

La politique des éditorialistes

Analyse NLP du récit politique dans les médias français

Projet Statapp

* *

*

Salomé Do (Sociologie et NLP, ENS / Sciences Po) et
Étienne Ollion (Sociologie, CNRS/ENSAE)

À la croisée de la sociologie politique et du traitement automatique du langage, ce projet prend pour objet les éditoriaux prononcés entre 2015 et 2019 par des journalistes français à la radio. Si la manière dont les journalistes effectuent leur travail fait l'objet de travaux contemporains bien détaillés, les éditorialistes devenus omniprésents sont moins étudiés.

L'objectif est de proposer une manière d'aborder les questions ci-dessus avec des méthodes récentes issues du traitement automatique des langues.

Données

Un jeu de données de 2200 éditoriaux retranscrits de radios nationales (Europe 1, RTL, France Inter, France Culture, Sud Radio) sera fourni aux étudiant.e.s.

Questions

Comment les éditorialistes décrivent-ils l'actualité politique ? Quels faits évoquent-ils, et quels autres occultent-ils ? Quels sont les acteurs qui peuplent leurs récits ? Quels mots, quelle tonalité utilisent-ils ? Prennent-ils explicitement position, et si oui selon quelle logique argumentative ?

L'objectif de ce projet, qui s'inscrit dans une recherche en cours menée par les encadrants, consiste à trouver des manières d'aborder ces problèmes et de les approfondir.

Méthodes

Plusieurs méthodes issues du TAL pourront être mobilisées.

- Apprentissage supervisé pour entraîner un algorithme à reconnaître certaines formes de texte
- *Name entity recognition* pour reconnaître les personnages cités
- *Sentence embedding* pour capturer les thématiques d'un texte

Bibliographie sélective

- Aalberg, T., Strömbäck, J., et de Vreese, C. H. (2012). « The framing of politics as strategy and game : A review of concepts, operationalizations and key findings ». In *Journalism*, 13(2) :162–178.
- Salomé Do, Etienne Ollion, « Learning to Analyze, Learning to Produce », *Presentation at the AI & Social Sciences Workshop*, Institute for Advanced Studies, Mars 2020.
- Arthur Boriello, « 'There is no alternative': How Italian and Spanish leaders' discourse obscured the political nature of austerity », in *Discourse and Society*, 2017.

- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. St., ... Kurzweil, R. (2018). Universal Sentence Encoder. <http://arxiv.org/abs/1803.11175>
- Arora, S., Liang, Y., & Ma, T. (2016). A simple but though Baseline for Sentence Embeddings. *Iclr*, 15, 416–424.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, January). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*.
- Wang, B., & Kuo, C.-C. J. (2020). SBERT-WK: A Sentence Embedding Method by Dissecting BERT-based Word Models, 14(8), 1–13.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural Architectures for Named Entity Recognition, 260–270. <https://doi.org/10.18653/v1/N16-1030>
- Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2019). Topic Modeling in Embedding Spaces.