

POMPAMO : outil de détection de candidats à la néologie

La POMPAMO est un élément d'une plateforme de veille lexicale actuellement en cours de développement. La méthode générale que nous mettons en œuvre vise à identifier les néologismes en confrontant des corpus, considérés comme des archives de pratiques linguistiques données, et des lexiques, considérés comme des simulations d'usages lexicaux attestés et correspondant à ces pratiques linguistiques.

La POMPAMO s'intègre dans une chaîne de traitement dédiée à la détection des néologies formelles comprenant quatre grandes étapes :

1. la collecte de documents textuels sur Internet à partir de mots clefs (restriction par nombre de mots, nom de domaine, etc.) (module « Pompadoc »)
2. la constitution de corpus homogènes et documentés (datation, type de textes, genre, domaine, etc.) (module « Poadoc »),
3. l'assortiment de ces corpus à des lexiques d'exclusion ad hoc et également documentés (datation, versionnage, etc.) (module « Pompamo »)
4. la structuration et la documentation des données obtenues à des fins d'analyses et de valorisation lexicographique (domaine, contextes, calcul des spécificités contextuelles en diachronie, etc.) (base de donnée relationnelle «Poamo »)

La matière première lexicographique produite sera susceptible d'alimenter des bases de données lexicales ouvertes telles que celles actuellement mises en place au CNRTL (lexiques TAL et dictionnaires informatisées) ou d'améliorer des lexiques de référence comme Morphalou (complétion, correction). Elle constituera également une ressource pour des recherches en veille lexicale (observatoire de la créativité lexicale).

Je commencerai mon exposé en vous présentant les différents modules de la plateforme de veille lexicale envisagée au sein de l'ATILF, puis, après un bref rappel théorique sur la néologie, je détaillerai le fonctionnement de l'outil POMPAMO. Prenant appui sur un exemple concret, je terminerai par une évaluation de la qualité des candidats à la néologie détectés.