

Paradigms, diachronic typology, and language classification: Bayesian phylogenetics and pronoun paradigms
(Oral presentation)

There is growing interest in the use of typology in historical linguistics. One particularly promising approach is to investigate relationships between tightly organised subsystems of language, as fractal-like proxies for overall relationships. One such case was the investigation of Northeast Caucasian languages by Cysouw & Forker (2009) who found close parallels between the phylogenetic classification given by a typological classification of the case systems and that given by traditional measures. However, the particular characteristics of Northeast Caucasian case systems prevent this particular application from being generalisable to most other language families. Therefore we focus on a linguistic subsystem – personal pronouns – which is found in all spoken languages.

This paper compares the architecture of attested pronominal systems. We develop a diachronic typology of the stepwise pathways between them, and then harness this to a phylogeny to infer the most parsimonious historical scenarios relating pronominal systems across a series of languages.

To get the right balance between informativeness and comparability, we focus on four (potentially) distinct forms of each personal pronoun: those for the three core grammatical relationships (A, S and O) and for the possessive pronoun (of course many languages will neutralise some of these distinctions). These four values then combine with the system of person/number combinations as well as other categories such as gender to give a paradigm. Within any pair of cells in the paradigm we code a number of relationships ranging from totally distinct (e.g. *we* and *our* in English), total syncretism (e.g. A=S in English *he*, S=O for Nen *bā* ‘he/she’; sg=du=pl in English *you*) to formal overlap (e.g. *me* and *my* in English) to formal increment (e.g. Warlpiri ergative *ngajulurlu* from Warlpiri absolutive *ngaju(lu)*). Taken together, these factors then give a vast ‘architectural design space’ formed by the product of all featural paradigms (in terms of feature combinations) with all formal relationships between all cells within them.

Diachronic pathways between possible paradigms in the design space can then be modelled by assuming they are achieved by (a) adding or subtracting features or combinations thereof (e.g. neutralising inclusive/exclusive, or developing a distinct ergative form) (b) changing the formal relationship between any two cells (e.g. from formal increment to formal overlap, or from distinct to total syncretism). Once the set of possible pathways has been exhaustively elaborated, we can give the evolutionary distance between any two paradigms by calculating the number of transformational steps needed to get from one to another. We can then use a Bayesian phylogenetic methods to infer a set of phylogenies based on the most parsimonious set of changes across the whole population of paradigms.

The method will be tested against existing classifications of one well-known language family (Austronesian) to determine its reliability, and then trialled as a heuristic classification for Australian and Papuan languages (Trans-New Guinea and Morehead-Maró, plus selected outgroups).

Cysouw, Michael & Diana Forker. 2009. Reconstruction of Morphosyntactic Function: Nonspatial Usage of Spatial Case Marking in Tsezic. *Language* 85(3), 588-617.