

Baniwa collection: Guide to archived materials

Sandra Cronhamn

December 2024

The Baniwa collection in the Lund University Humanities Lab Archive contains linguistic data on the Arawakan language Baniwa, collected between 2019–2022 as part of a PhD project by Sandra Cronhamn. Baniwa is part of the Baniwa–Koripako language continuum which is spoken by a few thousand people in Northwestern Amazonia, as seen in Figure 1 (Ramirez, 2020; Cronhamn, 2025).

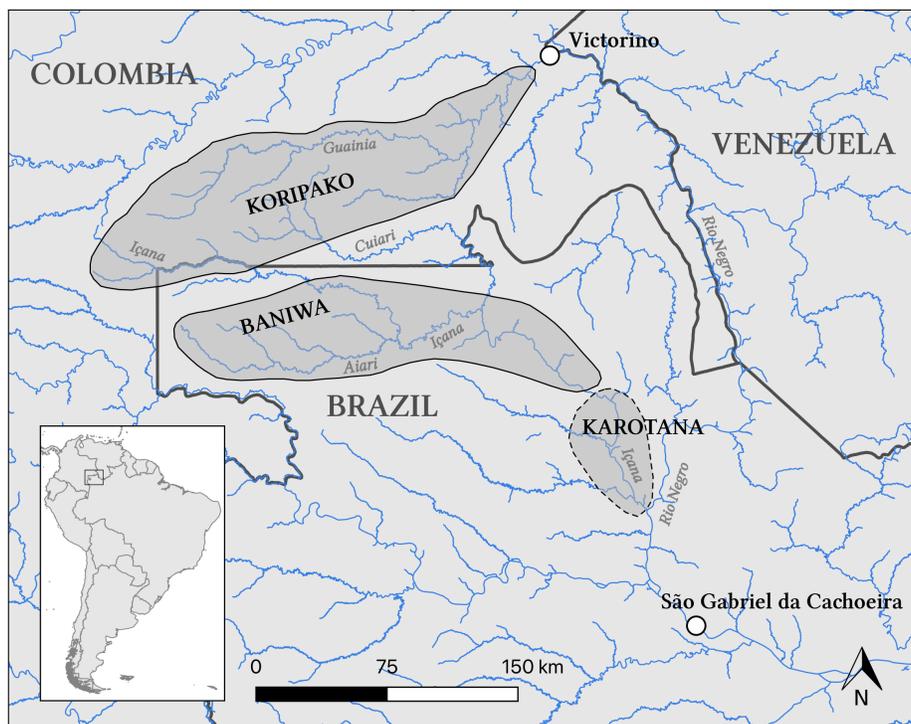


Figure 1: Geographical extension of Baniwa-Koripako dialects (Cronhamn, 2025; based on Ramirez, 2020, 39)

This document presents an overview of the materials in the collection. They are all available on request in the archive.

1 Structure of the archive

The structure of the archive is illustrated in Figure 2.

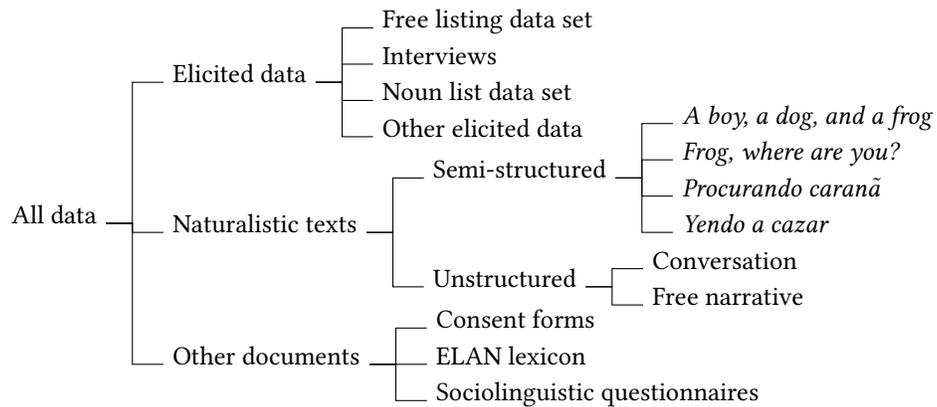


Figure 2: Structure of the archive

The primary data in the archive consists of audio and video recordings. Apart from recordings, the archive also contains transcriptions of some recordings as well as some other supporting files. The contents of the archive are described in more detail in § 2.

1.1 Recording ID:s

All sessions were recorded in audio; most were also recorded in video. Each recorded session is marked with an ID, which consists of the following parts:

1. **The ISO 639-3 language code** (BWI for Baniwa; Eberhard, Simons & Fennig, 2024)
2. **The initials of the researcher in charge of the data collection** (sc for Sandra Cronhamn)
3. **The date of the recording** (YYMMDD)
4. **A shorthand for the data type** (NARR for unstructured narrative data, BOY for *A boy, a dog, and a frog* (Mayer, 1967), FREE for the Free listing experiment, etc.)

5. **An anonymous identifier letter representing the speaker** (assigned randomly; if there are several speakers in the recording, only the letter of the main or first speaker is used in the ID)
6. **A chronological session number**

For example, `BWL_SC_221005_ELIC_K_02` is a recording of an elicitation session of the Baniwa language, collected by Sandra Cronhamn, recorded on the 5th of October, 2022, with the speaker *K*, and it is the second session of this kind with this speaker recorded on this particular day¹.

2 Contents

2.1 Elicited data

Free listing data set

This folder contains the recordings of the Free listing experiment (Franjeh, 2012), which was conducted in order to get an idea of the most prototypical members of each classifier (see more in Cronhamn, 2025, 22 ff.). The folder also features csv-files containing the data, both the individual and the compiled lists.

Interviews

This folder contains interviews with consultants. In contrast to other elicited materials, the interviews featured here are primarily on peoples' backgrounds, and are not part of the core collection of linguistic data.

Noun list data set

This folder contains the recordings of word list elicitation sessions. Transcriptions of the recordings are included as txt-files. The folder also contains a csv-file featuring the compiled Noun list data set, which was used to analyze classifier–noun combinability (Cronhamn, 2025, 22 ff.).

Other elicited data

This folder primarily features recordings of “general” elicitation sessions (see Cronhamn, 2025, 21–22), that is, elicited data that does not belong to any specific data set.

It also contains recordings of a few elicited data sets that ended up not being prepared for analysis: the Shape Classifier Task (Seifart, 2003), the Topological Relations task (Bowerman & Pederson, 1992), and the Banana task (developed by Arthur Holmer).

¹In Cronhamn (2025), the language code and researcher's initials are omitted in references to the data sets for lack of space.

2.2 Naturalistic texts

This folder contains naturalistic data, both unstructured and semi-structured, of a few different kinds.

The semi-structured data consists of recordings of retellings of picture book narratives. The following four picture books were used:

- *A boy, a dog, and a frog* (Mayer, 1967)
- *Frog, where are you?* (Mayer, 1969)
- *Procurando caranã* (Obert, Oliveira & Triana, 2017)
- *Yendo a cazar* (Vuillermet & Desnoyers, 2013)

The unstructured texts consist of both conversational and free narrative data.

Most recordings are paired with an eaf-file for use in ELAN (Sloetjes & Wittenburg, 2008; Max Planck Institute for Psycholinguistics, The Language Archive, 2023). Some of the naturalistic texts have been transcribed and translated into Portuguese. In addition, the following three texts have been interlinearized:

- BWI_SC_190405_BOY_E_01
- BWI_SC_200308_NARR_B_02
- BWI_SC_200308_NARR_J_01

2.3 Other documents

This folder contains secondary documents, such as the scanned and signed consent forms of all participants, and the scanned sociolinguistic background questionnaires of all participants. It also contains a copy of the lexicon from ELAN, featuring all morphemes that occur in the interlinearized texts.

3 Access

All data is under restricted access, but can be made available upon request. If you want to gain access to any of the materials, send an e-mail to sandra.cronhamn@gmail.com. Please specify which materials you would like to get access to, as well as what you want to use the materials for.

4 File formats

The data in the archive is stored in the following formats:

- Video recordings: mp4
- Audio recordings: wav
- Text: txt
- Table-structured text: csv
- ELAN-files: eaf

5 Citing the collection

If you need to cite the materials rather than the publication (Cronhamn, 2025), please cite the collection as follows:

Cronhamn, Sandra (2025). *Baniwa collection: A collection of linguistic data on the Arawakan language Baniwa, spoken in Northwestern Brazil*. Lund: Humlab, Lund University Humanities Lab. <https://hdl.handle.net/10050/84e2a81b-501a-41a9-a59d-9c62becbc6a1>. Accessed on [insert date here].

References

- Bowerman, Melissa & Eric Pederson (1992). “Topological relations picture series”. In: *Space stimuli kit 1.2*. Ed. by Stephen C. Levinson. Nijmegen: Max Planck Institute for Psycholinguistics.
- Cronhamn, Sandra (2025). “A grammar of Baniwa classifiers”. Doctoral dissertation. Lund University.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (2024). *Ethnologue: Languages of the World*. URL: <http://www.ethnologue.com> (visited on 05/27/2024).
- Franjeh, Michael (2012). “Possessive classifiers in North Ambrym, a language of Vanuatu: explorations in semantic classification”. Doctoral dissertation. SOAS, University of London.
- Max Planck Institute for Psycholinguistics, The Language Archive (2023). *ELAN*. Version 6.7. Nijmegen. URL: <https://archive.mpi.nl/tla/elan>.
- Mayer, Mercer (1967). *A boy, a dog, and a frog*. New York: Dial Books for Young Readers.

- Mayer, Mercer (1969). *Frog, where are you?* New York: Dial Books for Young Readers.
- Obert, Karolin, João M. B. de Oliveira & Glauber Triana (2017). “Procurando caranã: Guided story for motion event elicitation”. Manuscript. São Paulo: Universidade de São Paulo.
- Ramirez, Henri (2020). *Enciclopédia das línguas Arawak: Acrescida de seis novas línguas e dois bancos de dados. Vol. I: Um gramática do Baniwa do Içana*. Curitiba, Brazil: Editora CRV.
- Seifart, Frank (2003). “Encoding shape: Formal means and semantic distinctions”. In: *Field research manual 2003, part I: Multimodal interaction, space, event representation*. Ed. by Nick J. Enfield. Nijmegen: Max Planck Institute for Psycholinguistics, 57–59.
- Sloetjes, Han & Peter Wittenburg (2008). “Annotation by category - ELAN and ISO DCR”. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Vuillermet, Marine & Antoine Desnoyers (2013). “A hunting story – Yendo a cazar: A visual stimulus for eliciting constructions that associate motion with other events”. Manuscript. Berkeley: Linguistic Department, UC Berkeley. URL: <http://tulquest.huma-num.fr/fr/node/46>.