



## Superset grammars and failed changes

Rob Truswell

University of Edinburgh

rob.truswell@ed.ac.uk

Joint work with Daniel Lassiter, University of Edinburgh

A large body of research in diachronic syntax has been concerned with gradual grammatical change, and the S-shaped trajectory along which many such changes proceed to completion (Kroch 1989, Yang 2002, Blythe & Croft 2012). However, many "failed changes" have been documented, where a novel grammar is adopted to a certain extent but never spreads throughout the population. In some cases, the change fails despite the fact that the novel grammar has a demonstrable parsing advantage over competitors. That is, the novel grammar (the "superset grammar") fails to spread despite being able to parse an approximate proper superset of the structures parsed by competing "subset grammars". I documented one instance of this pattern in Truswell (2021): there is evidence that northern Early Middle English flirted with a grammar which generated matrix verb-third and embedded verb-second orders, in addition to most standard Middle English V2 orders. However, this very expressive grammar failed to spread, being shortlived and apparently confined to a single dialect. This is an instance of what we call the "superset problem": over-expressive grammars often do not spread. There are many other examples of failed changes, and many of them also show the superset problem.

This talk reports on work in progress in which we compare models of learning in the light of evidence that superset grammars can fail. First, we note that the superset problem, as well as the basic existence of failed changes, pose acute challenges to the model of learning in Yang (2002) based on the Linear Reward-Penalty update schema (Bush & Mosteller 1951). This schema always favours more expressive grammars, and underpins Yang's prediction that changes should never fail on their own terms ("Once a grammar is on the rise, it is unstoppable", p. 132). Any failed change, on Yang's model, must be reducible to exogenous factors, such as contact.

We compare the Linear Reward-Penalty schema to Bayesian learning models, while holding the rest of Yang's framework constant. Bayesian models do not favour overexpressive models (compare "Bayesian Ockham's Razor", Jefferys & Berger 1991; and the "size principle", Tenenbaum & Griffiths 2001). They therefore avoid the most problematic predictions of the Linear Reward-Penalty schema. However, they do not automatically predict the distribution of failed changes, and in particular do not relate the distribution of failed changes to expressivity of a grammar. In the final part of the talk, we evaluate the prospects for extragrammatical accounts of failed changes, focusing on the additional challenge faced by learners, of associating grammatical formatives with well-defined semantic/pragmatic functions.