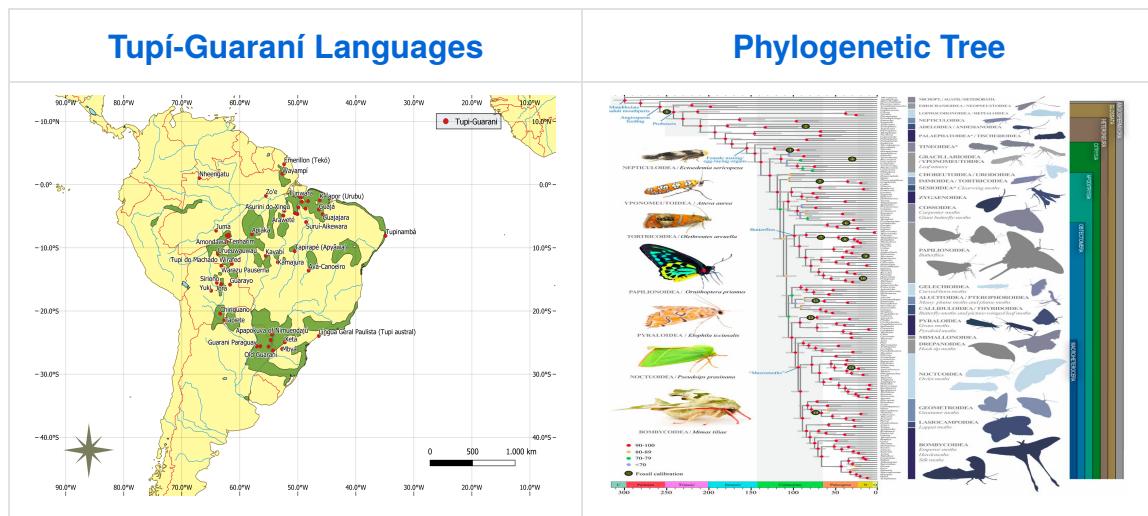


Linguistweets Conference (ABRALIN)

This page contains the [Twitter presentation](#) and additional comments to it with references. The content was presented (tweeted) by Fabrício Ferraz Gerardi, Tiago Tresoldi and Stanislav Reichert on the 05 of December 2020. This is ongoing work in the initial phase.

1 (Introduction)

Tupí-Guaraní (TG) is the largest linguistic family of South America. We want to know more about its spread (peoples and languages). Historical linguistics has tools to investigate this issue aided by digital data, computers, and methods from evolutionary biology.



See Nunn (2011).

2 (Data)

Using phylogenetic tools (BEAST2, beastling), we build linguistic trees from open-access reusable data in CLDF, lifted with EDICTOR from TuLeD (285 concepts, 16211 words, 41 TG langs of 78 in TuLeD). 2832 cognate sets detected with LexStat, 40% manually reviewed (ongoing work).

TuLeD

CLDF

EDICTOR

TuLeD

Welcome to TuLeD (pre-release v0.10)

Collaborators

The list of collaborators can be found here.

Info and Results

Some results of your analysis can be found in the TuLeD dashboard and information about the dataset can be found here.

How to cite TuLeD Online

Parés, Jordi; Ferrer, Jordi; García, Jordi; Martí, Lluís; & Vilà, Mireia. (2023). CLDF data source for South American language families. TuLeD. DOI: 10.5281/zenodo.7500000

Terms of use

This software is released under Creative Commons License. We encourage users of this software to share it with others for the application for the available data and research purposes, subject to applicable laws.

Acknowledgements

Fundació Universitat de Tarragona (FUT) and the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 84363).

GRANT AGREEMENT NUMBER

EU Grant Agreement Number: 84363

DOI

TuLeD is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike license.

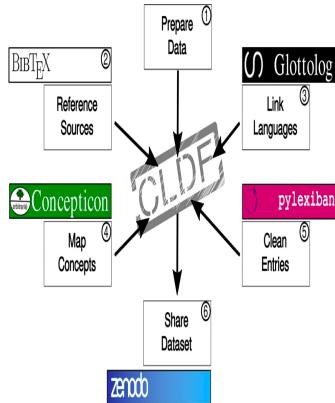
ETC

Probability Distance

Distance

Geographical

CLDF



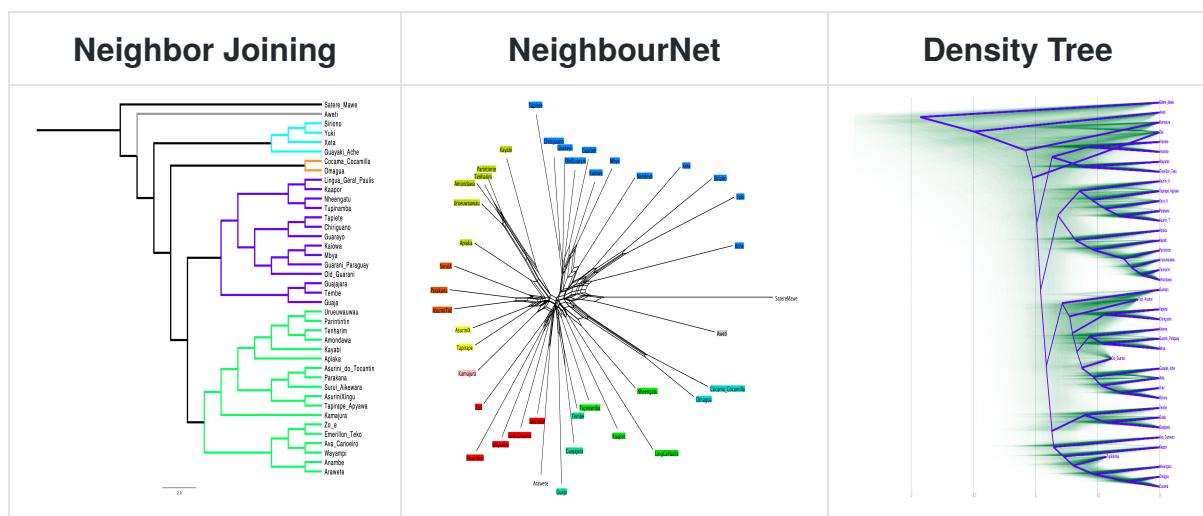
EDICTOR

ID	ELEMENT	CONCEPT		FORM	LEMMA	SEGMENTATION	ALIGNMENT	TOKENS
		TYPE	NAME					
10001	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10002	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10003	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10004	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10005	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10006	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10007	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10008	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10009	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10010	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10011	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10012	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10013	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10014	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10015	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10016	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10017	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10018	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10019	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10020	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10021	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10022	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10023	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10024	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10025	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10026	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10027	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10028	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10029	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10030	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10031	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10032	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10033	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10034	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10035	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10036	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10037	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10038	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10039	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10040	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10041	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10042	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10043	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10044	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10045	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10046	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10047	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10048	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10049	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10050	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10051	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10052	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10053	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10054	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10055	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10056	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10057	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10058	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10059	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10060	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10061	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10062	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10063	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10064	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10065	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10066	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10067	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10068	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10069	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10070	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10071	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10072	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10073	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10074	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10075	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10076	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10077	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10078	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10079	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10080	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10081	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10082	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10083	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10084	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10085	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10086	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10087	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10088	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10089	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10090	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10091	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10092	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10093	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10094	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10095	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10096	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10097	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10098	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10099	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10100	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10101	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10102	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10103	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10104	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10105	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10106	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10107	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10108	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10109	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10110	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10111	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10112	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10113	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10114	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10115	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10116	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10117	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10118	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10119	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10120	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10121	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10122	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10123	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10124	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10125	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10126	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10127	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10128	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10129	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10130	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10131	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10132	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10133	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10134	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10135	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10136	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10137	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10138	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10139	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE
10140	Adjective	ADJDT1	adjetivo					21 143 ADJECTIVE

By interpreting results considering extra-linguistic data we mean looking at the trees and interpreting them also considering, e.g., archeological or ethnographical data. The case of Ka'apor is illustrative. Ka'apor seems to have had a relatively long contact with either Tupinambá or Língua Geral (cf. Ribeiro 1996; Correa da Silva 2011), nonetheless, based on phonological criteria, Ka'apor is grouped with [Wayampi](#), [Emerillon](#), [Anambé](#) of Ehrenreich, and [Guajá](#), languages by Rodrigues and Cabral (2011) that are today far apart. In spite of the position of Ka'apor in the tree, we have linguistic data (Rodrigues and Cabral 2011), and the myths of the Ka'apor giving clues to their location more to the west in the past (cf. Huxley 1963; Ribeiro 1996; Balée 1994).

4 (Explaining a model)

Evolutionary methods: we start from simple (NJ) to highly complex ones for millions of trees (Bayesian MCMC, covarion model, relaxed clock, independent rates of variation). Basic calibrations from literature. Challenge: deliver the best single tree that summarizes the results.



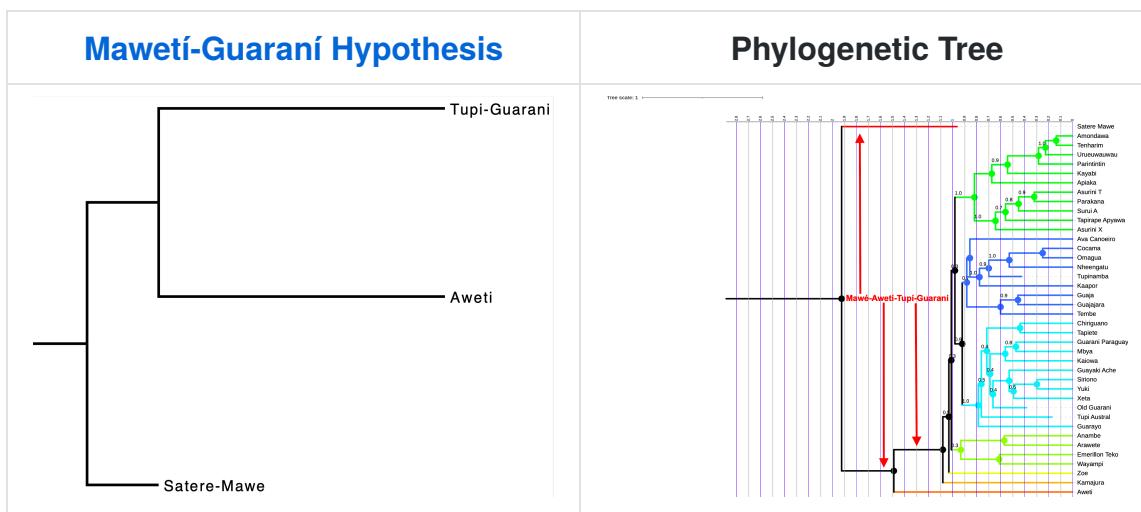
Neighbor Joining (NJ) is the clustering algorithm most frequently used in phylogenetics, which does not require an evolutionary clock. Neighbor joining is more of a quick method of approximating the minimum evolution tree, usually with a confidence inversely proportional to the distance from the root (i.e., it tends to identify recent splits better than older ones). It is a common first approach for data exploration and is expected to be not too different from a Bayesian one.

A NeighbourNet (NN) is a kind of split graph that is commonly used to visualize conflicting signals, such as those resulting from lexical borrowings. It is a quick method for identifying reticulate relationships and, while it does not group languages in terms of genealogy, can be used to visually estimate splits and rates of variation and non-overlapping evolutionary trajectories. The one here reported has a delta score of 0.3972 and a q-residual score of 0.006793.

The Density Tree plots thousands of the best scoring trees on top of each other (in green), highlighting conflicting signals that can be due to a number of reasons, from limited data to a high ratio of borrowings. The summary of the best trees, trees with higher probabilities, is plotted in blue. It also allows us to visually assess how likely some splits and their dates are. In our preliminary density tree, above, where the "root canal" is highlighted in blue, we can see that the groups for Guarani and Tupinambá (Ferraz Gerardi and Reichert 2020) seem clearly distinct, with some probability that it happened earlier than the date of root canal. Some languages also show some conflicting signal: in this analysis, for example, Old Guarani is closer to the Siriono group, but there is a strong signal (visible by the green shade) for a closer membership with the Mbya group. The positions of Zo'e and Kamajurá, as expected, also show a lower degree of confidence.

5 (Preliminary RESULTS)

Mawe-Aweti-TG hypothesis recognized. Fast expansion of Guaraní (light blue) until 400 years ago. 4 major groups identified. Blue group supports ethnogr/hist/ling evidence: dialect continuum, common area of origin, close contact. Fits archaeological dates (Almeida & Neves 2015).



The phylogenetic tree above is a consensus tree, computed with a 20% burn-in, which is the best preliminary results so far, accepting some of the monophyletic restrictions given in the literature and confirmed in our previous analyses. Sateré-Mawé forms the first out-group, followed by Awetí, with full confidence (the posterior support at 1.00 for both) (Silva 2011, Meira & Drude 2015). The groups highlighted by the root canal in the density tree are confirmed and better organized, with some lower support on the initial moments of expansion. These lower support might be improved with more data and longer analyses, but might also support hypotheses on a rapid expansion in different groups. As for dates, this consensus tree suggests that the expansion of the group, after the ancestor of Awetí had already split, started around 1,100 years ago, with a rapid speciation until circa 500 years ago. The less resolved relationship is between the various Guarani languages, which might indicate a continuous contact. The horizontal axis represent time in a scale of 100 years.

6 (To do)

Collect more historical info, improve cognacy. Cognacy alone is not all; include other linguistic data. Constrain and calibrate model(s) (dates, geography, ratios, monophyletic groups). More info, references, and comments: <https://tular.org/tgtweet>. Thank you!

Selected Bibliography

Almeida, F. O. D., & Neves, E. G. (2015). Evidências arqueológicas para a origem dos Tupi-Guarani no leste da Amazônia. *Maná*, 21(3), 499-525.

Balée, W. L. (1994). Footprints of the forest: Ka'apor ethnobotany-the historical ecology of plant utilization by an Amazonian people. Columbia University Press.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C. H., Xie D., et al. (2014). BEAST2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol.* 10(4).

Ferraz Gerardi, F & Reichert, S. (2020). The Tupi-Guaraní language family: a phylogenetic classification. To appear in *Diachronica*.

Greenhill, S. J., Heggarty, P., & Gray, R. D. (2020). Bayesian Phylolinguistic. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The Handbook of Historical Linguistics* (pp. 226-253). Hoboken, New Jersey: Wiley-Blackwell.

Huxley, F. (1963). *Selvagens amáveis:(um antropologista entre os índios Urubus do Brasil)*. Brasiliiana.

List, J. M. (2017, April). A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 9-12).

Maurits, L., Forkel, R., Kaiping, G. A., Atkinson, Q. D. (2017) BEASTling: A software tool for linguistic phylogenetics using BEAST 2. PLOS ONE.

Meira, S., & Drude, S. (2015). A summary reconstruction of Proto-Maweti-Guarani segmental phonology. Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, 10(2), 275-296.

Nunn, C. L. (2011). The comparative approach in evolutionary anthropology and biology. University of Chicago Press.

Rodrigues, A. D., & Cabral, A. S. A. C. (2002). Revendo a classificação interna da família Tupí-Guaraní. Línguas Indígenas Brasileiras. Fonologia, Gramática e História, Atas do I Encontro Internacional do GTLI da ANPOLL, 1.

Rodrigues, A., & Cabral, A. S. (2012). Tupían. In L. Campbell & V. Grondona. The Indigenous Languages of South America: a comprehensive guide.

Rodrigues, A. D., & Dietrich, W. (1997). On the linguistic relationship between Mawé and Tupí-Guaraní. Diachronica, 14(2), 265-304.

Silva, Beatriz Carretta C. D. (2011). Mawé/Awetí/Tupí-Guaraní: relações lingüísticas e implicações históricas. University of Brasília. Unpublished PhD thesis.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; et al. (15 March 2016). "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data. 3: 160018.
doi:10.1038/sdata.2016.18.