

Observatoire du discours parlementaire :

Développer de nouveaux outils de TAL
et les mettre à la disposition du grand public

Lucas Piessat*, Julien Velcin**, Gauthier Bravais*

*Skoli **Université de Lyon (ERIC, Lyon 2)

Action MADONA, Symposium MaDICS, 11-12 juillet 2022

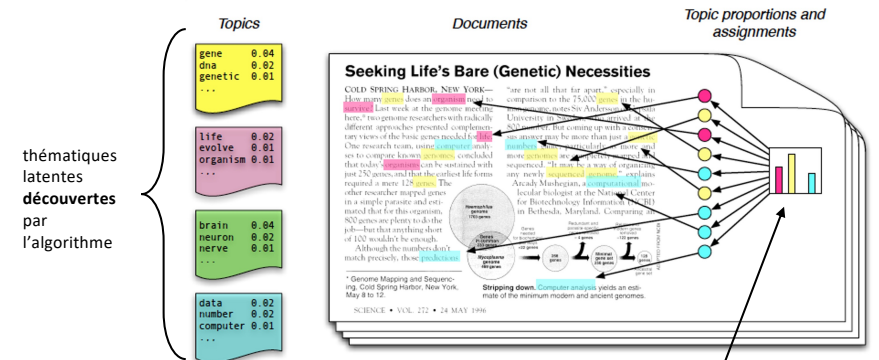
On constate un vide

- Accès à des masses d'information importantes sur la vie publique
- Pas (peu) d'outil disponible pour traiter efficacement (quantitativement) ces masses d'information par les analystes
- Or les recherches avancent très vite en TAL (ex. de la traduction auto, des chatbots...)
- L'information parlementaire est résumée à un indice de présence, infantilisant et qui ne prend pas en compte la diversité de l'activité parlementaire.
- La régularité des débats parlementaires et leur densité en empêchent un suivi pertinent. La dé-parlementarisation de la vie publique suite à la réforme du quinquennat a largement contribué à cette tendance que les législatures de 2022 battent en brèche.

Outils de TAL pour le « distant reading »

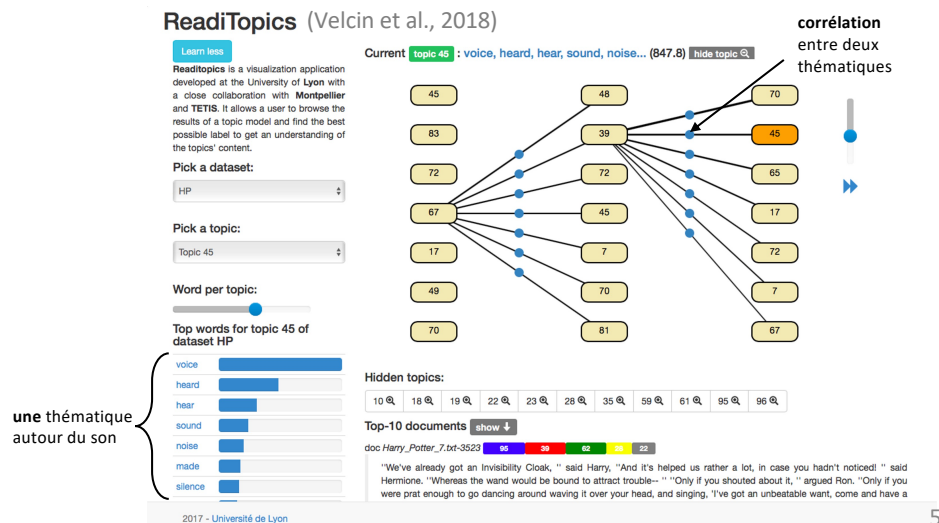
- Approches dites « non supervisées » (*clustering, topic modeling*)
- Avantages :
 - capable de traiter d'importantes quantités de données, sans à priori
 - construit une forme de résumé basé sur les co-occurrences de mots
 - permet ensuite de réaliser des comparaisons selon le vocabulaire employé (voir par exemple Velcin et al., 2017)
- Mais :
 - pas évident à déployer de manière efficace
 - problèmes d'interprétabilité des « topics »
 - instabilités possibles, suivant la nature et le volume des données
 - des attentes différentes en fonction de l'utilisation / utilisateur

Topic modeling



Modèle LDA (Blei et al., 2003)

Un texte est un mélange de thématiques



5

Quelques applications possibles pour l'analyse du discours parlementaire

- Analyse au niveau d'un député
- Analyse au niveau d'un groupe
- Analyse temporelle
- D'autres types d'analyses sont possibles :
 - corrélation entre les thématiques (statique, temporel)
 - écart entre le vocabulaire employé par un député et le vocabulaire général
 - analyse temporelle fine : qui « lance telle ou telle thématique »
 - structurer une thématique en sous-thématiques (facettes)
- etc.

6

Pistes de recherche et applicatives

- Rendre réellement compréhensibles les thématiques et les structurer de manière intelligibles (perspective XAI), par exemple sur la base de « signatures » caractéristiques (Gautrais et al., 2017)
- Développer une bibliothèque permettant de déployer ce type de solution, un cran plus loin que Readitopics (Velcin et al., 2018)
- Analyse automatique [/étude] de l'argumentaire parlementaire, en particulier l'évolution des éléments de langage, de l'agenda politique, du renouvellement des catégories politiques qui polarise l'espace public (écologie, nationalisme, ...). Le renouvellement de l'AN avec une majorité relative est un temps politique particulièrement intéressant pour comprendre ces dynamiques. (Braud, 2010)

7

Vers une forme d'engagement

Stratégie

Engagement comme contrat/promesse : La démocratie postule l'information et l'association des citoyens à la gestion des affaires communes. Aujourd'hui la densité de l'information rend impossible cette auto-saisine par le peuple de ses intérêts. En "contractant en stock le flux d'informations" on rend l'activité parlementaire préhensible.

Cadrage

Engagement comme enrôlement/activisme : Nous visons à travers une meilleure compréhension des lignes de fractures idéologiques qui traversent la représentation nationale à repolitiser la notion de "représentant du peuple". Accompagner la "re-parlementarisation" de la vie politique en rendant l'activité parlementaire et ses fractures plus lisibles aux citoyens, journalistes, activistes...

Interface

Engagement comme attention portée / considération : Au niveau de la restitution, la médiation et la publicité opérées par un partenaire média apparaissent indépassables pour optimiser la réception des analyses. D'un point de vue de l'interface, si le travail d'UI/UX n'est pas encore réalisé, le caractère "explorable" (Bret, 2011) apparaît décisif pour activer le lecteur.

8

Références

- (Blei et al., 2003) Blei, D. M., Ng, A. Y., & Jordan, M. I.: Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022, 2003.
- (Davidson et al., 2020) Davidson I., A. Gourru, J. Velcin and Y. Wu: Behavioral differences: insights, explanations and comparisons of French and US Twitter usage during elections. *Social Network Analysis and Mining (SNAM)*, 10: 6.
- (Gautrais et al., 2017) Gautrais C., Cellier P., Quiniou R., Termier A.: Topic Signatures in Political Campaign Speeches. *EMNLP 2017*: 2342-2347
- (Gourru et al., 2018) Gourru, A., Velcin, J., Roche, M., Gravier, C., & Poncelet, P.: United we stand: Using multiple strategies for topic labeling. In *International Conference on Applications of Natural Language to Information Systems (NLDB)*, pp. 352-363, 2018, Springer.
- (Velcin et al. 2017) Velcin J., J.C. Soulages, S. Kurpiel, D.L. Otavio, M. Del Vecchio et F. Aubrun: Fouille de textes pour une analyse comparée de l'information diffusée par les médias en ligne : une étude sur trois éditions du Huffington Post. *Atelier Journalisme computationnel @EGC*, 2017.
- (Velcin et al., 2018) Velcin, J., Gourru, A., Giry-Fouquet, E., Gravier, C., Roche, M., & Poncelet, P.: Readitopics: make your topic models readable via labeling and browsing. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018, demo track)*, pp.5874-5876, Stockholm, Sweden.