

Big data – small data – abstractness – concreteness: Critical approaches to methods in Corpus Linguistics

Workshop at BICLCE2017 (7th Biennial International Conference on the Linguistics of Contemporary English) in Vigo, 28-30 September 2017

Convenors: Ole Schützler & Julia Schlüter, University of Bamberg
(ole.schuetzler@uni-bamberg.de; julia.schlueter@uni-bamberg.de)

With Gries (2013: 4) we hold the view that strongly quantitative approaches and less quantitative (or qualitative) approaches to corpus research are not – or should not be – at odds, but must be seen and used in combination, as two sides of the same coin. However, the linguistic community still seems to be divided into two camps: those who use a variety of (relatively abstract) sophisticated statistical techniques – recent examples being Gries & Bernaisch (2016), Ruetten et al. (2016), Szmrecsanyi et al. (2016), Wolk & Szmrecsanyi (2016), Grieve (2014) and papers in Krug & Schlüter (eds., 2013) – and those who predominantly apply methods at a lower level of abstraction.

In this workshop we aim to bridge this apparent gap. We invite researchers to present case studies that work towards a 'best practice' in corpus linguistics, illustrating how state-of-the-art quantitative methods can be combined with better-established, traditional approaches and thus remain transparently grounded in the features at their basis. As the title suggests, we are particularly interested in dichotomies such as 'big' (and 'dirty'?) vs 'small' (and 'tidy'?) data, 'abstract' vs 'concrete' or the like. We strongly encourage what we have called 'critical approaches' – that is, in focusing on methodological aspects, contributors should address some of the following questions:

- How did they arrive at the chosen methodological solution? In particular, how did the research question or the data structure inform methodological decisions?
- Which alternative approaches to the data were tried and dismissed, or could have been considered?
- How exactly do the different levels of the analysis (e.g. abstract vs concrete, inferential vs descriptive) contribute to the whole?
- What caveats apply to the quality of the data on which analyses are based?
- What is (technically) difficult about the chosen methods?
- To which other research questions or types of data could the presented methods be applied?
- How can the chosen methods be (better) explained and made (more) plausible to the community?

The workshop thus seeks to foster discussion within the community and expressly encourages (self-)critical evaluations and discussions, based on detailed analyses and specific research questions. Relevant methods may include (but are not restricted to) types of inferential statistics, multi-dimensional approaches, and visualization techniques. In the spirit of the BICLCE conferences, case studies should focus on phenomena in contemporary English(es).

The workshop will consist of full papers, which will be allotted 20 minutes for presentation (plus 10 minutes for discussion). The deadline for submission of abstracts (up to 500 words, excluding title, references and keywords) is 15 February 2017. Notification of acceptance will be sent out by 15 March 2017. Abstracts should be sent to corpus-methods.eng-ling@uni-bamberg.de. We plan to publish the papers from this workshop and further commissioned contributions in an edited volume, or in a special issue of a high-ranking journal.

References

- Gries, Stefan Th. 2013. *Statistics for linguistics with R: A practical introduction*. 2nd edition. Berlin: Mouton de Gruyter.
- Gries, Stefan Th. & Tobias Bernaisch. 2016. Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide* 37(1): 1–25.
- Grieve, Jack. 2014. A multi-dimensional analysis of regional variation in American English. In: Tony Berber Sardinha & Marcia Veirano Pinto (eds.), *Multi-dimensional analysis 25 years on: A tribute to Douglas Biber*. Amsterdam/Philadelphia: Benjamins. 3–33.
- Krug, Manfred & Julia Schlüter (eds.). 2013. *Research methods in language variation and change*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2): 109–137.
- Ruette, Tom, Katharina Ehret & Benedikt Szmrecsanyi. 2016. A lectometric analysis of aggregated lexical variation in written Standard English with Semantic Vector Space models. *International Journal of Corpus Linguistics* 21(1): 48–79.
- Wolk, Christoph & Benedikt Szmrecsanyi. 2016. Top-down and bottom-up advances in corpus-based dialectometry. In: Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.), *The future of dialects: Selected papers from Methods in Dialectology XV*. (Series: *Language Variation* 1). Berlin: Language Science Press. 225–243.