Mixed Methods approaches in Corpus Pragmatics?
A case study about Epistemic and Evidential Markers in French

Jérôme Jacquin – University of Lausanne

Corpus linguistics has long been associated with the quantitative analysis of textual data. Today, linguists are more willing to admit not only the diversity of language materials on which to apply these exploratory techniques but also the reality and interest of mixed methods articulating quantitative and qualitative approaches (e.g. Riazi 2016; Teddlie and Tashakkori 2011; Angouri 2010). While the first limitation has been overcome through the emergence, aggregation and (still very relative) mutualization of oral or even multimodal corpora (e.g. Adolphs and Carter 2013; Avanzi, Béguelin, and Diémoz 2016; Baldauf-Quilliatre et al. 2016), the second limitation has been bypassed by a more general epistemological reflection on the aims of corpus linguistics (e.g. Egbert, Larsson, and Biber 2020; Mayaffre 2005; Rastier 2004).

That said, the pragmatic dimension is still largely underexplored in corpus linguistics, at least compared to the more traditional phonological, morphosyntactic, transphrastic and textual levels, despite some notable exceptions coming from the still-emerging field of "corpus pragmatics" (e.g. Aijmer and Rühlemann 2014; Romero-Trillo 2008; Rühlemann 2019). This is possibly due to the diversity and nature of the observables involved, as well as the theoretical and methodological heterogeneity that characterize the field of pragmatics. Some influential currents, such as ethnomethodology-inspired conversational analysis, are actually firmly opposed to quantification (Schegloff 1993).

For this talk, I will focus on evidential markers and – more generally – the linguistic domain of 'epistemicity' (Boye 2012), which have been studied quite extensively in French Linguistics, but mostly from a semantic and/or syntactic perspective and by using either invented or decontextualized examples. Corpus studies remain scarce, they often focus on one or few markers, and most of the time they do not integrate pragmatic factors such as the sequentiality of interaction (i.e. cotextual features), the discourse genres and constraints (i.e. contextual features) and the multimodal dimension of communication (i.e. polysemiotic features). The paper stems from an ongoing, 4-year SNF research project whose goal is to study French epistemic and evidential markers as they emerge in a 28h video-recorded corpus documenting public debates, TV debates, and work meetings. After a short introduction to the main theoretical and methodological options adopted by the project to analyze – both quantitatively and qualitatively – a wide range of French epistemic/evidential markers, the paper provides a case study on 328 tokens of verbs and adverbs of appearance (verbs: *sembler, paraître, avoir l'air, avoir l'impression et donner l'impression*; adverbs: *apparemment, évidemment, manifestement, visiblement*). Quantitative analysis focuses on the general distribution of the verbs and adverbs by discourse genres, communicative roles, and sequential positioning; Multimodality – notably (shifts of) gaze direction and gestures associated with the markers – is also considered. Qualitatively, the paper examines the most frequent expressions in the corpus (e.g. *il me semble (que)* [it seems to me (that)]), by scrutinizing variations of meaning and functions related to sequential positioning (e.g. initiative versus reactive position) and types of scope (i.e. in terms of factuality). The talk concludes with some general remarks about the interest of combining qualitative and quantitative methods for the study of epistemic and evidential markers at the interface of semantics and pragmatics.