

Introduction to this special issue

Stefan Th. Gries
University of California, Santa Barbara

1 Introduction

If one asks a corpus linguist how long the field has been around, two answers are heard most often. One would say that corpus linguistic methods have been around for quite some time, would point to early Bible concordances or Käding's (1897) work, would adduce European comparative linguists and American structuralists from the first half of the 20th century as additional examples, etc. The other would say that corpus linguistics really only began to take shape with, on the European stage, Firth's (1951) work on collocation or the work on the Survey of English Usage and/or, on the American stage, Fries's (1952) work on spoken American English, etc.

Regardless of which of these points of view one holds – they are probably both correct from some points of view and corpus linguists might adopt either one over where necessary to make a particular rhetorical move – it is probably no exaggeration to say that it is only over the last 20 years or so, that corpus linguistics has really taken off and developed into one of the most widely-used methods in linguistics. This is visible on many different levels:

- on the level of **resources**: technological developments took place that facilitated the creation of the first mega corpora of the kind exemplified by the British National Corpus;
- on the level of the role that corpus data play in the development and refinement of more comprehensive **theories of language** i.e. in work going beyond mere description. While such developments are still resisted by some – as is the view of corpus linguistics as a 'mere' methodology – (cf. Worlock Pope's (2010) the special issue of the *International Journal of Corpus Linguistics* on the so-called bootcamp discourse) the ways in which corpus linguistics on the one hand and cognitive linguistics and psycholinguistics on the other hand feed into each other is hard to ignore or resist;

- on the level of **statistical methodology**: the overall developmental trend in linguistics towards more quantitative methods can – finally! – also be seen in corpus linguistics. In fact, I have argued elsewhere that, since corpus linguistics is essentially based on nothing but distributional and quantitative data, the field should have been the one to lead the current quantitative revolution in linguistics rather than leaving this honor to, mainly, psycholinguistics ...;
- on the level of **competences by practitioners** of the field: many practitioners in the field have long been constrained by a few commercial corpus analysis tools, which limited researchers' ability to think outside of the (software tool) box, the field is now shaping up and many researchers turn to more versatile, powerful, and elegant tools such as the Natural Language Toolkit (cf. <<http://www.nltk.org>>) or programming languages (cf. Gries 2009 for one example), which finally allows the field to handle the complex types of data in more appropriate ways than was possible before.

By now, corpus linguistics is well established: the field has several international peer-reviewed journals, its own book series with international publishers, a lively conference circuit, and corpus-based methods have contributed to research in most sub-disciplines of linguistics. This also means that researchers don't have to include in their papers justifications or even defenses of why they are using corpus data anymore – corpus linguistics has succeeded to become many of its methods are now mainstream (in a positive sense).

2 This special issue

In spite of its impressive success story, corpus linguistics is still in need of maturation and further evolution, and this special issue is devoted to this topic. When I was invited to guest-edit a special issue of the *Brazilian Journal of Applied Linguistics (BJAL)* on corpus linguistics, I quickly decided to *not* edit the typical kind of issue in which 'standard' research articles present nice and significant results – my goal became to edit a special issue that outlines where the field of corpus linguistics should go next, an issue that, so to speak, provides direction to the field just as good plenary addresses would do. I thought it was particularly fitting that such a special issue would appear in an open-access journal, which makes the contributions more accessible than copyright restrictions of some commercial journals often allow for so I was delighted that the editorial team of BJAL accepted this plan.

The next step consisted of identifying a range of fields which I considered benefited much from, and contributed much to, corpus linguistics as well as persuading a range of prominent scholars in these fields to contribute to this special issue a paper that answered the following question:

In your area of research and in your work with corpora – and I am writing to you because of your work in _____ – where do you think the field of corpus linguistics has to go and/or mature, and why? What are developments in terms of resources, standards, technology, methods, etc. that you think are essential and/or at least desirable, and why, or what can we do then?

I was very lucky to receive affirmative and encouraging responses from high-profile colleagues for a number of linguistic areas or sub-disciplines, which are listed in Table 1. Each of the papers outlines answers to the above guiding questions in its own way, usually providing a short state-of-the-art overview, followed by perspectives, recommendations, lists of desiderata, case studies, and much more that should give the field food for thought for the foreseeable future – they certainly did that for me.

As a final note, a heartfelt ‘thank you!’ is due to my associate editor at *BJAL*, Heliana Ribeiro de Mello, without whom this special issue would not have materialized. And, I would of course also like to express my sincere thanks to the contributors, who agreed to contribute to a special issue with a somewhat unusual focus and who sent in thoughtful and inspiring papers that clearly outline how corpus linguistics can evolve further in ways that no single author ever could. If this special issue gets you thinking and planning, they deserve all the credit for that.

TABLE
Overview of this special issue

Area/sub-discipline: corpora and...	Authores
... quantitative research/methods	R Harald Baayen (University of Tübingen)
... metaphor research	Tony Berber Sardinha (Catholic University of São Paulo)
... sociolinguistics	Tyler Kendall (University of Oregon, Eugene)
... multi-modal data	Dawn Knight (University of Nottingham)
... historical linguistics	Merja Kytö (Uppsala University)
... second/foreign language learning	Fanny Meunier (Catholic University of Louvain)
... discourse pragmatics	Massimo Moneglia (University of Florença)
... cognitive linguistics	John Newman (University of Alberta, Edmonton)
... dialectology	Benedikt Szmrecsanyi & Christoph Wolk (Freiburg Institute for Advanced Studies)

References

- FIRTH, J.R. *Papers in linguistics*, 1934-1951. Oxford: Oxford University Press, 1951.
- FRIES, C. C. *The structure of English: an introduction to the construction of English sentences*. New York: Harcourt Brace, 1952.
- GRIES, St.Th. *Quantitative corpus linguistics with R: a practical introduction*. London / New York: Routledge, Taylor & Francis Group, 2009.
- KÄDING, F.W. *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz: no publ., 1897.
- WORLOCK POPE, C. (Ed.). The bootcamp discourse and beyond. Special issue of the *International Journal of Corpus Linguistics*, v. 15, n. 2, 2010.