

Subgrouping in a dialect area: A Bayesian phylogenetic analysis of Mixtecan

Sandra Auderset^{1,2} and Eric W. Campbell²

¹MPI for Evolutionary Anthropology, ²University of California Santa Barbara

Keywords: Mixtecan, Bayesian phylogenetics, subgrouping, dialect area

We present the first Bayesian phylogenetic analysis of Mixtecan, a language family of southern Mexico. We show that this method produces valuable results and new insights with respect to subgrouping beyond what the comparative method and dialect geography can provide. Our findings point to potential new subgroups that should be further investigated. We demonstrate that some unexpected groupings raise important questions about the effects of different methods of primary data gathering and organization.

Gaining a more detailed understanding of subgrouping within Mixtecan is also important for the creation of community language materials, as misclassifications on lower levels abound in reference catalogs such as Glottolog (Hammarström et al. 2021) and Ethnologue (Eberhard et al. 2021).

Mixtecan is classified by most as a branch of the Otomanguean stock (Campbell 1997; Kaufman 2006; Campbell 2017), but its internal structure is much less clear. This is mostly due to the highly diversified Mixtec group, characterized by most scholars as a dialect continuum consisting of chains of mutually intelligible varieties (Jiménez Moreno 1962; Josserand 1983: 457–458). It has sometimes been suggested that the comparative method or phylogenetics cannot be felicitously applied to dialect areas, because the varieties remain in contact and there is too much noise from internal loans (Bradley & Josserand 1982; Geraghty 1983; Ross 1997; inter alia). However, Bowerman (2013) and Greenhill et al. (2009) have shown that the impact and prevalence of contact on genealogical relationships is often overestimated and poses less of a problem for comparative reconstruction and Bayesian phylogenetics than has been argued. Our study adds further support to these claims.

We collected a standardized word list with 209 entries of basic vocabulary from 137 varieties (131 Mixtec, 4 Triqui, 2 Cuicatec), which we coded for cognacy resulting in over 18,000 data points distributed over more than 1,000 cognate sets. Using these data, we ran a Bayesian phylogenetic analysis with BEAST2 (Bouckaert et al. 2019) and extracted a consensus tree.

Our results recover the three main groups of Mixtecan, that is Triqui, Cuicatec, and Mixtec. Within Mixtec, we find partial confirmation for the only earlier subgrouping proposal available (Josserand 1983). This proposal divides the Mixtec languages into 12 groups based on isoglosses identified by vowel changes. In our study, we recover five of the proposed groups completely (Northern Alta, Northeastern Alta, Northern Baja, Mixtepec, Coast). For the other seven groups, our model either suggests that they be merged into one group (parts of Southern Baja+Guerrero, Central Baja+Western Baja) or be split into two or more groups (Western Alta I + II +II, Southern Baja I + II). We also identify specific varieties whose group membership

needs to be re-evaluated, explaining their misplacement partly in terms of data gathering processes.

This assessment serves as the basis for a new subgrouping proposal that can be evaluated by further research based on sound changes, reconstruction of the proto-language and intermediate stages, and novel data gathered in ongoing documentation projects.

References

- Bouckaert, Remco, Timothy G Vaughan, Joelle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15(4)
- Bowern, Claire. 2013. Relatedness as a factor in language contact. *Journal of Language Contact* 6. 411–432
- Bradley, C. Henry & J. Kathryn Josserand. 1982. El protomixteco y sus descendientes. *Anales de Antropología* 19(2). 279–343
- Campbell, Eric W. 2017. Otomanguean historical linguistics: Exploring the subgroups. *Language and Linguistics Compass* 11(7)
- Campbell, Lyle. 1997. *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press
- Eberhard, David M., Gary F. Simons, Charles D. Fennig. 2021. *Ethnologue: Languages of the World*. Dallas: SIL International. Online version: <http://www.ethnologue.com>
- Geraghty, Paul A. 1983. The History of the Fijian Languages. *Oceanic Linguistics Special Publications* 19. Honolulu: University of Hawaii Press.
- Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society of London B: Biological Sciences* 276. 2299–2306
- Hammarström, Harald, Robert Forkel, Martin Haspelmath, Sebastian Bank. 2021. *Glottolog* 4.5. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://glottolog.org>
- Jiménez Moreno, Wigberto. 1962. Estudios mixtecos. In: Fray Francisco de Alvarado, *Vocabulario en lengua mixteca*, 9–105. México: Instituto Nacional Indigenista e Instituto Nacional de Antropología e Historia S.E.P
- Josserand, J. Kathryn. 1983. *Mixtec dialect history*. New Orleans: Tulane University PhD dissertation
- Kaufman, Terrence. 2006. Oto-Manguean languages. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, vol. 9, 118–124. Oxford: Elsevier
- Ross, Malcolm D. 1997. Social networks and kinds of speech-community event. In: Roger Blench and Matthew Spriggs (eds.), *Archaeology and Language I*, 209–261. London: Routledge