A Quantitative Analysis of Morphological Features of the Greater Southwest

The Taracahitic subgroup of Uto-Aztecan (UA) is made up of Ópata-Eudeve, Cahitan and Tarahumara-Guarijío, which are found across the Southwest and Northern Mexico. Taracahitan has been well established by lexicostatistical studies,[1] which have unanimously found tight genetic unity of Ópata-Eudeve, Cahitan and Tarahumara-Guarijío. In her analysis of UA subgrouping through shared phonological innovations, Hill (2011) found no shared innovation connecting the previously classified Taracahitan languages as a clade. Hill's (2011) findings suggest that the robust unity of Taracahitan languages using lexicostatistical methods is more areal in nature than genetic. This position is reinforced by the geographic proximity of these languages, where only Ópata-Eudeve is not geographically contiguous, but is still quite close. Moreover, Foley (2005) points out that the peoples of the Southwest (not including the Pueblos) have been largely open to linguistic change in the period since Spanish contact. Thus, a contact driven convergence of both basic vocabulary and morphological would not seem completely out of place.

In this paper I follow Dunn et al.'s (2005) and Donahue et al.'s (2011) examination of morphological features among Papuan languages in analyzing the spread of morphological features in the languages surrounding the Taracahitan languages. I also specifically follow Donahue et al. (2011) in viewing the spread of these features as being largely indicative of contact. Here, I examine the areal distribution of morphological and structural features across the areas of the Great Basin, southern California, southern Plains, Pueblos, Southwest and Northern Mexico.[2] The languages of these areas are included in a number of language families, including: UA, Yuman, Caddoan, Keresan, Kiowa-Tanoan, Apachean and several isolates. I also include several members of the Aztecan subgroup of UA and Cora and Huichol. For comparison outside of the areas involved I also include: Guaraní, Turkish, Meithei and Swahili, which are not known to have been historically interactive with the languages examined here and are geographically quite distant. The distribution of morphosyntactic features amongst these languages should reproduce the lexicostatistically derived UA subgroups, however the position of the non-UA languages will be indicative of the impact that contact has had on the UA languages. Thus, Taracahitan languages will likely cluster with the non-UA Yuman and Seri languages, which are geographically contiguous, and historically interactive.

The database I use here consists of structural features compared across 52 languages. The features were initially selected from amongst those in WALS that have a value for Yaqui (as it is a well represented language in the database) and added to based on features used in Hill (2016). While WALS encodes features in multi-state form, Hill (2016) only uses binary-state coding. Thus, certain features such as WALS's *Antipassive* feature has been condensed from trinary (Implicit patient, Oblique patient, No antipassive) into binary (present, absent) because Hill (2016) does not have information for WALS's trinary values. However, where possible I have retained multi-state distinctions. Any value for a given feature is coded with a letter (a-g), while languages which have no data for a given feature are coded with (?). I use SplitsTree (Hudson & Bryant 2006) to generate a tree based on relative similarity among the languages. Within a given feature, any two or more languages with the same letter are connected and any language with no value, or (?), is treated as connected with every language, thus giving it a null value. As this project is focused on the areal distribution of features across the areas involved, I treat inherited, diffused and independently innovated features equally. Because a number of the languages in the aforementioned areas are poorly documented, the languages and features used in this study are selected by convenience, based on the amount of reliable materials that are accessible. I selected features that had values for at least 75% of the languages included and then removed any language for which there were not values for at least 75% of the selected features.

This project is still in the data collection phase, however, there are 29 features across 52 languages which meet the criteria for inclusion and by the presentation of this paper I will have at least 40 features for comparison. This project has direct implications for the subgrouping of the UA family, in terms of giving more credence to the Taracahitic languages as an areal unit rather than a genetic one. In a more general sense, this project argues against the notion that lexicostatistical methods are sensitive enough to be used to reliably classify languages where intra-familial contact is prevalent.

---

[1] Miller (1984), Cortina-Borja & Valiñas (1989), Cortina-Borja et al. (2002), Haugen, Everdell & Kuperman (2015, 2016) etc.

[2] Several of these have been previously discussed as linguistic areas in their own right (see Campbell 1997 for further discussion).

Bibliography

Cortina-Borja, Mario, Jane Stuart-Smith, and Leopoldo Valias-Coalla. 2002. "Multivariate Classification Methods for Lexical and Phonological Dissimilarities and their Application to the Uto-Aztecan Family." *Journal of Quantitative Linguistics* 9 (2): 97-124.

Cortina-Borja, Mario and Leopoldo Valias C. 1989. "Some Remarks on Uto-Aztecan Classification." *International Journal of American Linguistics* 55 (2): pp. 214-239.

Donohue, Mark, Simon Musgrave, Bronwen Whiting, and Sren Wichmann. 2011. "Typological feature analysis models linguistic geography." *Language* 87, no. 2: 369-383.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309 (5743): 2072-2075.

Foley, William A. 2005. "Personhood and linguistic identity, purism and variation." *Language documentation and description* 3: 157-180.

Haugen, Jason D., Michael Everdell, and Benjamin A. Kuperman. 2015 "A New Approach to Uto-Aztecan Lexicostatistics? Presentation at WAIL, University of California, Santa Barbara, CA, May 7-8, 2015.

Haugen, Jason D., Michael Everdell, and Benjamin A. Kuperman. 2016 "Bayesian Phylogenetic Classification of the Uto-Aztecan Languages Using Lexical Data" Presentation at 13th Annual MLK Day Linguistics Symposium, The Ohio State University, OH, January 15-16, 2016.

Hill, Jane H. 2011. "Subgrouping in Uto-Aztecan." *Language Dynamics and Change* 1: 241-278.

Hill, Jane. *Languages of California and the Great Basin. In Languages of Hunter-Gatherers and their Neighbors: Database, Claire Bowern, Patience Epps, Jane Hill, and Patrick McConvell.* https://huntergatherer.la.utexas.edu. Accessed 10/5/2016.

Huson, Daniel H., and David Bryant. 2006. "Application of phylogenetic networks in evolutionary studies." *Molecular biology and evolution* 23, no. 2: 254-267.

Miller, Wick R. 1984. "The Classification of the Uto-Aztecan Languages Based on Lexical Evidence." *International Journal of American Linguistics* 50 (1): pp. 1-24.