

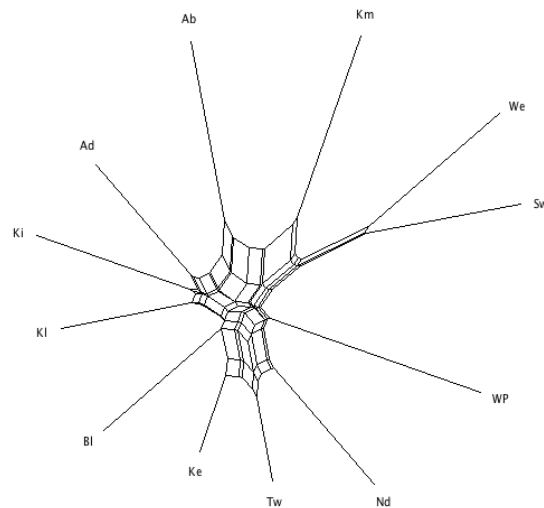
Internal classification of Alor-Pantar using computational methods applied to the lexicon

Laura Robinson and Gary Holton
University of Alaska Fairbanks

The standard method for establishing internal genetic relationships within a language family relies on tracing shared innovations, usually phonological ones. In many language families, however, continued contact and convergence renders the family tree model inadequate for the purposes of classification (cf. Krauss 1973 on Athabaskan, Sidwell 2009 on Austroasiatic, Windfuhr 2009 on Iranian). In this paper we discuss one such problematic case from the Alor-Pantar family and propose an alternate method of subgrouping employing computational phylogenetic methods applied to lexical data.

The genetic relatedness of the non-Austronesian languages of Alor-Pantar was suggested on the basis of lexicostatistical evidence (Stokhof 1975) but has only recently been established using the comparative method (Klamer et al. 2009, Holton et al. to appear). However, the identified sound changes are typologically common and do not delineate neat subgroups. For example, the change pAP *r > l subgroups Adang with Western Pantar, while the change *s > h subgroups Adang with Blagar and Klon. These two changes typify the larger data set, complicating standard approaches to subgrouping. Moreover, many of the changes do not run to conclusion but instead are only partially realized across the lexicon.

We apply the NeighborNet algorithm (Huson & Bryant 2006) to recently-collected lexical data consisting of nearly 400 items from twelve languages. Crucially, these data are coded for cognacy (based on the results of the comparative method) rather than apparent similarity, allowing internal borrowings to be excluded from the analysis. Nonetheless, the resulting network (shown below) still exhibits a high degree of reticulation (delta score 0.38), indicating the presence of undetected borrowing or incomplete sound changes.



There is much variation across the lexicon, however. We perform a factor analysis of the relative contribution of each lexical character, allowing us to identify classes of lexical items which are more or less resistant to intra-family borrowing. This allows us to posit subgroups based on the most stable lexical items.

This paper shows that computational cladistics applied to the lexicon can be a useful tool for internal classification. In contrast to lexicostatistical methods (Swadesh 1950), our approach relies explicitly on the comparative method to identify cognate relationships. While approaches based solely on typological characters (e.g., Dunn et al. 2008) have been criticized for revealing more about areal rather than genetic tendencies (cf. Donohue et al. 2011), in the case of closely related languages such as Alor-Pantar group, typological approaches offer little insight since the languages share a common phonological and morphological profile. In these cases, computational methods using lexical data can be a powerful tool supplementing the comparative method.

- Donohue, Mark, Simon Musgrave, Bronwen Whiting & Søren Wichmann. 2011. Typological feature analysis models linguistic geography. *Language* 87(2):369-83.
- Dunn, Michael, Stephen C. Levinson, Eva Linström, Ger Reesink & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in island Melanesia. *Language* 84(4):710-59.
- Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2): 254-67.
- Klamer, Marian, Gary Holton & František Kratochvíl. 2009. The languages of Alor-Pantar (Eastern Indonesia): A (re)assessment. Paper presented at the 11th International Conference on Austronesian Linguistics. Aussois, France, June 22-25.
- Krauss, Michael E. 1973. Na-Dene. In T.A. Sebeok (ed.) *Linguistics in North America*, 903-78. (Current Trends in Linguistics 10). The Hague: Mouton.
- Sidwell, Paul. 2009. How many branches in a tree? Cua and East (North) Bahnaric. In Bethwyn Evans (ed.) *Discovering history through language. Papers in honour of Malcom Ross*. Canberra: Pacific Linguistics.
- Stokhof, W. A. L. 1975. *Preliminary notes on the Alor and Pantar languages (East Indonesia)*. Canberra: Pacific Linguistics.
- Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16(4): 157-167.
- Windfuhr, Gernot, ed. 2009. *The Iranian Languages*. (Routledge Language Family Series.) London: Routledge.