



**HAL**  
open science

## Introduction

Henry Tyne, Virginie André, Alex Boulton, Christophe Benzitoun, Yan Greub

► **To cite this version:**

Henry Tyne, Virginie André, Alex Boulton, Christophe Benzitoun, Yan Greub. Introduction. Henry Tyne; Virginie André; Alex Boulton; Christophe Benzitoun; Yan Greub. French through Corpora: Ecological and data-driven perspectives in French language studies. Introduction., Cambridge Scholars Publishing, pp.341, 2014. hal-00940071

**HAL Id: hal-00940071**

**<https://hal.science/hal-00940071>**

Submitted on 12 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Henry Tyne, Virginie André, Alex Boulton, Christophe Benzitoun & Yan Greub. (2014). Introduction to In H. Tyne, V. André, A. Boulton, C. Benzitoun & Y. Greub (eds), *Ecological and Data-Driven Perspectives in French Language Studies*. Newcastle: Cambridge Scholars, p. 250-256.**

This book looks at the French language through corpora, and comprises four parts dealing respectively with diachrony, syntax, sociolinguistics and issues arising in the learning and teaching of French. Each part is headed by a chapter that provides an overview of the given field in relation to the themes running through the book. Other than contributing to our general understanding of the French language today, this book specifically addresses the use of corpora for the study of language and the links between tools, methods and analyses. How do we use corpora? What are the underlying theoretical and/or methodological considerations? How have these changed our way of formulating linguistic descriptions? What are the implications for descriptive accounts of French? What are the applications of corpus studies?

Each chapter focuses on specific aspects of French and addresses (often indirectly) issues such as the ways in which researchers use existing resources, the reasons for producing new resources, or questions arising from different types of data use. One aim is to challenge or complete existing work, not least in relation to the possibilities that are now made available through corpus use.

Corpora provide data, and a common theme throughout the book is one of empiricism. A distinction is sometimes drawn between corpus-based and corpus-driven approaches, and this is apparent here: a corpus-based study looks to gather findings that test a certain idea or model, whatever its scope or ambition; a corpus-driven study is more ecological insofar as it aims to build conclusions on the sole basis of the findings (Tognini-Bonelli 2001). These are perhaps two ends of a continuum, and in many cases there will be a continuous interaction between data consultation and the questions one has, each influencing the other at all stages, from corpus compilation through to final outcomes. But in all cases it is important not to lose sight of the fact that language is not just a neutral collection of data, hence the inclusion of the concept ecology in the title.

Ecology has become something of a buzz-word of late in many spheres. The rationale for the ecological leitmotif here derives from a shared interest in language and environment: ecology is to do with understanding language not as an abstract system but as an integral part of human existence (Haugen 1972). Ecology also refers to the authentic nature of the data and issues to do with their collection, transcription and editing, mark-up and tagging, or any other treatment resulting from human intervention whether deliberate or not.

Corpora are used nowadays in many areas of linguistics (and beyond), often for purposes not anticipated by the original designers. Burnard (2002: 67), for example, readily admits that the linguists behind the landmark BRITISH NATIONAL CORPUS (BNC) never imagined that it would be bought by individuals, nor that it could be of any direct use for teachers and learners. Nonetheless, where large corpora are now available freely on line, teachers and learners tend to make up the majority of users, as is the case for the BYU or Lextutor interfaces to several

large corpora in different languages. In some cases, corpora have been used as simple repositories of linguistic ‘facts’, almost as archives or databases to dip in to when needed and in whatever way required. This approach can be useful in philology, for example, where the corpora of available texts are often limited and difficult to gather together into a single coherent corpus. Such activity may be considered peripheral to corpus linguistics as an identifiable discipline – or even by some as not true corpus linguistics at all. It does, however, have the virtue of quietly filtering corpus linguistics into ‘the rest of linguistics’, thereby making it more accessible to a wide research community and extending the influence of its methods and results among the general field. This has also given rise to a debate about whether corpus linguistics represents ‘merely’ a methodology precisely because it is open to so many uses in so many fields (McEnery et al. 2006: 8), or something rather more (Tognini-Bonelli 2001), given its tremendous impact on all that we know about language use – not for nothing do McCarthy (2008) and others talk of a corpus revolution.

Turning specifically to French, Gadet (2009: 115) points out that while “for many contemporary linguists, ‘doing linguistics’ means working with corpora [...] the study of French distinguishes itself in this respect from that of other languages”. She gives two reasons for the comparative underuse of corpora in French studies: the fact that sizeable corpora for French have been late in coming in comparison to other languages (in Europe, at least); and the non-centrality of many existing French corpora (it is of note that there is no French national corpus). Gadet does concede, though, that the “broadening of the data” over the last few decades has brought about a number of interesting changes, as different research methodologies and theoretical backgrounds find common ground. This is particularly true perhaps for the study of spoken grammar (Blanche-Benveniste 2007: 129), an area where French corpus work has proved particularly innovative through the work of the GARS (*Groupe aixois de recherches en syntaxe*) school of linguists. Blanche-Benveniste notes that this is beneficial for grammatical description of the language as a whole, both in terms of suggesting new methods and for discovering new phenomena.

Though we might optimistically imagine the future will hold greater collaboration between sectors of linguistics, positions can become entrenched and exchanges rather heated, as in the “bootcamp debate” (see Worlock Pope 2010) between two camps of corpus linguists – those who are keen to see interaction with cognitive linguistics, and those who are sceptical of any compromise on what may be perceived as an epistemological or even ethical divide between empiricism and intuition/introspection. Corpus linguists may be understandably wary, having been branded as “butterfly collectors” (Chomsky 1979) lacking theory – corpus linguistics as a field simply “doesn’t exist” claims Chomsky in an interview reported in Aarts (2001: 5). On the other hand, it is perhaps worth remembering that linguistics has been through a number of pendulum swings over the last century or so (for reasons of technical and methodological possibilities and indeed for cultural or political reasons as much as for theoretical reasons), and that it may be more profitable to seek common ground and collaboration between different approaches (e.g. Fillmore 1992). In other words, should butterfly collectors hunt in a protected area or should they open up to different areas and points of view? Should linguists who do not go in for butterfly collecting become interested in the findings of their collector colleagues, and vice versa? Answers to these questions (any many more) represent a major challenge for the ecology of the linguistics world. As Labov

(1971: 119) pointed out several decades ago now, “it is not necessary for everyone to use the same methods – indeed it is far better if we do not [...] Data from a variety of distinct sources and methods, properly interpreted, can be used to converge on right answers to hard questions”. Labov also draws our attention to the “cumulative principle” of linguistic research, whereby “the more that is known about language, the more we can find out about it” (p. 98). And so it is that this book offers a series of studies that come together in their concern for furthering our understanding of the French language, of its uses, its forms, its variation, its acquisition, etc., as well as in promoting ecological approaches to using corpora for studying these questions.

### **Bibliography**

- Aarts, B. 2001, “Corpus linguistics, Chomsky and fuzzy tree fragments”, In C. Mair & M. Hundt (eds.), *Corpus Linguistics and Linguistic Theory*, Amsterdam: Rodopi, 5-13
- Blanche-Benveniste, C. 2007, “Corpus de langue parlée et description grammaticale de la langue”, *Langage et société* 121-122(3), 129-141
- Burnard, L. 2002, “Where did we go wrong? A retrospective look at the British National Corpus”, In B. Kettemann & G. Marko (eds.), *Teaching and Learning by Doing Corpus Analysis*, Amsterdam: Rodopi, 51-70
- Chomsky, N. 1979, *Language and Responsibility* (based on conversations with Mitsou Ronat), New York, Pantheon
- Fillmore, C. 1992, “Corpus linguistics or computer-aided armchair linguistics”, In J. Svartik (ed.), *Directions in Corpus Linguistics*, Berlin: Mouton de Gruyter, 35-60
- Gadet, F. 2009 “Stylistic and syntactic variation: introduction”, In K. Beeching, N. Armstrong & F. Gadet (eds.), *Sociolinguistic Variation in Contemporary French*, Amsterdam: John Benjamins, 115-120
- Haugen, E. 1972, *The Ecology of Language: Essays by Einar Haugen* (ed. A. S. Dil), California: Stanford University Press
- Labov, W. 1971, “Some principles of linguistic methodology”, *Language in Society* 1, 97-120
- McCarthy, M. 2008, “Accessing and interpreting corpus information in the teacher education context”, *Language Teaching* 41(4), 563-574
- McEnery, T., R. Xiao & Y. Tono 2006, *Corpus-Based Language Studies: An Advanced Resource Book*, London: Routledge
- Tognini-Bonelli, E. 2001, *Corpus Linguistics at Work*, Amsterdam: John Benjamins
- Worlock Pope, C. (ed.) 2010, “The bootcamp discourse and beyond”, *International Journal of Corpus Linguistics* 15(3)