

Vaulting ambition

John Nerbonne (University of Groningen and University of Freiburg)

The three major subfields that once contended to explain grammar comprehensively – (Theoretical) Grammar, Corpus Linguistics, and Computational Linguistics – are backing away from the claims and even the goals of comprehensiveness. Generative grammatical theory has restricted its recent interest to “narrow” or “core” processes, while alternatives such as Construction Grammar seem to forswear comprehensive studies, at least implicitly, in favor of concentration on one or two constructions at a time (pace programmatic statements such as Kay’s 1995). Corpus Linguistics has always been more modest, and while they draw on ever more impressive amounts of data, up to tens of billions of tokens, many contemporary studies focus on explaining the variation among attested structures, and exact studies have come to rely on multivariate statistics, sometimes producing models for limited phenomena that nonetheless have a dozen or more (significant) variables. Finally, while parse accuracy (really parse and disambiguation accuracy) was once a hotly contested field (in the 1990s) in Computational Linguistics, progress has stagnated, even with models that include millions(!) of independent variables. Current attention is devoted more to developing analysis methods for non-standard data and for languages with poor resources.

These observations about three important subfields suggest that grammar studies have had overly ambitious goals. A less comprehensive agenda will be proposed, one closest to Corpus Linguistics.