

Title: Quantifying phonological complexity and its distribution

Abstract Category: oral/poster

Theme Session: Quantitative Linguistic Typology: State-of-the-Art and Beyond
(Convenor: Harald Hammarström)

Contrary to an earlier orthodoxy that “all languages are equally complex” it is now generally agreed that spoken languages vary considerably in their complexity (Miestamo 2008, Sampson et al 2009, McWhorter 2011). Of course, either position presupposes that some measure or measures of complexity can be meaningfully calculated. With respect to phonological properties several well-developed notions have been proposed as appropriate indices of complexity (see, e.g. contributions to Pellegrino et al 2009). These include simple measures such as the size of consonant, vowel and tone inventories and the variety of permitted syllable structures, as well as somewhat more elaborate calculations taking into account the relative frequency with which contrasts are exploited, relative entropy of phonological information, or relative frequency of occurrence of simpler versus more complex elements. In this paper a number of these measures will be used, first of all, to support the view that phonological complexity in languages is not uniform but varies considerably. Rather, languages appear to be broadly distributed along any scale of phonological complexity (but see Pellegrino et al 2011). Two issues will be particularly explored. The first is primarily methodological and concerns how large geographically and genetically diverse samples of languages can be assembled in a standardized fashion so that comparisons across the languages are valid. Exemplification will include discussion of problems involved in harmonizing analyses of segment inventories and syllable structures, and calculating the length of phonological words. Inventories can be more reliably assessed if closer-to-surface rather than abstract forms are compared. Word-length is best compared when standardized text samples are available, rather than relying on lexical entries. The second issue concerns what factors might underlie the distribution of phonological complexity. Explanatory principles proposed include, among other factors, population size and isolation (e.g. Trudgill 2004, 2011, Hay & Bauer 2007), environmental setting (e.g. Munroe et al 1996 and later, Fought et al 2004 and later), social behavior patterns (e.g. Ember & Ember 2010), and decline of phonological diversity correlated with decline of genetic diversity as modern human populations dispersed from an African origin (Atkinson 2011). The last of these proposals can probably be dismissed since it is not motivated by any independent rationale, and fails to offer any explanation for the initial assumption of maximal complexity in Africa. The varied environmental and social factors suggested seem more promising but generally raise the same question: what time point in the history of the languages concerned is relevant? For example, catastrophic population crashes in Australia and the Americas following European colonization have no obvious effects on the phonological complexity of those languages which survived, nor has the phonological complexity of Spanish been clearly affected by its spread to tropical, sub-tropical and near-polar climates. Using maximally large sample sizes of languages may enable robust correlations to emerge from the noise contributed by such historical accidents. At the least, significant correlations can be demonstrated between some aspects of phonological typology and broad climatic/vegetational zones.