

Collocation networks in the language of crime journalism

Réseaux de collocations dans le langage du journalisme d'investigation

David Brett

Résumé : Les procédés standard pour le traitement des collocations, qui impliquent l'élaboration de listes de collocations par paires, sont loin d'être optimaux pour l'étude de la connectivité, c'est-à-dire l'analyse de la probabilité que ces collocations soient co-occurentes ou non. Cet article explore une stratégie alternative qui a suscité beaucoup d'intérêt ces dernières années. Il s'agit de l'utilisation des procédés du type *Social Network Analysis*.

En utilisant des procédés standards, des listes de collocations (*concgrams*) ont été extraites à partir d'un corpus contenant un million de mots, tiré de journaux d'investigation. Le logiciel Gephi a été ensuite utilisé pour transformer ces listes de collocations dans un réseau.

On a identifié un petit nombre de paires de collocations considérées comme des cas isolés, à savoir des paires de collocations co-occurentes uniquement l'une avec l'autre. La majorité appartient au composant géant formé de paires où au moins un membre est co-occurent avec au moins un autre mot. Ont été identifiés également des modules (des regroupements de collocations ayant un degré de connectivité élevé) appartenant à des aires thématiques spécifiques. Le corpus a été ensuite réexaminé pour voir dans quel contexte ces regroupements apparaissent et à quel point cette technique aide à détecter la thématique d'un texte donné.

Mots-clés: analyse des réseaux sociaux, collocation, réseaux de collocations, langage journalistique, journalisme d'investigation.