

Projet de synthèse

Portail web pour le processus de contextualisation et caractérisation des jets textuels

Contexte

Dans le cadre de ce projet, l'objectif est de développer un Framework informatique pour l'analyse des « jets textuels » dans les productions écrites lors de rédactions numériques. Un jet textuel peut se définir comme une séquence d'écriture produite entre deux pauses lors la rédaction d'un texte. En effet, lors de la rédaction, les scripteurs alternent des moments de pause, sans écriture, avec des moments de transcription continue du texte. Les pauses se produisent parce que les rédacteurs doivent réaliser des gestes moteurs complexes, parce qu'ils n'ont plus d'information pour continuer leur texte, ou parce qu'ils doivent l'évaluer ou le réviser. Au regard des deux derniers types de pauses, qualifiées de pauses cognitives, nous formulons l'hypothèse selon laquelle la segmentation par les pauses dénote des régularités syntaxiques et sémantiques dont la prévisibilité est du moins en partie calculable.

L'étude des caractéristiques des jets textuels est importante pour notre compréhension des mécanismes cognitifs en jeu lors de l'accomplissement des tâches liées à l'expression écrite mais aussi pour comprendre, d'un point de vue linguistique, la construction progressive d'un texte, c'est-à-dire le processus de textualisation.

Le projet se situe à l'articulation entre les méthodes de fouille de données, les études linguistiques d'une part et des configurations discursives d'autre part. Sur la base de données textuelles multi-vues (texte + textualisation), l'objectif sera d'utiliser des méthodes de Machine Learning (non-supervisés et supervisés) pour une détection automatique des pauses et les caractériser. Des méthodes de traitement de texte (TF-IDF, word2v,...) et de projections de données (NMF, t-SNE,...) seront utilisées dans le projet.

Les données

Les données multi-source collectées pour ce projet sont issues de l'enregistrement en temps réel du processus d'écriture de plusieurs types de production langagière : récits et argumentations produites par des élèves (entre CE et 6ème), textes produits par des étudiants de Licence dans des conditions écologiques ou expérimentales, traductions anglais-français, écrits académiques produits par des étudiants de Master, rapports éducatifs de la protection de l'enfance.

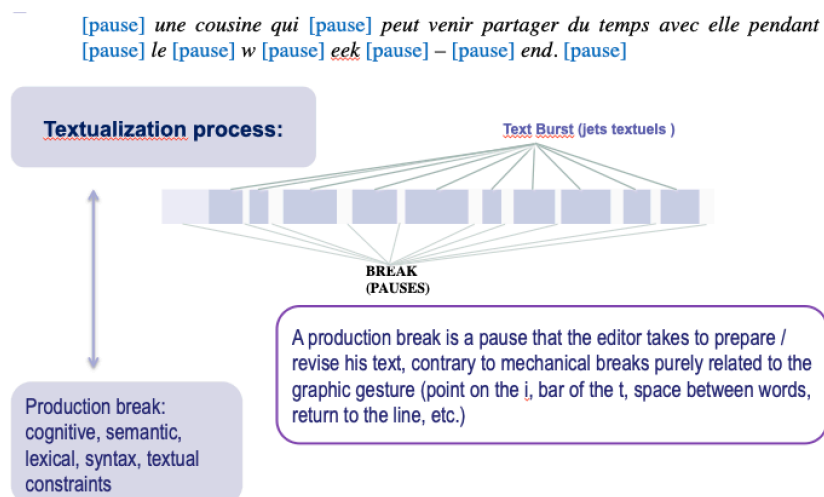


Figure 1. Exemple de jets textuels lors d'écriture d'un texte

La figure 1 représente un exemple de texte et des pauses obtenues lors de l'écriture (pauses plus longues de 2 secondes).

Le travail sera fait en langage Python avec les bibliothèques numpy et scikit-learn avec un développement d'une interface web.

Résultats attendus:

Développement d'une plate-forme web qui permettra de détecter automatiquement les pauses durant l'écriture et les expliquer.

Compétences souhaitées :

- Traitement de données
- Développement Python et /ou R
- Développement Web

Contact

Nistor GROZAVU : nistor.grozavu@cyu.fr