy_object]                              |   |                       | NMLZ-BN-BN           | cmp                | b-b    | yes  | Mod.Head    |
| 24 | arete, pendiente   | earring       | wapesodn'oro                     | wa.pe.sodn'.oro [NMLZ.(interior_)ear.flappy_object gold]                    | oro: from Spanish                       | NMLZ-BN-BN-N          |                      | cmp                | b-i    | yes  | Mod.Head    |
| 25 | cera de oído       | earwax        | wapeoktasörö                     | wa.pe.ok.tasörö [NMLZ.(interior_)ear.period/SPAT:separation.?)              | wa.pe.ok                                | cera de c ?           |                      | gen                |        | yes  | Mod.Head    |
| 26 | ceja               | eyebrow       | wakmberekutopa                   | wa.kmbere.ku.to.pa [NMLZ.forehead.head.SPAT:down.CLF:rod]                   |   | huakber               | NMLZ-BN-BN-SPAT-CLF  | clf                |        | yes  | Base.SPAT.C |





An Van linden\*

# When the alienability contrast fails to surface in adnominal possession: bound nouns in Harakmbut

<https://doi.org/10.1515/ling-2022-0039>

Received March 4, 2022; accepted April 7, 2023; published online October 26, 2023

**Abstract:** This article investigates the nature and behavior of independent, bound and deverbal nouns at various levels of linguistic organization in Harakmbut (isolate, Peru), and assesses the explanatory potential of the alienability contrast for the data observed. While the distinction between bound and independent nouns is to

to study binominal lexemes from a crosslinguistic perspective. Out of the 78 Harakmbut data items collected for 72 entries of this list, 15 instantiate compounding. Table 4 presents the distribution of these 15 items over the two noun classes in N1 and N2 positions. Independent nouns (I in Table 4) are only rarely found as N2, whereas bound nouns (B in Table 4) frequently occur as N2. Although this preference is statistically significant,<sup>17</sup> the scarcity of the data calls for caution. The tendencies in Table 4 need to be confirmed with more data.

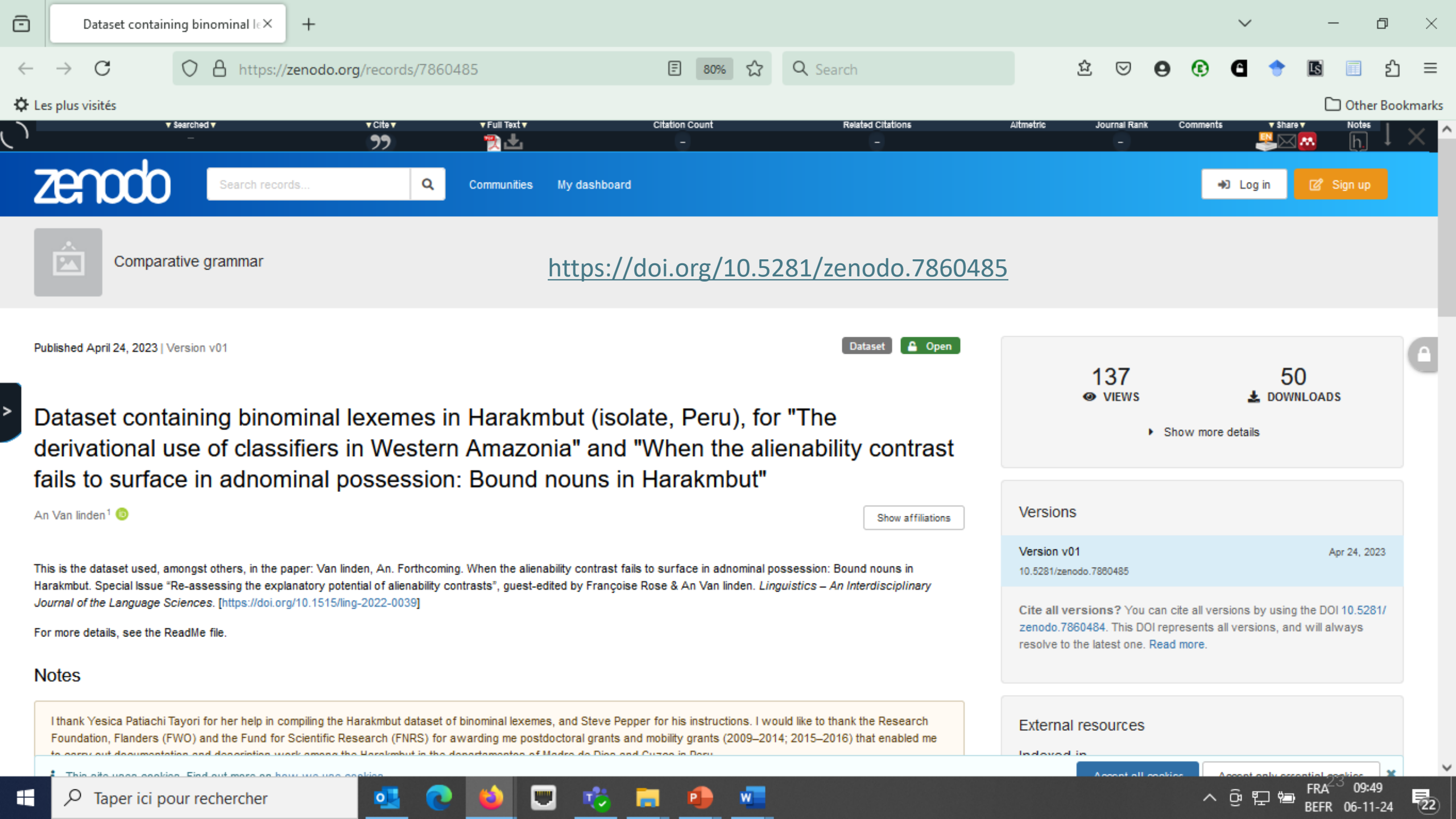
**Table 4:** Types of N-N compounds according to morphological class of N1 and N2 in Pepper’s list of 100 complex concepts.

| Noun classes in N-N | I-I | I-B | B-B | B-I |
|---------------------|-----|-----|-----|-----|
| Harakmbut (15)      | 0   | 5   | 9   | 1   |

Whenever you run statistics on linguistic data, *Linguistics* requires the data to be published (in my case: binomial tests) → to ensure replicability and adequate evaluation of the article

**Data availability statement:** The dataset underlying this article can be found in the Zenodo file: <https://doi.org/10.5281/zenodo.7860485>.





Search records...

Communities

My dashboard

Log in

Sign up

Comparative grammar

<https://doi.org/10.5281/zenodo.7860485>

Published April 24, 2023 | Version v01

Dataset Open

# Dataset containing binominal lexemes in Harakmbut (isolate, Peru), for "The derivational use of classifiers in Western Amazonia" and "When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut"

An Van Linden<sup>1</sup>

Show affiliations

This is the dataset used, amongst others, in the paper: Van Linden, An. Forthcoming. When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut. Special Issue "Re-assessing the explanatory potential of alienability contrasts", guest-edited by Françoise Rose & An Van Linden. *Linguistics – An Interdisciplinary Journal of the Language Sciences*. [<https://doi.org/10.1515/ling-2022-0039>]

For more details, see the ReadMe file.

## Notes

I thank Yesica Patiachi Tayori for her help in compiling the Harakmbut dataset of binominal lexemes, and Steve Pepper for his instructions. I would like to thank the Research Foundation, Flanders (FWO) and the Fund for Scientific Research (FNRS) for awarding me postdoctoral grants and mobility grants (2009–2014; 2015–2016) that enabled me to carry out documentation and description work among the Harakmbut in the department of Madre de Dios and Cuzco in Peru.

137 VIEWS

50 DOWNLOADS

Show more details

## Versions

Version v01

Apr 24, 2023

10.5281/zenodo.7860485

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.7860484](https://doi.org/10.5281/zenodo.7860484). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

## External resources

Indexed in

Accept all cookies

Accept only essential cookies

# References

- Hart, Raymond. 1963. Semantic components of shape in Amarakaeri grammar. *Anthropological Linguistics* 5(9). 1–7.
- Helberg Chávez, Heinrich Albert. 1984. *Skizze einer Grammatik des Amarakaeri*. Tübingen: Tübingen University PhD dissertation.
- Helberg Chávez, Heinrich Albert. 1990. Análisis funcional del verbo amarakaeri. In Rodolfo Cerrón Palomino & Gustavo Solís Fonseca (eds.), *Temas de lingüística amerindia*, 227–249. Lima: Concytec.
- Pepper, Steve. 2020. *The typology and semantics of binominal lexemes. Noun-noun compounds and their functional equivalents*. Oslo: University of Oslo PhD dissertation.
- Rose, Françoise & An Van linden. 2022. The derivational use of classifiers in Western Amazonia. Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes* [Empirical Approaches to Language Typology 62], 237–276. Berlin: De Gruyter Mouton [<https://doi.org/10.1515/9783110673494-008>]

# References

- Tripp, Robert. 1976a. *Los verbos amarakaeri*. Lima: Instituto Lingüístico de Verano.
- Tripp, Robert. 1976b. *Sustantivos verbales y frases de sustantivos verbales en amarakaeri*. Lima: Instituto Lingüístico de Verano.
- Tripp, Robert. 1976c. *Las relaciones señaladas por -po y -nõk en amarakaeri*. Lima: Instituto Lingüístico de Verano.
- Tripp, Robert. 1995. *Diccionario amarakaeri-castellano*. Yarinacocha: Ministerio de Educación & Instituto Lingüístico de Verano.
- Van linden, An. 2023. When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut. In *Re-assessing the explanatory potential of the alienability contrast*, Françoise Rose & An Van linden (eds.), Special issue, *Linguistics* 61 (6), 1533–1561. [<https://doi.org/10.1515/ling-2022-0039>] + DATASET @zenodo [<https://doi.org/10.5281/zenodo.7860485>]
- Van Praet, Wout, Liesbeth Degand & An Van linden. Accepted for publication. As if grammar, discourse and prosody don't interact: A comparative study of hypothetical manner clauses in English and Dutch. *Folia Linguistica*.