Classifying Tupí-Guaraní Languages: Problems and Methods

TG-Lex Team

January 11, 2018

1 Abstract

According to Ethnologue [9], there are 456 languages spoken in South America. Spanning Brazil, Paraguay, Bolivia, French Guiana, Peru and counting more than 5 million speakers, Tupían languages are by far the most widely spoken languages on this continent in terms of territory [13]. Tupí-Guaraní (TG) is the largest Tupían subfamily with ca. 48 languages. While Paraguaian Guarani has ca. 5 million speakers, many of these languages, according to the most recent Brazilian census in 2010 [6], have less than 100 speakers and a dozen of these have probably died out since then. Tupí-Guarani languages are also territorially widespread like no other language family in South America. These languages seem to have originated ca. 3000 years B.P., possibly in what is today the Brazilian state of Rondonia. As of today, the exact place of their origin and the routes of their spread through their historical territories remain topic of an ongoing scientific debate [12, 15].

The last ten years have seen a significant increase in the amount of studies of Tupí-Guaraní languages. Nonetheless comprehensive descriptions are still at large for many of these languages, which share very interesting and rare typological and grammatical characteristics. These characteristics are of importance, since they allow us not only to make inferences about the evolution of these languages and to refine our knowledge of linguistic typology, but also to gain knowledge about linguistic areas in South America [3], and to be able to classify languages using quantitative methods in a more reliable way (cf. [10]).

Comparative linguistic work may throw some light on contact and migrations, thus emphasizing or showing disagreement with the interpretation of archaeological data [14]. There is no consensus on the TG homeland and their dispersal in both archaeological and linguistic research literature [12]. It is, however, clear that the answer to the questions of how, where and when the dispersion occurred will eventually come from combined insights in both disciplines [15, 5], as it has been the case for other language families ([1]).

Our team has collected 498 words from 40 Tupí-Guaraní languages and 4 additional Tupían languages to be used as outliers in order to classify Tupí-Guaraní languages lexically and phonologically paying special attention to the lowest levels of the tree (subgrouping). In order to classify the lexicon, the first stage of our work applies Bayesian phylogenetic methods taken from molecular biology. We also use computational methods commonly employed in historical linguistics to produce an

alignment of words in the various languages in order to calculate lexical distances [7].

Another part of our team's work comprises a separate attempt to classify these languages based on grammatical properties, updating and extending a paper by Wolf Dietrich ([4]), specially those that are very characteristic of this family such as the relational prefixes marking contiguity [11] and a morpheme that allows different word classes to be used as arguments of predicates [2], since these languages exhibit omnipredicativity [8] at least to some degree. Our research aims to combine archaeological and recent data from genetic studies with linguistic data in order to propose a more solid ground for their classification and spread.

References

[1] D. W. Anthony. *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world*. Princeton University Press, 2010.

[2] A. Cabral. Observações sobre a história do morfema-a da família tupí-guaraní. *Des noms et des verbes en tupi-guarani: état de la question*, pages 133–162.

[3] M. Crevels and H. Van der Voort. The guaporé-mamoré region as a linguistic area. In P. muysken, editor, *From linguistic areas to areal linguistics*, pages 151–179. John Benjamins, Amsterdam, 2008.

[4] W. Dietrich. *More evidence for an internal classification of tupi-guarani languages*. Gebr. Mann Verlag, 1990.

[5] W. Dietrich. As línguas tupí-guaraní bolivianas e as de rondônia: novas hipóteses sobre as origens. Forthcoming.

[6] IBGE. Os indígenas no censo demográfico 2010. `https://indigenas.ibge.gov.br/images/pdf/indigenas/folder_indigenas_web.pdf`. Accessed: 2017-27-12.

[7] G. Jäger. Phylogenetic inference from word lists using weighted alignment with empirically determined weights. *Language Dynamics and Change*, 3(2):245–291, 2013.

[8] M. Launey. The features of omnipredicativity in classical nahuatl. *STUF-Language Typology and Universals*, 57(1):49–69, 2004.

[9] M. P. Lewis, G. F. Simons, C. D. Fennig, et al. *Ethnologue: Languages of the world*, volume 16. SIL international Dallas, TX, 2009.

[10] A. McMahon and R. McMahon. *Language classification by numbers*. Oxford University Press on Demand, 2005.

[11] S. Meira and S. Drude. Sobre a origem histórica dos "prefixos relacionais" das línguas tupí-guaraní. *Cadernos de Etnolinguística*, 5(1), 2013.

[12] F. S. Noelli. The tupi expansion. In H. Silverman and W. Isbell, editors, *The Handbook of South American Archaeology*, pages 659–670. Springer, 2008.

[13] A. Rodrigues. Tupían. In L. Campbell and V. Grondona, editors, *The indigenous languages of South America: a comprehensive guide*, volume 2. Walter de Gruyter, 2012.

[14] A. D. Rodrigues. Hipótese sobre as migrações dos três subconjuntos meridionais da família tupi-guarani. In *Atas do II Congresso Nacional da ABRALIN*, pages 1596–1605, 2000.

[15] G. Urban. On the geographical origins and dispersion of tupian languages. *Revista de Antropologia*, pages 61–104, 1996.