FRIAS

Lunch Lecture 2017/18:
Quantitative vs. qualitative approaches across sciences

11 January 2018

Reflecting on the quantitative turn in linguistics

Bernd Kortmann

UNI FREIBURG

Structure

1. Introduction
2. The Quantitative Turn

   2.1 Corpus linguistics
   2.2 Two applications: Aggregation and metricization
   2.3 Probabilistic linguistics

3. From Corpus to Cognition? (or: Do corpora mirror psychological reality?)
4. Major Caveats
5. A Quantitative Crisis in Linguistics?
6. Conclusion

4. Major Caveats

- **don't** see language or some variety of a given language exclusively through the lens of (available relevant) corpora, as this may artificially narrow (or block) the object of study
- **don’t** take induction (e.g. in corpus-driven research) and statistics to be objective *per se*
- **don’t** take the search for correlations to be an end in itself
- **don’t** confuse correlations with causes
- **do** everything that is necessary (!) for achieving a maximum of methodological transparency, rigour, statistical significance, robustness, reproducibility, falsifiability and, ultimately, explanatory power and mileage for linguistic theory-building (e.g. no statistical cherry-picking), **but**

- **don't** do everything that is statistically possible just because you can do it (heaping minor, or worse: irrelevant, detail on minor/irrelevant detail), even less as a remedy for an imperfect data set or inconclusive data analyses

  → **don’t** multiply statistical testing beyond necessity
  → **no** statistics-driven research! (statistic machinery must not determine the research question), i.e.
  → **don't** let the tail wag the dog

- **do first** formulate intelligent research questions and a solid research- and theory-grounded set of hypotheses, which can **then** be statistically tested/falsified, yet

- **don’t** take statistical compatibility with a given hypothesis immediately as (sufficient) proof

- **don't** commit the “from-corpus-to-cognition fallacy” and conduct, as is appropriate for the research question, experimental studies alongside corpus studies (-> multi-method design)

- however powerful and promising the corpus revolution and quantitative turn may be (or be felt to be): **don’t** forget the rich inventory of theories and **(largely qualitative)** methods which (schools of) linguists have developed and refined over many decades for the analysis of natural language and communication (-> a caveat which applies also vice versa)

- **never** forget the human factor behind everything in communication and language: the intentions, needs, constraints of natural language users in spontaneous verbal interaction

5. A quantitative crisis in linguistics?

- Workshop at ISLE 5 (London, 17-20 July, 2018):
  Sönning/Werner (Bamberg): “The ‘quantitative crisis’, cumulative science, and English linguistics”

-> some focal problems identified in the broader discourse:

- overreliance on a single influential theory determining an entire research paradigm
- non-reproducibility of studies
- high rates of false-positive findings in published research
- lack of transparency as regards methodology and analysis
- negligence of replication studies as “unoriginal” (and unprestigious)

- overreliance on a single influential theory determining an entire research paradigm? **-> NO**

- with regard to the four methodological problems:

  all relevant in linguistics, too, but there is reason for optimism

  → not least due to awareness raising at a fairly early point and

  → the readiness in the publishing of linguistic research to go by the principles of open science, e.g.

    - accessibility of data & analyses
    - reproducibility of studies and statistical tests
    - no fear of publishing “negative” results

6. Conclusion

- Overall, the QT in linguistics has been a largely positive development. It has many strengths and great potential **always provided** corpus analyses and statistical techniques are selected and conducted/applied cautiously and in a highly reflected manner,

- heeding constraints, challenges and dangers, such as
  - the limits of what corpora can tell us about cognition
  - the risks of simplistic / naïve statistical analysis: cherry-picking, confusing correlations (at worst: spurious correlations) with causes

…but there is still some way to go

- The crucial point and task for linguists committed to the QT:
  „to boldly go where the others already are“

- In the concert of the quantitative sciences, linguistics is still a (somewhat little naive) newcomer, but if it wants to be taken seriously it needs to stand up to the rigorous standards of these sciences -> this is still a quite hard and long way to go

- Besides basic and advanced statistical training as part of degree and doctoral training programmes, besides statistics-savvy linguists, the members of each and every linguistics department should also have the possibility of consulting with professional (ideally linguistics-savvy) statisticians!

- If these conditions are fulfilled, linguistics will become an even more respected showcase of the **Digital Humanities**, and may truly succeed in **bridging the disciplinary boundaries** to the STEM sciences, especially to the behavioral and neuro-sciences.

- Finally: the QT in linguistics has NOT been to the detriment of qualitative approaches; rather: a productive relationship characterized by mutual respect, reinforcement and benefit

THANK YOU!

References

- Antes, Gerd. Die Medizin im Datenrausch. *FAZ* Nr. 1/2. Januar 2018/S. 9.
- Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne Zeschel. 2010. Cognitive corpus linguistics: Five points of debate on current theory and methodology*. Corpora* 5(1), 1–27.
- Biber, Douglas. 2010. Corpus-Based and Corpus-Driven Analyses of Language Variation and Use. In: Heine/Narrog, eds. 159-191.
- Blumenthal-Dramé, Alice. 2012. *Entrenchment in Usage-Based Theories: What Corpus Data Do and Do not Reveal about the Mind*. Boston/Berlin: De Gruyter Mouton.
- Bod, Rens. 2010. Probabilistic linguistics. In: Heine/Narrog, eds. 633-662
- Divjak, Dagmar, Natalia Levshina and Jane Klavan. Cognitive Linguistics: Looking back, looking forward. *Cognitive Linguistics* 27(4): 447-463.
- Gries, Stefan Th. 2015. Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions. *Language & Linguistics* 16: 93-117.
- Gries, Stefan Th. 2013. Elementary statistical testing with R. In: Krug/Schlüter, eds. 361-381.
- Heine, Bernd and Heiko Narrog, eds. 2010. *The Oxford Handbook of Linguistics Analysis.* Oxford: Oxford University Press.

- Kortmann, Bernd and Kerstin Lunkenheimer, eds. 2012. *The Mouton World Atlas of Variation in English.* Berlin/New York: De Gruyter Mouton.
- Krug, Manfred and Julia Schlüter, eds. 2013. *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press.
- Kunter, Gero. 2017. Processing complexity and the alternation between analytic and synthetic forms in English. Postdoctoral thesis, University of Düsseldorf.
- Mannila, Heikki, Terttu Nevalainen and Helena Raumolin-Brunberg. 2013. Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables. In: Krug/Schlüter, eds. 337-360.
- McMahon, April and Warren Maguire. 2013. Computing linguistic distances between varieties. In: Krug/Schlüter, eds. 421-432.
- Szmrecsanyi, Benedikt and Bernhard Wälchli, eds. 2014. *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*. Berlin: de Gruyter
- Szmrecsanyi, Benedikt. 2013. Analyzing aggregated linguistic data. In: Krug/Schlüter, eds. 433-455.
- Wälchli, Bernhard and Benedikt Szmrecsanyi. 2014. Introduction: The text-feature-aggregation pipeline in variation studies. In: Szmrecsanyi/Wälchli, eds. 1-25.