Organisation :

Sascha DIWERSY   **Céline POUDAT**

UNTANGLING ASSOCIATIONS

ADVANCES IN COLLOCATION AND KEYWORD ANALYSIS

22 sept. 2023
Stefan Th. Gries
Université Paul-Valéry Montpellier 3

For the past few decades, assessing the adequacy of association measures (AM) - be it in the domain of keywords, collocations or collostructional analysis - has been one of the most important strands of corpus linguistic research. Most of the work in this area focussed on finding the one best measure as a trade off between statistical appropriateness and ease of computational implementation, thus reflecting the practice of corpus linguists predominantly favouring data sets based on one single AM for the sake of simplicity. This one-AM-fits-all approach, however, suffers from the fact that the most consensual and widespread AMs, such as log-likelihood ratio (G2), conflate different strands of information (viz. frequency and strength of attraction), whereas data sets used to investigate association phenomena (keywords of target corpora as well as lexical or lexico-grammatical cooccurrence) should be designed to integrate several distinct dimensions, as has been recently pointed out by Stefan Th. Gries (2019, 2021).

Starting from Gries's proposal of the approach called "tupleization", this conference will be the occasion to discuss the present state of the art and possible innovations within the realm of methodological frameworks and studies encompassing keyword, collocation or collostruction analysis. It gathers scholars from different areas of research ranging from corpus linguistics and NLP to Digital Humanities and Textometrics as practised in the French tradition of Discourse Analysis.

STATISTICS FOR LINGUISTICS WITH R

Programme

- **10h-11h**, Stefan Th. Gries, Tupleization in corpus linguistics: how and why
- **11h05-11h45**, Martin Hilpert, Why are grammatical elements more evenly dispersed than lexical elements? A reanalysis with a new dispersion measure
- **11h45-12h25**, Ludovic Lebart, Dealing with low frequencies or high discrepancies of lexical frequencies: How to adapt the tools of textual data analysis to corpora of poems and lyrics
- **12h30- 15h15**, Lunch break

- **14h20-15h**, Bénédicte Pincemin, The Specificity Measure in Textometry: a Hermeneutic Use of the Fisher's Exact Test
- **15h-15h40**, Christof Schöch, Evaluating Measures of Keyness: A Perspective from Computational Literary Studies
- **15h40-16h20**, Ludovic Tanguy & Filip Miletic, Measuring semantic specificities across corpora: looking for semantic shifts in Quebec English
- **16h20- 17h**, Coffee break

- **17h-18h**, Panel discussion

The conference will be held on Friday 22 September 2023 at Paul-Valéry University Montpellier (campus Saint-Charles) and online via Zoom
Access link: https://univ-montp3-fr.zoom.us/j/95367941544).
Attendance is free of charge.

References
Gries, Stefan Th. (2019). 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). International Journal of Corpus Linguistics 24 (3), 385–412.
Gries, Stefan Th. (2021). A new approach to (key) keywords analysis: Using frequency, and now also dispersion. Research in Corpus Linguistics 9 (2), 1–33.

INSCRIPTIONS

ADRESSE