Areality through Migration: Investigating the Structure of Numeral Classifiers in the Eastern Himalayan Region Reveals Historic Contact Events

Numeral classifiers exhibit striking distribution in certain hotspots around the world, serving as strong areal markers due to their highly recessive nature genetically and areally (Nichols, 2003:299). By investigating one such known hotspot, the Eastern Himalayan Region (henceforth: EHR ; often called North-East India in the prior literature) – I show that the surface proliferation of a grammatical feature can actually be the result of migration and distant contact events, rather than an indication of intense, sprachbund-like convergence as was previously assumed for this feature. In doing so, I problematize the notion of an 'areal feature,' and the methods used to define them.

The EHR is a complex linguistic zone with four distinct language families - Trans-Himalayan, Indo-Aryan, Austroasiatic and Kra-Dai - all with long-standing similarities in history, culture and ecological environment (Konnerth et. al, 2020). Previous work identifying the EHR as a linguistic area sees classifiers as a defining areal feature (Moral, 1997), with recent work (Chelliah and Lester, 2017 ; Cathcart et. al, 2020) pointing to contact as the explanation for the prominence of classifiers in the region, or their emergence in Eastern Indo-Aryan, respectively. By surveying 22 of the languages in Assam (the most populous state of the region), I find evidence to the contrary, and present the first detailed investigation of the spread of the feature in this region that reveals clues about the linguistic prehistory of the EHR, and South Asia more broadly.

By analyzing the structure of the classifier systems in the area, I find that contact is an unlikely explanation for the high frequency of classifier languages in the area. The premise of my analysis is the fact that classifiers here are not lexical borrowings from other languages in the region, but instead have cognates in their own distant linguistic relatives. Hence, a situation of language contact must be similar to that elucidated by Matras and Sakel (2007) where structural, instead of lexical borrowing takes place, replicating a *pattern* from the donor language: a manifestation of metatypy (Ross, 1996, 2007). The key elements I identified for this analysis were Noun (N), Numeral/Quantifier (Q), and Classifier (CL) as established by previous scholars (Jones, 1970; Allan, 1977). When examining the order of these key elements, I find that the structure of the classifier phrase in *every* language of the area is predicted by its genetic affiliation, reflecting inheritance, rather than contact. This argument is further supported by the fact that the languages of the same family outside of the area (such as Standard Thai, Odia and Vietnamese) do not show a difference in the order of key elements as those in the EHR. All of this points towards contact being an insufficient explanation for the perceived areality of numeral classifiers. Some interesting oddities remain – Eastern Indo-Aryan and Khasian do share the order of key elements, and Tibeto-Burman is an outlier in all senses – the order of its classifier phrase is distinct from all languages in the area as well as its own genetic relatives outside of the area.

What then, explains the distribution of numeral classifiers in this region? I suggest that what has been considered an areal feature of the EHR can be better explained by two distant contact events and the subsequent migration of those communities into the region. This hypothesis is supported by previous work in historical linguistics – Peterson (2010, 2017) highlights classifiers as a potential contact feature between Eastern Indo-Aryan and Proto-Munda. Moreover, Dockum (2016) points towards metatypy between Southwest Tai and parts of Austroasiatic, which could have been an earlier contact event that feeds into this. I weigh the linguistic data against contributions from allied fields such as archaeogenetics, anthropology, and history, to build a timeline of when speaker groups could have migrated into the EHR, and elucidate the kind of contact that has occurred in the region. Understanding how patterns of grammatical features in the EHR can tell us about its history allows us to build a better picture of how contact and migration shape the linguistic diversity we see in other such complex linguistic zones with high genetic diversity, such as the Daly River region, Western South America, or Mainland South-East Asia.

References

Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311. https://doi.org/10.1353/lan.1977.0043.

Cathcart, Chundra A., Andreas Hölzl, Gerhard Jäger, Paul Widmer & Balthasar Bickel. 2020. Numeral classifiers and number marking in Indo-Iranian: A phylogenetic approach. *Language Dynamics and Change* 11(2). 273–325. https://doi.org/10.1163/22105832-bja10013.

Chelliah, Shobhana & Nicholas Lester. 2016. Contact and convergence in the Northeast. In *The Languages and Linguistics of South Asia: A Comprehensive Guide*. Walter de Gruyter GmbH & Co KG.

Dockum, Rikker 2016. Numeral classifiers in areal perspective: Khmer and Thai 'syntactic borrowing' revisited. *Austroasiatic Syntax in Areal and Diachronic Perspective. Workshop of the Austroasiatic Working Group.* September 5-7, 2016. [Slides]

Jones, Robert B. 1970. Classifier Constructions in Southeast Asia. *Journal of the American Oriental Society*. American Oriental Society 90(1). 1–12. https://doi.org/10.2307/598428.

Konnerth, Linda, Stephen Morey, Mijke Mulder, Mark W. Post & Kellen Parker van Dam. 2020. Languages and Peoples of the Eastern Himalayan Region and the North East Indian Linguistics Society: Taking stock. Himalayan Linguistics 19(2). https://doi.org/10.5070/H91151002.

Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical Borrowing in Cross-Linguistic Perspective* (Empirical Approaches to Language Typology [EALT]). Berlin, New York: Mouton de Gruyter. https://doi.org/10.1515/9783110199192.

Ross, Malcolm. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford University Press, USA.

Ross, Malcolm. 2007. Calquing and Metatypy. *Journal of Language Contact*. Brill 1(1). 116–143. https://doi.org/10.1163/000000007792548341.

Nichols, Johanna. 2003. Diversity and Stability in Language. In *The Handbook of Historical Linguistics*, 283–310. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756393.ch5.

Moral, Dipankar. 1997. North-East India as a linguistic area. *Monkhmer Studies* 43-54.

Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2). https://doi.org/10.5070/H99223479.

Peterson, John. 2017. Fitting the pieces together - Towards a linguistic prehistory of eastern-central South Asia (and beyond). *Journal of South Asian Languages and Linguistics*. Berlin, Germany: Walter de Gruyter GmbH 4(2). 211–257. https://doi.org/10.1515/jsall-2017-0008.