

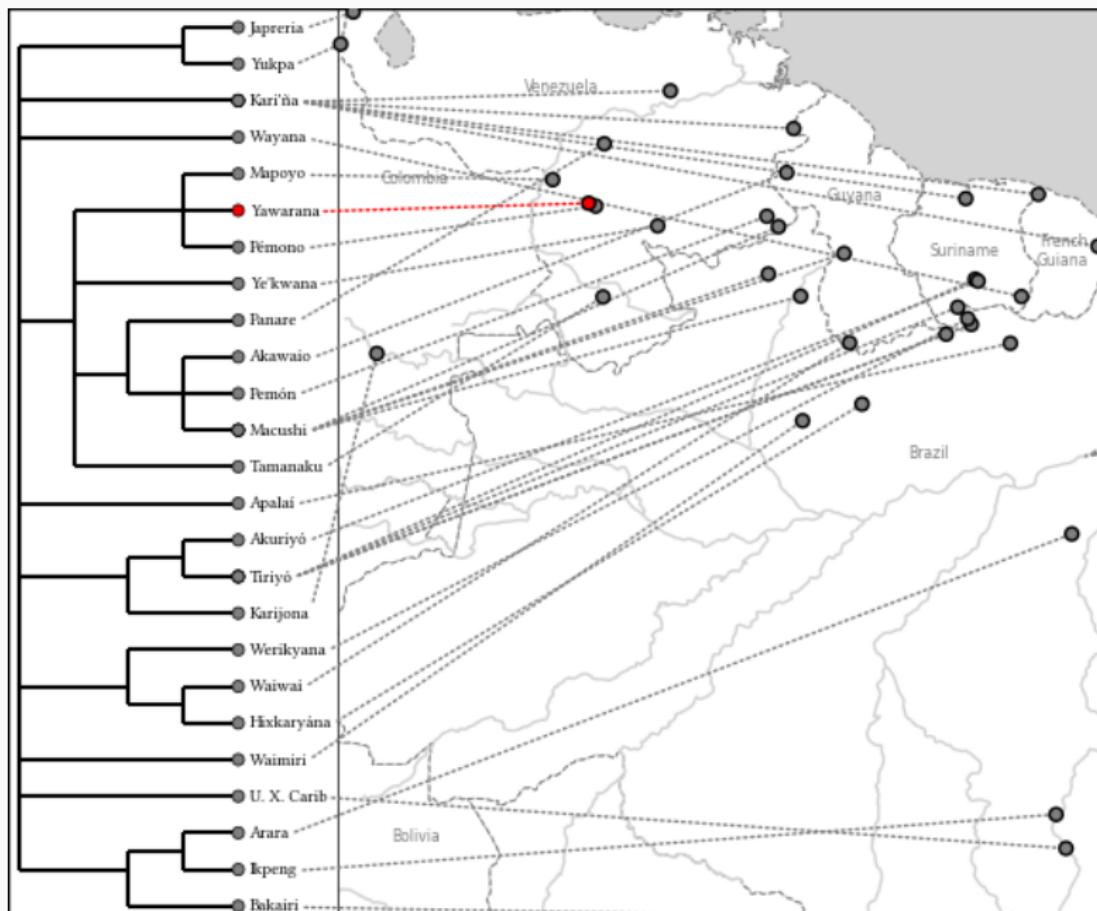
A digital sketch grammar of Yawarana

Florian Matter

Amazonicas IX; June 9, 2023

- a member of the Cariban family (Venezuelan branch?)
- spoken in the upper Orinoco (and Puerto Ayacucho)
- critically endangered, ongoing revitalization efforts
(Conde Pérez and Bolívar 2023)

Geanological and geographic location



- existing work
 - wordlist (Koch-Grünberg 1928)
 - wordlist (Méndez-Arocha 1959)
 - BA thesis (Colina Amaro 1991)
 - comparison with Pémono (Mattéi-Müller 2003)
- NSF-funded project '**Documenting Linguistic Structure and Language Change in Yawarana**'
 - documentary corpus (Cáceres Arandia 2022)
 - dictionary in preparation (Mattéi-Müller and Cáceres Arandia 2023)
- collaboration on digital grammar sketch

My approach to digital grammars

- ideally: a standard, widely accepted, ontology for descriptive linguistics
 - would enable fully machine-readable (and comparable) descriptions
 - could be rendered into human-readable output?
- **grammars are prose interspersed with data** (Nordhoff 2012b)
 - also true for other linguistic documents

My approach to digital grammars

- everything referenced in linguistic texts are abstract entities
 - language
 - ...
 - wordform
 - ...
 - phone
- entities are stored in a database
 - every kind of entity has a table
 - every entity has an ID
- these entities are referenced in the descriptive prose
 - no actual data in text, only pointers
 - (on-the-fly) compilation to readable output

Benefits

- output:
 - multiple formats
 - interactive / non-linear
 - media-rich
- consistency
- explicitness
- “reproducibility”
- accessibility of data for other researchers
- consistent formatting (not trivial)

Implementation

- my weapons of choice:
 - CLDF for data
 - markdown for prose
 - CLLD for non-linear consumption
- **the suggested approach could be implemented with other components**

Implementation: CLDF into Markdown

- Robert Forkel implemented a text component for CLDF
- markdown link notation is co-opted for referencing entities from the database
 - [label] (<http://www.target.com>)
 - [label] (FormTable#cldf:form-1)
 - rendered with [Jinja2](#) templates
- I added functionality with [lingdocs](#):
 - simpler database references ([f] (form-1))
 - cross- and example references
 - tables (as CSV files)
 - multi-file documents
 - different output formats (LaTeX and HTML created with [pandoc](#))
 - general-purpose rendering application for data-rich linguistic documents

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinal)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinfl)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-në*. [todo](do we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives. [todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes [todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinal)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: commands in text

```
# Nouns [label](nouns)
```

```
## Pronouns [label](sec:pronouns)
```

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

```
[table](pronouns)
```

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

```
[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)
```

```
[ex](histyarirdi-160,histpajirdi-114,ctovarmafl-443?example_id=2marking)
```

Implementation: commands in text

```
# Nouns [label](nouns)
```

```
## Pronouns [label](sec:pronouns)
```

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-në*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

```
[table](pronouns)
```

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](1marking); [exref](2marking) illustrates the same distribution for [mp](me2).

```
[ex](histyarirdi-723,convrismaj-46?example_id=1marking)
```

```
[ex](histyarirdi-160,histpajirdi-114,ctovarmafl-443?example_id=2marking)
```

Implementation: commands in text

Nouns [label](nouns)

Pronouns [label](sec:pronouns)

The personal pronouns of [lg](yab) are shown in [ref](tab:pronouns).

The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mp](ejne-1-2pro?nt) does not have the /k/ found elsewhere in the family.

It is likely a reflex of an old copula + infinitive *|*eti-nē*. [todo](do) we have parallel cases elsewhere?)

Regarding plural marking, it should be noted that [mp](kontomopl?nt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.[todo](tajne)

[table](pronouns)

Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbinalf)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinf)).

The occurrence of bound [mp](u1) on members of all three parts of speech is illustrated in [exref](lmarking); [exref](2marking) illustrates the same distribution for [mp](me2).

[ex](histyarirdi-723,conrisamaj-46?example_id=1marking)

[ex](histyarirdi-160,histpjirdi-114,ctovarmafl-443?example_id=2marking)

Implementation: HTML preview

The screenshot shows a Microsoft Word document with the following sections:

- TOC:**
 - 1 Introduction
 - 2 Phonetics and phonology
 - 3 Parts of speech in Yawaraña
 - 4 Nouns**
 - 5 Pragmatically marked construction
 - 6 Verbal roots and stems
 - 7 Detransitive voice
 - 8 Verbal inflection
 - 9 Postpositions
 - 10 Particles, ideophones and suffixes
 - 11 Negation
 - 12 Phrases
 - 13 Auxiliarized constructions
 - 14 Simple verbal clauses
 - 15 Nonverbal predictions
 - 16 Questions
 - 17 Multiclausal
 - 18 Word order variation
- Code Editor:**

```
1 # Nouns [label](nouns)
2
3 ## Pronouns [label](sec:pronouns)
4 The personal pronouns of [tg](yab) are shown in [ref](tab:pronouns).
5 The system shows the usual Cariban inclusive/exclusive ([gl](1+2) and [gl](1+3)) distinction, though the 1+2 pronoun [mpl](e)ne-1-2pro? does not have the /k/ found elsewhere in the family.
6 It is likely a reflex of an old copula + infinitive *i*efi-nə*. [todo](do we have parallel cases elsewhere?)
7 Regarding plural marking, it should be noted that [mpl](kontomoplnt), which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives. [todo](tajne)
8
9 [table](pronouns)
10
11 Reduced forms of the first and second person pronouns occur as proclitics/prefixes[todo](proclitics or prefixes?) attaching to nouns to indicate possessor (see [ref](sec:nominalperson)), attached to verbs to indicate the A or P argument (see [ref](verbint)), or attached to postpositions to indicate the argument of the postposition (see [ref](sec:postinfl)).
12 The occurrence of bound [mpl](u1) on members of all three parts of speech is illustrated in [exref](1marking); [exref](2marking) illustrates the same distribution for [mpl](ne2).
13
14 [ex](histyarirdi-723,convsrisoma]-467example_id=1marking)
15
16 [ex](histyarirdi-160,histpajirdi-114,ctovarmafl-443?example_id=2marking)
17
18 The third person demonstrative pronouns or articles are shown in [ref](tab:pronouns3). [todo](is there a 4-way distinction? [cf. Ye'kwana?])
```
- Table 1: Pronouns**

SG	PL
1	uñ̄e
1+2	ejñ̄i
1+3	ans
2	mñ̄e monkontomo
3	tñ̄eñ̄i tñ̄eñ̄santomo
- Reduced forms of the first and second person pronouns occur as proclitics/prefixes **?** attaching to nouns to indicate possessor (see [1.2.4](#)), attached to verbs to indicate the A or P argument (see [1.2.5](#)), or attached to postpositions to indicate the argument of the postposition (see [1.2.6](#)). The occurrence of bound u'-i on members of all three parts of speech is illustrated in [\(1\)](#); [\(2\)](#) illustrates the same distribution for mñ̄-z'.
- (1) a. iré nasa chipéki u-samorí u-yipí-**?** in-chari
iré nasa chipéki u-samo-ri u-yipi-**?** in-chari
3ANA.JNAN thus *** 1-cry-IPFV 3-mountain-PERT see-JMN
That's why I'm crying seeing my hills (auto) (histyaridi; 723)
b. muñ̄i uñ̄el nñ̄emekiri

Implementation: from text to PDF (via LaTeX)

4.1 Pronouns

The personal pronouns of Yawarana are shown in Table 4.1. The system shows the usual Cariban inclusive/exclusive (1+2 and 1+3) distinction, though the 1+2 pronoun *ejnë* does not have the /k/ found elsewhere in the family. It is likely a reflex of an old copula + infinitive **eti-në*. Regarding plural marking, it should be noted that *-kontomo*, which appears on the second person plural pronoun, is usually restricted to verbs, while *-santomo* is only found with third person pronouns and demonstratives.

Table 4.1: Pronouns

	SG	PL
1	<i>würë</i>	
1+2	<i>ejnë</i>	
1+3	<i>ana</i>	
2	<i>mëré</i>	<i>monkontomo</i>
3	<i>tëwëi</i>	<i>tëwisanntomo</i>

Reduced forms of the first and second person pronouns occur as proclitics/prefixes attaching to nouns to indicate possessor (see Section 4.2.4), attached to verbs to indicate the A or P argument (see Chapter 8), or attached to postpositions to indicate the argument of the postposition (see Section 7.2). The occurrence of bound *u-*'1' on members of all three parts of speech is illustrated in (1); (2) illustrates the same distribution for *më-*'2'.

- (1) a. *uyarë wîrë përemeküri*
u-yare wîrë përemeki-ri
1-alone 1PRO talk-IPFV
'I just talk.'
- b. *irë nwa chipëkë usamori uyipë incharë*
irë nwa chipëkë u-samo-ri u-yipë-Ø in-charë
3ANA.INAN thus *** 1-cry-IPFV 1-mountain-PERT see-IMN
'that's why I'm crying seeing my hills (auto)'

Implementation: tables

pronomens.csv - LibreOffice Calc

The screenshot shows a LibreOffice Calc spreadsheet titled "pronomens.csv". The table has four columns labeled A, B, C, and D. Column A contains row numbers from 1 to 12. Columns B and C contain two entries each, while columns D and E are empty. The data is as follows:

A	B	C	D	E
1	[gl](sg)	[gl](pl)		
2	[gl](1)	[mp](wire-1pro?nt)		
3	[gl](1+2)	[mp](ejne-1-2pro?nt)		
4	[gl](1+3)	[mp](ana-1-3pro?nt)		
5	[gl](2)	[mp](mere-2pro?nt)	[mp](monkontomo-2pl-pro?nt)	
6	[gl](3)	[mp](tewi-3pro?nt)	[mp](tewisantomo-3pro-pl?nt)	
7				
8				
9				
10				
11				
12				

Sheet 1 of 1 Default English (USA) Average: Sum: 0 400%

Implementation: CLDF dataset

- morphs and morphemes
- lexemes, stems, and wordforms (with POS)
- inflectional categories and values
- derivational processes
- phonemes
- “examples” / text records / utterances
- texts
- speakers
- can be found [on github](#)
- (fairly rudimentary) ontology bundled in [CLDF-LDD](#)

Implementation: interactive data-rich web app

- the CLLD web framework allows for rapid development of powerful interactive web apps for linguistic projects
- I developed (reusable and modular) CLLD plugins:
 - `clld-markdown-plugin` (w/ R. Forkel)
 - `clld-document-plugin` (chapters, example references, tables...)
 - `clld-morphology-plugin` (morphological structure)
 - `clld-corpus-plugin` (texts, link entities to tokens in corpus)
- bundled in template called `InDiCoGram`

Implementation: interactive web app

Go live

Comparison: other approaches

- Abesabesi grammar (Lau 2022; Lau 2021)
 - structure: XML description + FLEx converted to better XML
 - writing: manually coding XML
 - consumption: web app
- online grammars of Eastern Cree (Junker 2000–2014) and Nunggubuyu (Thieberger, Musgrave, and Baker 2018; Musgrave and Thieberger 2012)
 - structure, writing, consumption: HTML

References

-  Cáceres Arandia, Natalia (2022). *Yawarana. documentation project collection*. Under curation. URL:
<https://ailla.utexas.org/islandora/object/ailla:254562> (visited on 10/14/2020).
-  Colina Amaro, Angel Arturo (1991). *Descripción del idioma yabarana*. Caracas.
-  Conde Pérez, Olga and María Fernanda de Bolívar (2023). “Iniciativa colectiva de rescate del yabarana (lengua Caribe) en un centro urbano”. Paper presented at the “Simposio de activismo lingüístico”, Amazonicas IX. Bogotá.

References

-  Junker, Marie-Odile, ed. (2000–2014). *The Interactive East Cree Reference Grammar*. URL:
<https://www.eastcree.org/cree/en/grammar/> (visited on 07/19/2021).
-  Koch-Grünberg, Theodor (1928). “Yauarána”. In: *Sprachen*. Stuttgart, pp. 233–245.
-  Lau, Jonas (2021). “A Digital Reference Grammar of Abesabesi. Towards a data format for digital reference grammars”. PhD thesis. University of Cologne.
-  — (2022). *Abesabesi Grammar*. URL:
<https://abesabesi.cceh.uni-koeln.de> (visited on 03/03/2022).

References

-  Mattéi-Müller, Marie-Claude (2003). “Pémono. eslabón perdido entre mapoyo y yawarana, lenguas caribes ergativas de la Guayana noroccidental de Venezuela”. In: *Amerindia* 28, pp. 33–54.
-  Mattéi-Müller, Marie-Claude and Natalia Cáceres Arandia (2023). “Yawarana dictionary”. Flex project with 3583 lexical entries. Database in the author’s possession.
-  Méndez-Arocha, Alberto (1959). “Un vocabulario Yabarana con Apuntes Fonéticos”. In: *Antropológica* 7, pp. 65–84.
-  Musgrave, Simon and Nick Thieberger (2012). “Language description and hypertext. Nunggubuyu as a case study”. In: *Electronic Grammaticography*. Ed. by Sebastian Nordhoff. Manoa: University of Hawai’i Press, pp. 63–77.

References

-  Nordhoff, Sebastian, ed. (2012a). *Electronic Grammaticography*. Manoa: University of Hawai'i Press.
-  — (2012b). “The grammatical description as a collection of form-meaning-pairs”. In: *Electronic Grammaticography*. Ed. by Sebastian Nordhoff. Manoa: University of Hawai'i Press, pp. 33–62.
-  Thieberger, Nick, Simon Musgrave, and Brett Baker, eds. (2018). *Nunggubuyu Grammar Online*. URL: <http://users.monash.edu.au/~smusgrav/Nunggubuyu/> (visited on 05/20/2020).