

Titre – Modulations mathématiques et descriptives pour l'émergence de parcours interprétatifs dans des corpus textuels. Application au Corpus Mendès France (1922-1982)

Résumé – La logométrie propose le rapprochement des pratiques lexicométriques – méthodes de comparaison statistique des textes – et des théories récentes de la linguistique textuelle à travers de nouveaux observables et de nouveaux modes d'interprétation. Dans ce cadre, nous proposons une mise en œuvre statistique de la sémantique textuelle interprétative sur un corpus de textes de Pierre Mendès France (1922-1982) : sont construits des parcours interprétatifs qui vont du global – le corpus – vers le local – ses partitions contrastives – en considérant l'influence de la situation historico-sociale sur les textes. Ainsi, deux grandes variables qui traversent le corpus sont testées : la variable chronologique et la variable générique. Nous établissons que le corpus est structuré par ses hautes fréquences et études, de manière privilégiée, le vocabulaire banal. On y applique deux traitements cooccurentiels : l'environnement cotextuel et la cooccurrence asymétrique. Le premier aboutit à la description de fonds textuels et intertextuels sur lesquels se détachent des formes sémantiques. La seconde décrit des contours rythmiques de variation lexicale qui semblent associés, dans notre corpus, à des visées argumentatives distinctes : informatives ou persuasives-explicatives.

Mots clés – logométrie, linguistique de corpus, classification émergente, cooccurrence, sémantique textuelle, émotion, argumentation, discours politique, Pierre Mendès France

Title – Mathematics and descriptive modulations for interpretative ways in text corpora. Application to the Mendès France corpus (1922-1982)

Abstract – Statistical natural language processing (SNLP) are quantitative methods of text comparison. With logometry, SNLP and recent linguistic theories are combined to produce new tools for textual observations. Our logometric practice is a statistical application of François Rastier's semantic theories on a corpus of Pierre Mendès France's discourses. Our statistical way determine particularity by generality : words by paragraph, paragraphs by text, texts by corpus. We want to describe how situation – historical or social practice – influe on textual structure. To study the distribution high frequency words in the corpus, our works use two innovent tools : the cotextual environnement and the asymmetric cooccurrence. The former produces "smallworlds" : description of lexical semantic networks in texts ; the latter describes rythms of lexical variations. In our corpus, these rythms are associated to distinct argumentative values : persuasive-explanatory vs informative.

Keywords – Statistical Natural Language Processing, Statistical Semantic, Lexical Semantic, Text Meaning, Cooccurrence, Emotion, Political Discourse, Pierre Mendès France