

Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE

Thomas François
Aspirant FNRS
CENTAL (Université catholique de Louvain)

La lecture constitue l'une des tâches clefs dans l'apprentissage d'une langue étrangère. Dès lors, le professeur de FLE est régulièrement amené à préparer une activité sur la base d'un texte adapté au niveau de sa classe. Bien que les manuels de FLE constituent une manne où puiser des textes calibrés pour un niveau donné, leur variété thématique reste limitée par comparaison avec l'Internet.

Cependant, la recherche en ligne d'un texte portant sur un sujet précis et qui soit adapté au niveau de chaque apprenant est consommatrice de temps et gagnerait à être automatisée. Des expériences (Collins-Thompson & Callan, 2005 ; Heilman et al., 2008) montrent que, pour l'anglais, l'utilisation de classifieurs statistiques permet d'estimer automatiquement la difficulté d'un texte. Nous proposons une méthodologie originale comparant, pour le français langue étrangère (FLE), diverses techniques de classification (la régression logistique, le bagging et le boosting) sur deux corpus d'entraînement et s'intéressant à la capacité prédictive de diverses variables linguistiques.

Bibliographie :

K. Collins-Thompson and J. Callan. (2005.) « Predicting reading difficulty with statistical reading models » in *Journal of the American Society for Information Science and Technology*, 56(13), 1448-1462.

Heilman, M., Collins-Thompson, K., and Eskenazi, M. (2008). « An Analysis of Statistical Models and Features for Reading Difficulty Prediction » in *Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Applications*. Columbus, Ohio, 71-79.