

Johanna Nichols (University of California, Berkeley)

## **Proper measurement of linguistic complexity (and why it matters)**

Abstract:

This paper addresses what I see as gaps in cross-linguistic work on complexity:

- A measure of the full linguistic complexity of a language is generally held to be unattainable at least with current resources, yet cross-linguistic comparisons require some assurance of reasonably comprehensive coverage.
- The kind of complexity that is favored by certain sociolinguistic factors is not what is usually surveyed in studies invoking the sociolinguistic work.
- Either the granularity of cross-linguistic complexity studies is too coarse, or the grammatical coverage is too narrow. Phonological and morphological complexity are very strongly inversely correlated and form opposite worldwide frequency clines, yet surveys of just one or the other, or both lumped together, are used to support cross-linguistic generalizations.
- Linguists need to be able to generate better hypotheses for psycholinguistic and neurolinguistic work, and identify promising targets for computational extraction of complexity figures from corpora.
- Measuring the complexity of polysynthetic languages is neglected.

I propose a tripartite metric that addresses these problems, using a set of different assays across different parts of the grammar and lexicon. Meeting current expectations of sustainability and replicability, the set is reusable, revealing, granular, and (at least mostly) amenable to computational implementation. I test its usefulness to typology and historical linguistics with several cross-linguistic surveys.