

dans le texte d’appel. Chaque pétition est classée dans une des neuf catégories proposées sur le site lapetition.be : Art et culture ; Droits de l’Homme ; Environnement, nature et écologie ; Humour/Insolite ; Loisirs ; Politique ; Protection animale ; Social ; Autres.

Dans un premier temps, nous avons mené une série d’investigations afin de comparer les pétitions en fonction des catégories, de la longueur du texte d’appels, du nombre de signataires, etc.

La figure 1 rend compte de variations importantes en termes de nombre de pétitions selon la catégorie retenue par le porteur de pétition. La catégorie “Autres” rassemble plus de 2400 pétitions sur la période étudiée et la catégorie “Humour/Insolite” rassemble moins de 400 pétitions. Le choix de la catégorie relève du promoteur de la pétition et témoigne d’une interprétation qui varie selon les porteurs de projet, la rubrique “Droits de l’homme” comporte une variété de textes d’appel qui souligne la difficulté d’appréhension de cette catégorie, ou bien son détournement par des porteurs de projet.

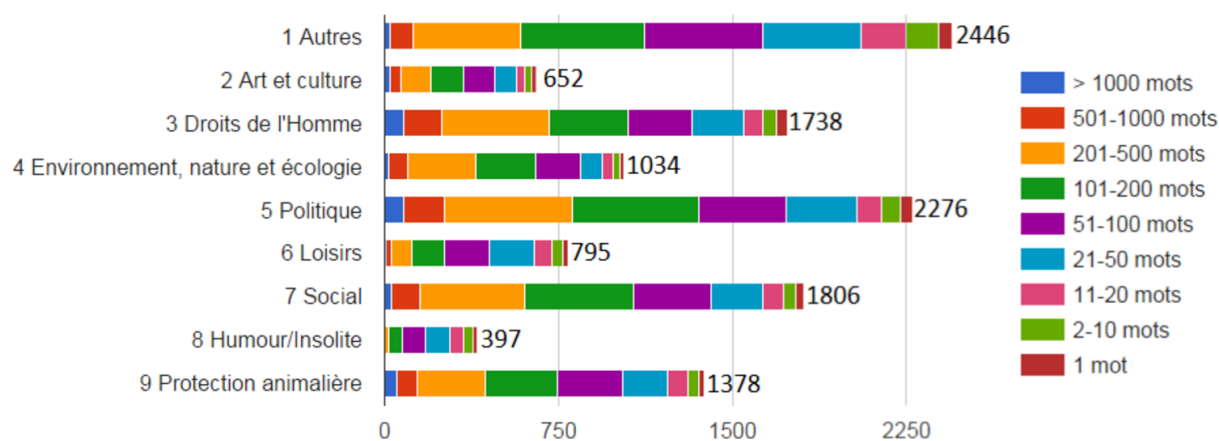


Fig. 1 - Distribution du nombre de pétitions par catégorie de pétition, réparties par nombre de mots¹

La variété des interprétations possibles des catégories proposées, les usages distancés ou non de ces catégories, leur possible détournement, de même que les éventuels mésusages par rapport aux présupposés et choix des concepteurs contribuent à expliquer que la catégorie “Autres” rassemble le plus grand nombre de pétitions. C’est également un facteur explicatif quant au nombre de mots (estimé ici de manière automatique à l’aide de la commande `wc -w` d’Unix) du texte d’appel qui varie de 1 à 8322, soulignant le caractère très hétérogène de cette rubrique. La catégorie “Politique” est la deuxième rubrique en termes de nombre de pétitions, elle se caractérise par le fait que plus de 60 % des textes d’appel de cette catégorie comportent plus de 100 mots. L’hétérogénéité du corpus en termes de nombre de mots exige cependant de la prudence quant à son interprétation, notamment dans le cas d’un nombre très restreint de mots. En effet, il semble que les pétitions qui ne comportent qu’un mot s’apparentent à des tests de la part des porteurs de projet, avec une seule signature (“+”, “ERREUR”, “test”, “f”, “rien”, “RAS”), à des URL (“<http://www.blog4ever.com/fichiers/94004070306065618.pdf>”) ou à des suppressions de pétition (“x” et “...” sont des textes et des titres de pétitions ayant recueilli 35 signatures). Il est en effet important de noter que les porteurs de projet ont la possibilité de changer le texte et le titre de la pétition a posteriori.

¹ Les données quantitatives ayant permis générer cette figure, comme les suivantes, sont disponibles sous forme de matériel supplémentaire de cet article, à l’adresse <http://jadt2016.treecloud.org>.

On remarque des répartitions similaires dans le plan, avec le long de l’axe horizontal, la catégorie “Protection animalière”, à droite, séparée des autres à gauche par les catégories “Loisirs” et “Humour/Insolite” en position intermédiaire. Les termes proches de cette catégorie “Protection animalière”, qui apparaissent lors de l’affichage des points-lignes dans TXM, sont “merci”, “faut”, “signer” et “association” pour les débuts de textes, “aidez”, “stop”, “svp” ou encore “facebook” pour les fins de textes.

En ce qui concerne l’axe vertical, la répartition semble refléter une distinction entre des thématiques vers le haut qui font appel à un lexique que nous pourrions qualifier “du quotidien” (“bonjour”), inscrit dans une forme familière d’échanges, de l’ordre de l’interpersonnel (catégories “Humour/Insolite”, “Loisirs”, “Autres” et “Art et culture”), avec les mots “tokio”, “hotel”, “marre”, “cher” et “signer” pour les débuts de textes et “moi”, “petition” et “belge” pour les fins de textes, et des thématiques vers le bas qui font appel à un registre lexical relevant du champ politique avec la présence de termes comme “projet”, “gouvernement”, “association”, “publique” et “ministre” pour les débuts de textes et “org”, “citoyens”, “loi”, “ministre” et “expression” pour les fins de textes.

4.2. Quelques thématiques des débuts et fins de pétitions

La construction de nuages arborés des débuts de textes (figure 5), qui regroupent les mots qui apparaissent fréquemment à proximité dans les débuts de textes, tout en faisant ressortir les plus fréquents par une taille de police plus importante et des couleurs vives (Gambette et Véronis, 2010), met en évidence la manière dont les auteurs de ces textes choisissent de s’adresser à leurs interlocuteurs.

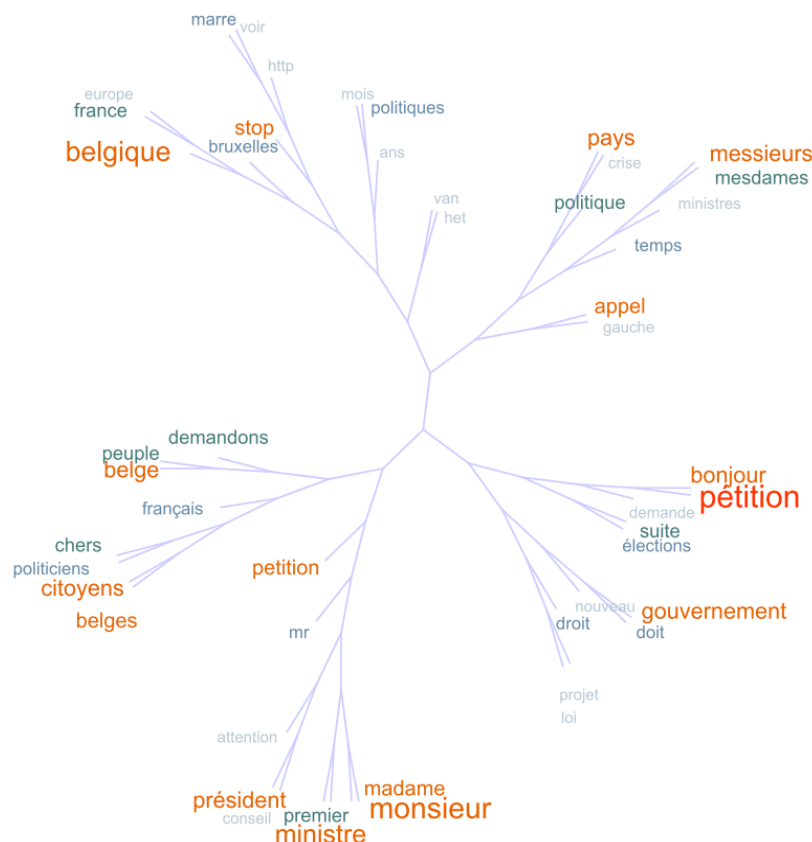


Fig. 5 – Nuage arboré des 10 premiers mots des textes de pétitions de la catégorie “Politique”.

On retrouve ici un paradoxe classique des pétitions (Contamin, 2001), qui s’adressent à un décideur (“ministre”, “ministres”, “président”, “madame”, “monsieur”, “politiciens”), mais constituent également un message directement adressé aux signataires (“chers”, “citoyens”, “peuple”). Plus qu’une pétition, il s’agit alors d’un véritable “appel”, qui peut être introduit par des mots comme “stop” ou “marre”.

Il faut toutefois relativiser la représentativité du nuage arboré de la figure 5. En effet, les deux termes les plus fréquents, “pétition” et “monsieur”, n’apparaissent que 138 et 108 fois respectivement dans ces débuts de 2276 pétitions. Une analyse plus fine avec retour au texte montre que 6,5 % des pétitions de la catégorie “Politique” mentionnent des décideurs politiques parmi les 10 premiers mots, et que 3,8% font référence aux signataires. Une grande part des pétitions politiques commence directement par une présentation de la problématique, qui fait appel à un vocabulaire spécifique, qui est donc essentiellement (sauf pour les mots “crise” et “élections”) absent des termes les plus fréquents.



Fig. 6 – Nuage arboré des 10 derniers mots des textes de pétitions de la catégorie “Politique”.

En ce qui concerne les fins de pétitions représentées dans le nuage arboré de la figure 6, on remarque de nouveau la présence des divers interlocuteurs des pétitionnaires, éventuellement à proximité de formules de politesse (“expression” et “salutations” proches de “ministre”). En plus de formules attendues d’incitation à signer ou de remerciement, on remarque tout particulièrement des composants d’URL (“http”, “www”, “be”, “fr”, “com”). Cela traduit les relations entre ces pétitions et d’autres relais de mobilisation politique. Au total, ce sont 4,1 % des textes des pétitions qui contiennent “http” ou “www” à la fin, contre 1,1 % au début. Au total, 11,5 % des textes de pétitions contiennent “http” ou “www”. On peut remarquer que les pratiques diffèrent là encore selon les catégories de pétition, puisque seules 7,3 % des pétitions de la catégorie “Loisirs” contiennent “http” ou “www”, et 17,5 % des pétitions de la catégorie “Protection animale”. La mobilisation en ligne par divers canaux dans les réseaux

dédiés à la protection animalière se confirme avec une sur-représentation statistiquement significative du mot “facebook” dans cette catégorie.

4.3. Le nombre de signatures lié au vocabulaire des débuts et fins de textes de pétition

Nous avons également construit à l’aide du logiciel TextObserver (Leblanc, 2010) les analyses factorielles des débuts et des fins (figure 7) de textes de pétitions de la catégorie “Politique” regroupés en fonction du nombre de signatures, dans les intervalles suivants : [1-49] (intervalle noté “1”), [50-199] (“2”), [200-499] (“3”) et [500-] (“4”). On note que dans les deux cas, l’axe horizontal permet de distinguer ces quatre catégories d’intervalles de signatures, de la gauche vers la droite. Associés aux catégories 1 et 2 des pétitions les moins signées, à gauche des AFC, on retrouve parmi les mots qui contribuent le plus à l’axe horizontal les pronoms personnels (“je”, “on”, et “vous” dans les débuts de textes, “on” et “nous” dans les fins de texte), ainsi qu’un vocabulaire familier (“[ras le] bol”, “marre”). Quant aux catégories 3 et 4 des pétitions les plus signées, à droite des AFC, elles sont associées à des composants d’URL (“facebook”, “over”, “blog”, “be”, “fr”, “http”), à des décideurs politiques (“sénat”, “présidente”, “bourgmestre”). On remarque également que des mentions de dates précises, qui font généralement référence à des événements factuels qui soutiennent l’argumentaire, apparaissent également à droite des AFC.



Fig. 7 – AFC des 10 mots de début (à gauche) et des 10 mots de fins (à droite) des textes de pétitions, regroupés en fonction de leur nombre de signatures (1 pour 1 à 49 signatures, 2 pour 50 à 199 signatures, 3 pour 200 à 499 signatures, 4 pour 500 signatures et plus)

5. Conclusion

Cet article a proposé quelques premières analyses du corpus de pétitions du site Lapetition.be, dont on a mis en évidence la complexité et l’hétérogénéité, tout en montrant que les débuts et fins de textes contiennent des signaux relatifs aux catégories thématiques et au nombre de signatures des pétitions. Ces signaux font ressortir un nombre de signature globalement plus

important pour des pétitions qui cibleraient précisément certaines problématiques tout en précisant les destinataires de la pétition, et en fournissant une URL de référence ou de mobilisation pour les signataires. Les données dont nous disposons sont riches et variées, et elles peuvent être regroupées en sous-corpus et exploitées, tant du point de vue linguistique que dans une perspective pluridisciplinaire. Les pistes d'exploitation sont donc très nombreuses, et conduiront à explorer plus en profondeur ces pétitions en ligne, pour caractériser plus finement le genre "pétition".

Remerciements

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche (ANR-14-CE29-0010).

Références

- Barats C., Leblanc J.-M. et Fiala P. (2013). Approches textométriques du web : corpus et outils. In Barats C. (dir.), *Manuel d'analyse du Web en sciences humaines et sociales*. Paris, Armand Colin.
- Barbe L., Delcroix E. (2009). Émergence et appropriation des dispositifs sociotechniques : le cas de Facebook. *Sciences de la Société*, vol.75:115-125.
- Baude O. et al. (2006). *Corpus oraux, guide des bonnes pratiques*. CNRS Éditions, Paris.
- Baude O. (2007). Aspects juridiques et éthiques de la conservation et de la diffusion des corpus oraux. *Revue française de linguistique appliquée*, vol.XII:85-97.
- Boure R. et Bousquet F. (2010). Enjeux, jeux et usages d'une pétition politique en ligne. La pétition Vauzelle. *Réseaux*, vol.6(164):127-159.
- Boure R. et Bousquet F. (2011). La construction polyphonique des pétitions en ligne. Le cas des appels contre le débat sur l'identité nationale. *Questions de Communication*, vol.20:293-316.
- Briassoulis H. (2010). Online petitions: new tools of secondary analysis ? *Qualitative Research*, vol.10(6):715-727.
- Contamin J.-G. (2001). *Contribution à une sociologie des usages pluriels des formes de mobilisation : l'exemple de la pétition en France*, thèse de doctorat de l'Université Paris 1.
- Évrard, É. (1990). L'étude des longueurs de phrases: un réexamen des méthodes. *Revue Informatique et Statistique dans les Sciences humaines*, vol.26:55-66.
- Gambette P. et Véronis J. (2010). Visualising a Text with a Tree Cloud. In Locarek-Junge H. et Weihs C., *Proc. of IFCS'09, Studies in Classification, Data Analysis, and Knowledge Organization 40*, pages 561-570.
- Heiden S., Magué J.-P. et Pincemin B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie-conception et développement. In *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Rome, pages 1021-1032.
- Kobus C., Yvon F. et Damnati G. (2008). Transcrire les SMS comme on reconnaît la parole. In *Actes de TALN 2008*, Avignon, pages 123-138.
- Latzko-Toth G. et Proulx S. (2013). Enjeux éthiques de la recherche sur le web. In Barats C. (dir.), *Manuel d'analyse du Web en sciences humaines et sociales*. Paris, Armand Colin.
- Leblanc, J.-M. (2010). Nouvelles fonctionnalités pour la visualisation des données textuelles et des résultats : Pour une approche ergonomique des dispositifs lexicométriques. In *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Rome, pages 1057-1067.
- Tournier M. (1973). Le vocabulaire des pétitions ouvrières de 1848 : étude des parentages statistiques. In Robin R., *Histoire et linguistique*, Paris, Armand Colin.
- Viennay S. et Melian C. (2004). La correction automatique du langage des nouvelles formes de communication écrite. *Bulag*, vol.29:183-196.