Synchronic and diachronic variation: 
the how and why of sociolinguistic corpora.

Dr. Kate Beeching, Senior Lecturer, 
Linguistics and French, Faculty of Humanities, Languages and Social Sciences, 
University of the West of England, 
Bristol, Frenchay Campus, Coldharbour Lane, 
Bristol BS16 1QY. 
Tel: 0117-344-2385; 
Fax: 0117-344-2820; 
E-mail: Kate.Beeching@uwe.ac.uk

Abstract

Tape-recordings have come of age. The spoken word has been captured, stored on tape and (in as yet 
regrettably restricted amounts) transcribed for a number of decades now and this is a situation which 
can only improve. Where good accompanying demographic data are available, it is beginning to be 
possible to trace linguistic change and to provide empirical evidence of the way that synchronic and 
diachronic variation may be linked. In the past, it has only been possible to guess at long-gone spoken 
manifestations through written, mainly literary, evidence. This paper aims to describe and illustrate the 
potential of (spoken) sociolinguistic corpora for research studies in both synchronic and diachronic 
variation, with reference to French, and to suggest ways in which useful research corpora may be 
established for future generations of scholars.

The Etude Sociolinguistique d’Orléans, currently available from the University of Leuven, Belgium, 
was conducted by a group of university teachers of French at Reading University from 1966-70. The 
resulting electronic corpus comprises approximately 109 hours of spoken French, 902,755 words of 
which have been orthographically transcribed, with a further 13 hours of phonetic transcription. These 
transcriptions are available at: http://bach.arts.kuleuven.ac.be/elicop. Although French researchers, 
notably Claire Blanche-Benveniste and the GARS team in Aix and Mireille Bilger in Perpignan, have 
collected and transcribed spoken French, their corpora are not (yet?) available on line. Beeching’s 
(1980-90) “Bristol” Corpus, which comprises 17.5 hours of orthographically transcribed speech, or 
155,000 words, involving 95 speakers, aged 7 to 90 years with a balanced range of educational 
backgrounds, though small in scale, may be accessed at 
http://www.uwe.ac.uk/facults/les/staff/kb/CORPUS.pdf.

Beeching (forthcoming b) suggests that pragmaticalisation is a process of language change during 
which words, used in everyday social interaction, shift in meaning or accrue a certain social semiotic, 
become habituated in that usage and are propagated because of a new fashion or prestige which is 
attached to them. Spoken corpora and corpus tools are an excellent heuristic in charting distributional 
frequencies or probabilistic factors. Andersen (2000) suggests that the upsurge of innit and like in the 
COLT Corpus of adolescent English may be more than age-grading. The present paper will present 
broadbrush preliminary evidence with respect to the evolution of selected pragmatic particles in 
French.