

Disentangling the variability of the perfect gram type

Category: oral or poster

Based on Dahl (1985) and Bybee & Dahl (1989) we assume that there are universal “gram types” in the domain of tense and aspect. This paper focuses on one gram type: the perfect (e.g. the English Perfect as in *I have bought a car* and the Indonesian *sudah*). Like Dahl (1985) we assume that perfects exhibit similar distributions in parallel texts which is why a first step in compiling a sample of perfect grams is to extract them from parallel texts by means of collocation measures (here the New Testament is used).

However, at the same time as perfect grams are cross-linguistically similar, they also display considerable cross-linguistic variability which can be roughly ordered into several thematic groups: (i) semantic: perfects can be associated with “iamitive” (forms expressing ‘already’), hodiernal past, evidential, experiential, and resultative; (ii) constructional: synthetic vs. analytic exponence, single marker vs. distributed exponence, auxiliary-based, participle-based, adverb; (iii) combinatorial: special negation strategy, mutual exclusion with some kinds of subordinate clause (e.g., Swahili *-me-* cannot occur in relative clauses), pluperfect (combination with past); (iv) diachronic: there are different grammaticalization sources and the perfect grams are at different stages on the grammaticalization cline (partly reflected by their frequency); and (v) example gram similar: the “Euro”-Perfect is different from the “Sino”-Perfect and from the “Malayo”-perfect; etc.

We compile a database of perfect variability where each factor of variation is treated as a feature of its own. Feature values are either measured in parallel texts or determined manually with reference grammar data. This results in a data table of 100 languages from all continents and 20 features from all five thematic groups outlined above. Unlike in conventional typological databases the default assumption is that many features are correlated since they are all related by being associated with perfect grams. Accordingly we do not use a stratified sample (there is no point in sampling languages without a perfect), but even closely related languages are included since their small differences may be relevant for better understanding the variability. The sample is determined by the availability of the two different kinds of data sources used: electronic parallel texts (N.T.) and reference grammars. After compiling the database we apply a posteriori sampling methods.

Rather than comparing each pair of features individually, statistical methods are used to reduce the dimensionality of variability. The methods used are inspired by dialectometry and register analysis. Our results so far suggest that grams do not fall into neatly delineated subtypes but rather form clusters with graded membership. These clusters correlate to a certain extent with areal distribution and genealogical affiliation (which are no input features for the aggregation analysis).

References:

- Bybee, Joan and Östen Dahl. 1989. The Creation of tense and aspect systems in the languages of the world. *Studies in Language* 13.51–103.
Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.