

Analyse linguistique d'un texte en cours de rédaction

Alexis Nasr

26 octobre 2023

Ce stage s'inscrit dans le cadre général de l'étude du processus de création littéraire. L'idée consiste à identifier les configurations linguistiques dans lesquelles un auteur effectue une opération d'édition, tel que l'effacement ou le remplacement. Pour cela, nous disposons des fichiers produits par un éditeur de texte qui enregistre toutes les opérations d'édition de l'auteur.

Les modèles de TAL réalisant une analyse linguistique, tel que la segmentation en mots, l'étiquetage en parties de discours, l'analyse morphologique et l'analyse syntaxique prennent en entrée du texte finalisé. L'objectif du stage est d'adapter ces modèles afin qu'ils soient en mesure d'analyser le texte au fur et à mesure de sa création. Ainsi, lorsqu'une opération de correction est effectuée, on est en mesure de décrire précisément le contexte linguistique de cette opération.

Le travail s'appuiera d'une part sur l'éditeur schnappi, qui permet de générer des fichiers retranscrivant tout l'historique de rédaction d'un texte ainsi que sur une extension de l'analyse syntaxique en transition, qui a été développée au sein de l'équipe talep [Dary :2021] [Dary :2022], qui permet de réaliser une analyse incrémentale.

La première partie du travail portera sur la prédiction de structure syntaxique proprement dite. La seconde portera sur le regroupement de contextes de correction : étant donné un certain nombre d'opérations de correction ainsi que la description de leur contexte linguistique, comment regrouper ces derniers pour faire apparaître d'éventuels motifs réguliers. En d'autres termes, est-ce que les auteurs ont tendance à effectuer des corrections dans des contextes se ressemblant ?

Bibliographie

- Dary :2021 Franck Dary, Alexis Nasr, The reading machine : A versatile framework for studying incremental parsing strategies. In The 17th International Conference on Parsing Technologies, 2021
- Dary :2022 Franck Dary, Maxime Petit, Alexis Nasr. Dependency Parsing with Backtracking using Deep Reinforcement Learning. Transactions of the Association for Computational Linguistics, 10, 888-903, 2022