

## "Are we nearly there yet, Mum?"

Geoffrey Sampson  
University of Sussex

School of Cognitive and Computing Sciences  
University of Sussex  
Falmer  
Brighton BN1 9QH

geoffs@cogs.susx.ac.uk

Tel. 01273 678525

Fax 01273 671320

### ABSTRACT

The reason why "corpus linguistics" is recognized as a distinct sub-discipline with a name of its own is that, until quite recently, much linguistic research was conducted without reference to empirical observation of real-life language use even when it concerned topics to which empirical observation seems relevant. The objections to that state of affairs in terms of scientific validity are well known; the success of the corpus "movement" over recent decades may suggest that the problem is now being dealt with satisfactorily. When Diana McCarthy and I used a historical basis for compiling our anthology of the field (*Corpus Linguistics*, now in press with Continuum of London and New York), the picture we found was one of slow growth from the pioneering work of Charles Fries in the 1950s onwards for several decades, giving way to almost explosive growth and diversification over about the last ten years.

This paper seeks to discover how strong the impact of corpus work on the wider discipline has actually been to date, by counting the proportion of items in random samples from linguistic and computational-linguistic journals over a period of up to fifty years which appear to make empirical claims, and which base such claims on data that are in principle intersubjectively observable. Like the answer often received by the child piping up from the back seat of the car, the answer to our question may be that in reality we have much further left to go than we imagine and hope.