

***That* or no *that* in English dialect corpora:
Grammaticalization, frequency and complexity in the emergence of grammar**

Sali Tagliamonte (sali.tagliamonte@utoronto.ca)
Department of Linguistics, University of Toronto

Helen Lawrence (hrl103@york.ac.uk)
Department of Language and Linguistic Science, University of York

Jennifer Smith (js40@york.ac.uk)
Department of Language and Linguistic Science, University of York

The contemporary complementizer system of English exhibits widespread variation between *that* and zero. Historical linguistic research on this area of grammar (e.g. Rissanen 1991, Storms 1966, Palander-Collin 1997, Warner 1982) reveals a dramatic trajectory of change that begins with near categorical *that* in Old English (Mitchell 1985), to near categorical *zero* in contemporary spoken data (Thompson & Mulac 1991a; 1991b).

The reflexes of this changeover persist in English dialects, as in (1-2), providing an opportunity to examine grammatical alternation-in synchronic perspective.

1. To prove \emptyset I could do it, yes, I had to prove *that* I could do it. (AYR/b)
2. I wish *that* forty or fifty years ago I'd as much confidence. I wish \emptyset I'd had it then. (MPT/@)

Numerous explanations have been proposed for this change including grammaticalization of epistemic phrases such as *I think* (Thompson & Mulac 1991); the 'complexity principle' (Rhodenburg 1998), lexical effects (Warner 1982, Poutsma 1929), processing (Underhill 1988), and verb frequency (Ungerer et al. (1984). While external influences such as formality and style are also implicated, the number of potential internal linguistic explanations make this variable an excellent candidate for elucidating the combined influence of processing structural and cognitive mechanisms in linguistic variation and change.

In order to explore this process we conduct a large-scale quantitative analysis of *that*/ \emptyset variation using GOLDVARB. The data come from a million word corpus from three peripheral dialect regions of Britain, thus providing an intermediate position on the trajectory of linguistic change between historical and contemporary research as well as the perspective of highly informal vernacular usage. Out of over 4000 tokens, *that* occurs a mere 8.7% of the time amongst the elderly speakers in these corpora, a frequency comparable to American university students in the 1990s, 13.6% (Thompson & Mulac 1991). Compare this to 98% in Middle English (Warner 1982): *that* has shown a dramatic decrease in use over the past 600 years.

However, when we assess the constraint ranking and significance of factors known to influence the choice of *that* throughout Middle and Early Modern English we find significant parallelisms, including the favouring effect of specific matrix verbs (e.g. *think*), intervening material, and the contribution of matrix and complement subject type. This highlights the universality of constraints on *that*/ \emptyset variability, regardless of the frequency of zero or sociolinguistic characteristics of the speakers. Moreover, the multi-dimensionality of the mechanisms underlying this variability is confirmed — most of the factors tested are significant. Their relative importance, however, is not shared across the board, suggesting other influences at work. Further analysis reveals a complex inter-play between lexical verb and verb frequency, above and beyond the influence of epistemic phrases, suggesting that grammaticalization may be only one aspect of the mechanism of change in this area of grammar.

We discuss how these results can be interpreted and show how the varying perspectives provided by synchronic data can contribute important evidence for understanding frequency and constraints in linguistic variation and how these relate to the emergence of grammatical structures.