
Identifying usage patterns in corpus data: A multifactorial account of epistemic and evidential markers in English and French

Karolina Krawczak*¹, Gilles Col*², Frédérique Mélanie-becquet³, and Thierry Poibeau³

¹Faculty of English, Adam Mickiewicz University, Poznan – Pologne

²FoReLL, Université de Poitiers – FoReLL, Université de Poitiers – France

³LaTTiCe, UMR 8094 (CNRS / ENS / Université Paris 3) – LaTTiCe, UMR 8094 (CNRS – France)

Résumé

This is a contrastive corpus-based analysis of a set of epistemic and evidential markers in British English and French. The markers under investigation are *I think*, *I believe*, *it seems*, *it appears* and *je pense*, *je croie*, *il semble* and *il paraît*. The goal is to determine the usage patterns for the markers across both languages through a combination of detailed manual analysis and multivariate statistical modeling. Patterns, in this context are understood as statistical tendencies.

The approach adopted here is known as the multifactorial feature analysis (Geeraerts *et al.* 1994; Glynn & Robinson 2014). The underlying assumption behind this method is that contextualized language use is an index of language structure. In other words, by way of generalization across many observations, it is possible to reveal frequency-based patterns of language use. These patterns are identified in terms of associations and disassociations of contextual clues.

The data for the study, amounting to 120 occurrences per marker, were extracted from the blog components of *Sketch Engine* with the use of regular expressions. All the examples were manually annotated for a range of syntactic and semantico-pragmatic variables. The metadata will be submitted to multivariate statistical modeling in order to identify the frequency-based patterns associated with the investigated markers of epistemicity and evidentiality.

We expect that the agentive markers *I think* and *I believe* and their respective equivalents in French will be more frequently associated in use with situations referring to personal and non-verifiable reality. Social reality and verifiable information, in turn, are here expected to be more readily correlated with the non-agentive markers in the two languages.

Geeraerts, D. et al. 1994. *The Structure of Lexical Variation*. Mouton.

Glynn, D. & J. Robinson (eds.). 2014. *Corpus Methods for Semantics*. Benjamins.

Mots-Clés: evidentiality, epistemicity, constructions, multifactorial analysis

*Intervenant