

Implémentation et évaluation d'indicateurs de la complexité linguistique d'un écrit en français

Encadrement : nicolas.hernandez@univ-nantes.fr

Lieu du TER : LS2N (Site de la FST)

Description

Contexte

L'apprentissage à distance (e.g. MOOC) souffre de problèmes de distanciation sociale. Les apprenants sont isolés les uns des autres et les formateurs n'ont pas les mêmes moyens qu'en présentiel pour observer les comportements et les signes de communication non-verbale utiles pour évaluation de l'apprentissage des apprenants. Pour pallier ces problèmes, les plateformes d'apprentissage cherchent à intégrer de nouveaux outils de communication et de suivi notamment pour prédire la complétion, la réussite ou le décrochage des apprenants. La plupart des études s'appuient sur l'analyse de traces d'utilisation des plateformes par les apprenants (e.g. nombre de clicks, connexions, vidéos visualisées, devoir soumis) mais très peu exploitent le contenu des forums. De plus, les travaux qui mettent en oeuvre des techniques de traitement automatique des langues (TAL) rendent comptes de résultats contrastés sur l'apport effectif du TAL pour la mesure de l'engagement des apprenants dans leur formation [1;2;3].

Sujet

L'objectif de ce travail est de poursuivre ces études sur la mesure de niveau d'engagement des apprenants dans leur formation sur la base de l'analyse des contenus de forums de MOOC en français.

A cette fin le présent sujet vise l'implémentation d'indicateurs de la complexité linguistique d'un écrit en français (e.g. complexité lexicale, syntaxique ou cohésive) et la mesure de leurs corrélations avec les performances scolaires [2;3].

A minima, l'étude de corrélations pourra être menée sur des textes générés artificiellement ou bien de genres et de registres différents (e.g. journalistique, littéraire, encyclopédique, tweets).

La littérature compte plusieurs travaux sur l'analyse la lisibilité ou de la complexité linguistique d'une production écrite en anglais [4;5;6;7;8] et quelques uns sur le français [9;10;11].

Profil des candidats ou des candidates

Deux personnes sont attendues. Chacune doit avoir une appétence confirmée pour le Traitement Automatique des Langues et l'Apprentissage Automatique. Des connaissances en Fouilles de données sont bienvenues.

Livrables

- un rapport qui prendra la forme d'un article scientifique
- ainsi qu'une mise à disposition des ressources produites (code source documenté réalisé et éventuellement données annotées) sous licence libre.

Bibliographie

- [1] Esther Félix, Exploration du contenu textuel des forums des MOOC pour analyser le niveau d'engagement des apprenants, rapport technique, LS2N Lab-STICC, 16 sept. 2021
- [2] Ed Fincham, Alexander Whitelock-Wainwright, Vitomir Kovanović, Srećko Joksimović, Jan-Paul van Staaldin, and Dragan Gašević. 2019. Counting Clicks is Not Enough: Validating a Theorized Model of Engagement in Learning Analytics. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. ACM, Tempe AZ USA, 501–510. <https://doi.org/10.1145/3303772.3303775>
- [3] Srećko Joksimović, Oleksandra Poquet, Vitomir Kovanović, Nia Dowell, Caitlin Mills, Dragan Gašević, Shane Dawson, Arthur C. Graesser, and Christopher Brooks. 2018. How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs. *Review of Educational Research* 88, 1 (Feb. 2018), 43–86. <https://doi.org/10.3102/0034654317740335> Publisher: American Educational Research Association.
- [4] S. Crossley, J. Ocumpaugh, M. J. Labrum, F. Bradfield, M. Dascalu, et R. Baker. Modeling Math Identity and Math Success through Sentiment Analysis and Linguistic Features. In *EDM*, 2018.
- [5] Kyle, K., Crossley, S. A., & Berger, C. (in press). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*. <https://www.linguisticanalysistools.org/taales.html>
- [6] Crossley, S. A., Kyle, K., & Dascalu, M. (in press). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating Semantic Similarity and Text Overlap. *Behavioral Research Methods*. <https://www.linguisticanalysistools.org/taaco.html>
- [7] Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods* 48(4), pp. 1227-1237. doi:10.3758/s13428-015-0651-7
- [8] Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral Dissertation). <https://www.linguisticanalysistools.org/taassc.html>
- [9] François, T. et Bernhard, D. (eds.) Recent Advances in Automatic Readability Assessment and Text Simplification. In *International Journal of Applied Linguistics (Special issue)*, 165:2, John Benjamins.
- [10] Projet ALLuSIF - AnaLyse de la Lisibilité et Simplification pour le Français <http://lilpa.unistra.fr/fdt/projets/anciens-projets/allusif/>
- [11] François T., Les apports du traitement automatique du langage à la lisibilité du français langue étrangère, Ph.D. Thesis, Université Catholique de Louvain. Thesis Supervisors : Cédric Fairon and Anne Catherine Simon.