

L'évolution diachronique du sens des mots est l'une des branches récemment abordée par le Traitement Automatique des Langues. Cette présentation permettra de présenter les principales approches expérimentées et les principales perspectives de travail actuelles pour améliorer les systèmes existants.

Après les premiers systèmes essentiellement basés sur le changement de fréquence (par exemple Gulordova & Baroni, 2011), des systèmes ont utilisé les représentations vectorielles par Word Embeddings (Kim et al., 2014) puis plus récemment par Contextual Embeddings (Hu et al., 2019 ; Martinc et al., 2019 ; Giulianelli et al., 2020). Ces derniers systèmes procèdent généralement en regroupant les représentations vectorielles contextuelles des différents usages en clusters de sens, puis détectent les changements selon différentes métriques (Monteïrol et al. 2021). Les systèmes actuels sont encore confrontés à de nombreuses limitations. Principalement, l'opacité des modèles neuronaux ne permet pas de caractériser ces évolutions, en particulier il est difficile, voire impossible, de relier les changements sémantiques à des caractéristiques linguistiques morphologiques, syntaxiques ou lexico-syntaxiques, ou de catégoriser les types de changements (extension, restriction, métaphore, métonymie, etc.). A cette fin, d'autres pistes peuvent être abordées, notamment en combinant les approches neuronales avec des des approches issues de la linguistique de corpus (par exemple Gries , 2012 ; Jansegers & Gries, 2017) qui permettent d'extraire les motifs lexico-syntaxiques les plus saillants d'un sens donné à partir d'un corpus d'occurrences et ainsi d'en identifier l'évolution. Une dernière piste encore inexploree consiste à exploiter les informations contextuelles des occurrences (date, type de source, domaine, traits diatopiques et diastratiques, etc.) pour caractériser et suivre l'évolution des usages.