

Analyser des pétitions en ligne : potentialités et limites d'un dispositif d'étude pluridisciplinaire

Christine Barats¹, Anne Dister², Philippe Gambette³,
Jean-Marc Leblanc¹, Marie Peres-Leblanc¹

¹Université Paris-Est, CEDITEC (EA 3119), Créteil – France

²Université Saint-Louis - Bruxelles – Belgique

³Université Paris-Est, LIGM (UMR8049), Champs-sur-Marne – France

Abstract

Online petitions are a worthy object of research, from the perspective of discursive materiality or of their relationship with the political process. However, this object happens to be difficult to grasp, considering how diverse data and variables may be. Analyzing an online petitioning website therefore outlines the heterogeneity of topics, titles or texts of petitions, as well as number signatures or comments. To investigate this tool, the researcher must make choices and use several tools of statistical textual analysis in order to uncover what characterizes petitions, while taking their diversity into account.

Key words : online petitioning, statistical analysis, data heterogeneity, genre.

Résumé

Les pétitions en ligne constituent un objet d'investigation riche pour le chercheur, que ce soit du point de vue de la matérialité discursive ou du rapport au politique. Cet objet s'avère cependant difficile à saisir, compte tenu de la variété des données et des variables. L'analyse d'un site de pétitions en ligne souligne ainsi l'hétérogénéité des rubriques, des titres et textes d'appel des pétitions, du nombre de signataires ou de commentaires. Pour examiner ce dispositif, le chercheur se doit de faire des choix et d'avoir recours à différents outils de statistique textuelle afin de mettre au jour ce qui caractérise les pétitions, tout en rendant compte de leur variété.

Mots clés : pétition en ligne, analyse statistique, formes d'adresse, hétérogénéité des données, genre.

1. Introduction

Le pétitionnement est l'une des pratiques les plus anciennes et communes des démocraties libérales. Néanmoins, force est de constater que, pendant longtemps, l'aspect linguistique de la pratique n'a pratiquement pas suscité l'intérêt des chercheurs, mis à part Tournier (1973) qui a analysé des pétitions ouvrières de 1848. Sans doute cela tient-il pour une part à la difficulté de l'accès aux données, qui est maintenant palliée par les pétitions en ligne (Boure et Bousquet, 2010 et 2011 ; Briassoulis, 2010). En effet, depuis plusieurs années maintenant, les pétitions en ligne se multiplient, et le nombre de signataires s'accroît. Si l'accès aux données est plus aisé que par le passé, cela ne se fait pas sans poser une série de questions méthodologiques, éthiques et juridiques. Nous appréhendons l'objet « pétition en ligne »

comme un dispositif socio-technique, c'est-à-dire relevant à la fois d'une pratique ancienne et dans le même temps actualisé par des dispositifs en ligne (Barbe, Delcroix, 2009 ; Barats et al., 2013). Dans le cadre d'un projet soutenu par l'ANR, un groupe de chercheurs a construit un dispositif méthodologique inédit fondé sur une collaboration entre des sociologues, des politistes, des juristes et des spécialistes du langage pour analyser un site de pétitions en ligne. Ce dispositif s'appuie sur l'accès à la base de données de l'un des principaux sites francophones de pétitions en ligne (lapetition.be), soit à des données concernant plus de 12.000 pétitions ayant récolté au total 3,25 millions de signatures sur la période comprise entre le 31 octobre 2006 et le 12 février 2015. Nous mettrons l'accent dans notre communication sur les difficultés de préparation des corpus compte tenu de la variété des données (titre de la pétition, texte de l'appel à signer la pétition, commentaires) et des variables (rubrique choisie par le porteur de la pétition, nombre de signataires, nombre de commentaires, période de signature, résidence du signataire, etc.). Nos hypothèses portent sur la présence d'éléments caractéristiques des pétitions en ligne et de l'écriture web (smileys, ponctuations que nous considérons ici comme des données alphanumériques, etc.). Peut-on parler de "genre" ? Quels en seraient les marqueurs ? Si l'accès à une base de plus de 3 millions de signatures constitue un potentiel sans précédent pour ce type de problématique, le chercheur est cependant confronté à de multiples difficultés : anonymisation, données manquantes, masses de données à traiter, etc. Après avoir mis l'accent sur les modalités de constitution et de préparation des corpus, liées à la collecte de données, nous présenterons quelques résultats statistiques liés à l'analyse des textes d'appel des pétitions, et en particulier aux formes d'adresse. Nous présenterons également certains aspects de l'analyse des commentaires postés.

2. Présentation du corpus

2.1. Un corpus complexe aux nombreuses variables

L'analyse des pétitions en ligne donne accès à de très nombreuses données et variables. Dans le cas des données, il s'agit de données alphanumériques liées au titre et au texte de l'appel à signer, ainsi qu'aux éventuels commentaires des signataires. Dans le cas des variables, il s'agit des informations sur la date et l'heure de signature, la localisation géographique du signataire, sa profession, etc. L'exigence d'anonymisation des données (Latzko-Thot et Proulx, 2016), comme elle se pose également dans d'autres types de recherche en linguistique telles que celles sur l'oral (Baude et al., 2006 ; Baude 2007), rend ainsi particulièrement difficiles la collecte et la préparation des données. En effet, outre la présence des variables citées, nombre de commentaires permettent d'en identifier le signataire. Le chercheur se doit ainsi de faire des choix quant aux données recueillies et analysées. Dans un premier temps, nous avons choisi de distinguer les textes d'appel de leurs titres, ainsi que des commentaires y afférents afin de tester nos hypothèses sur la présence de marqueurs spécifiques au genre "pétition" et recourir à des outils de statistique textuelle. Nous avons dans un premier temps choisi de procéder à une normalisation minimale des données, afin de tenir compte de l'usage de majuscules, indiquant une forme d'emphase ou d'interjections. Dans le cas des commentaires, nous avons souhaité conserver les textes bruts.

2.1. Catégories, textes d'appels et nombre de signataires

Le corpus global des pétitions retenues pour l'analyse, toutes catégories confondues, comprend plus de 12 522 pétitions, chacune signée au moins une fois (certaines pétitions n'ont pas de signataire, même pas celle de leur promoteur) et comprenant au moins un mot

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.

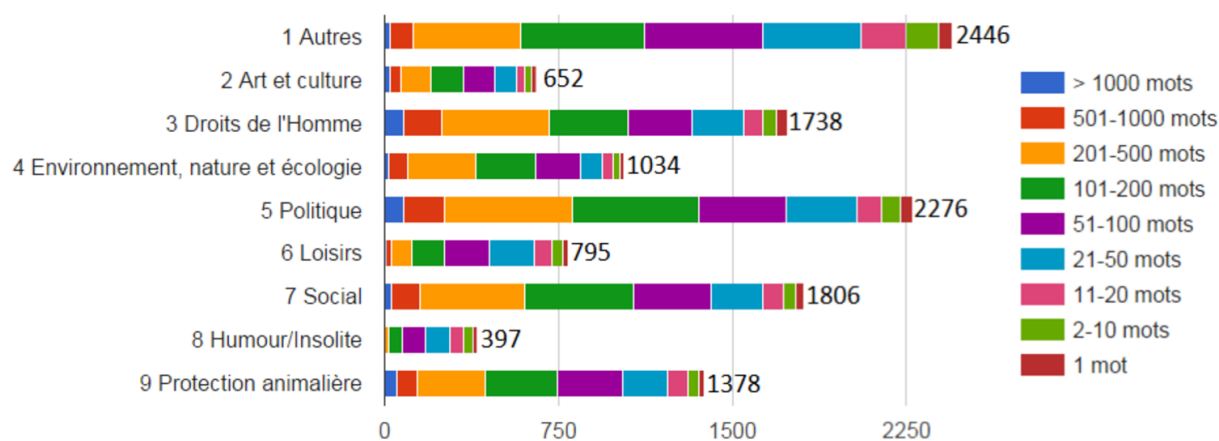


Fig. 1 - Distribution du nombre de pétitions par catégorie de pétition, réparties par nombre de mots¹

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>.

La normalisation effectuée en considérant, pour chaque catégorie, le pourcentage de pétitions dont le nombre de mots appartient à un intervalle donné (figure 2), permet de constater que les pétitions des catégories “Humour/Insolite” et “Loisirs” ont une distribution de nombre de mots similaires, avec le plus petit nombre moyen de mots (respectivement 64 et 117 mots). Les catégories “Autres”, “Art et culture” et “Protection animale” apparaissent ensuite, avec environ 60 % des textes d’appel ayant entre 51 et 500 mots. Enfin, environ 60 % des textes d’appel des catégories restantes (“Droits de l’Homme”, “Environnement, nature et écologie”, “Politique”, “Social”) ont plus de 100 mots, les textes les plus longs figurant dans la catégorie “Droits de l’Homme” (283 mots en moyenne pour cette catégorie). Cette hétérogénéité en termes de nombre de mots constitue un indice de la variété des textes de pétition.

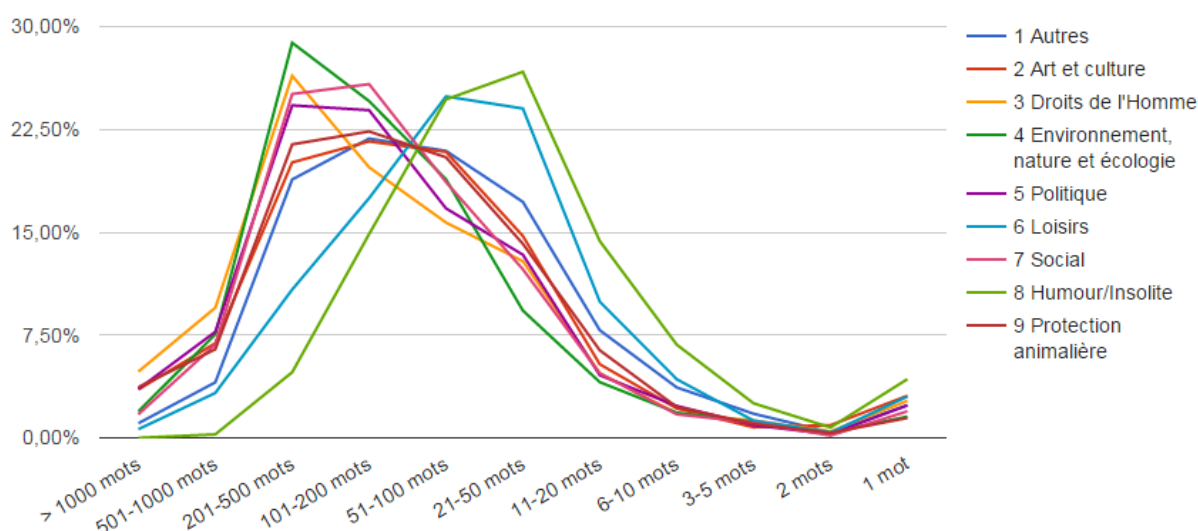


Fig. 2 - Pourcentage, au sein de chaque catégorie, de pétitions par nombre de mots

Nous avons testé s’il y avait une relation entre la longueur des textes des pétitions et leur nombre de signatures, mais nous n’obtenons pas de lien statistiquement significatif (voir figure 3).

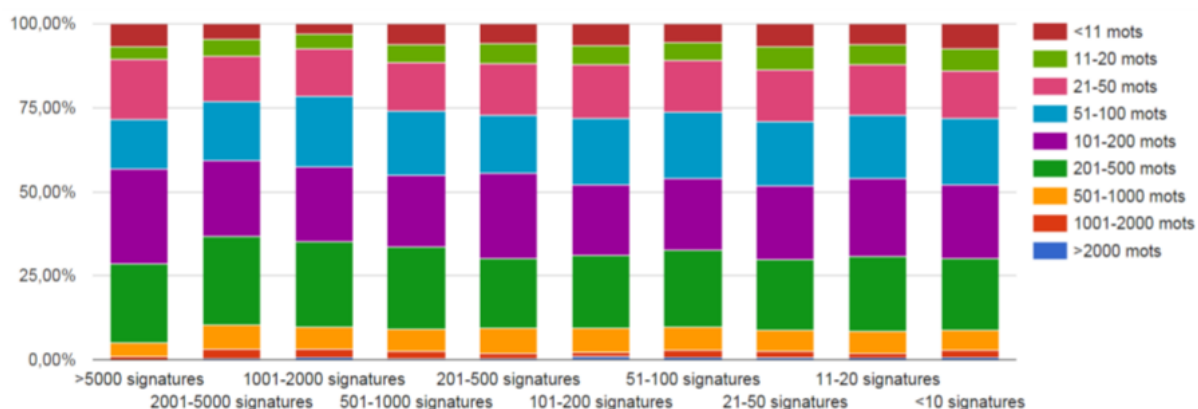


Fig. 3 - Pourcentage de pétitions dans chaque intervalle de nombres de mots en fonction du nombre de signatures de la pétition

3. Hétérogénéité des données

Comme nous l’avons déjà mentionné ci-dessus, les données collectées présentent une forte hétérogénéité et variabilité, que ce soit :

ANALYSER DES PÉTITIONS EN LIGNE : POTENTIALITÉS ET LIMITES D'UN DISPOSITIF...

- dans le choix de la rubrique par le porteur de pétition, comme tend à l'indiquer le nombre très important de pétitions dans la rubrique "Autre" ;
- dans la formulation du titre et du texte de l'appel, dont la longueur et les formes varient ;
- dans l'identification du destinataire explicite de l'appel (mention "s'adresse à") : la plupart du temps, le promoteur de la pétition indique le nom de la personne qui recevra la pétition in fine ("Gouvernement Bruxellois", "ministre de l'immigration"), mais de nombreux porteurs de projets indiquent dans cette catégorie le public-cible susceptible de signer la pétition. On a ainsi "Schaerbeekoï(s)" pour une pétition contre une taxe dans la commune de Schaerbeek, pétition en fait destinée aux autorités communales, "tout public" pour une pétition contre la fermeture d'un camping à la côte belge ou encore "victimes décedées" (nous conservons les graphies originales dans toutes les citations) dont il paraît évident que la pétition ne va pas leur être transmise ;
- dans les champs sélectionnés par le porteur pour constituer le formulaire pour le signataire. Le porteur de pétition choisit en effet les entrées du formulaire de sa pétition (avec ou sans l'indication du pays, de la profession, la possibilité de laisser un commentaire, etc.). Notons également que le formulaire a évolué dans le temps, ce qui n'est pas sans conséquence pour l'interprétation des résultats. Cela rend notamment difficile une approche diachronique visant à contraster les commentaires ;
- dans les commentaires : les signataires qui laissent des commentaires le font par rapport au texte de l'appel. Mais ils ont la possibilité de lire les commentaires déjà présents sur le site, et de ce fait, certains commentaires y répondent, validant le contenu précédent ou s'y opposant parfois. Ces commentaires ne doivent donc pas être analysés de manière isolée, mais comme entrant dans un champ discursif plus large. Par ailleurs, il arrive que certains signataires s'inscrivent en faux par rapport à la pétition qu'ils signent. On peut citer cet exemple dans une pétition en faveur du rattachement de la Wallonie à la France ("rattachement a la France", catégorie "politique") :

"Vous faite tous chier maintenant!!!Francophone ou neerlandophone , SI tu n'aime pas ton pays part ! Et toi ,SI TU AIME LA FRANCE , VA Y HABITER !
MAIS MOI , TU NE M'Y EMENERA PAS !!!
Belge et fier de l'etre !".

Le signataire est loin, ici, d'adhérer au rattachement à la France prôné par le texte de l'appel. Ainsi, tous les commentaires ne doivent pas d'emblée être analysés comme autant de données argumentées en faveur du texte de la pétition ;

- dans la maîtrise de la langue des signataires : comme l'illustre bien l'exemple de commentaire ci-dessus, les normes orthographiques standards sont loin d'être respectées. De plus, nombreux sont les commentaires agrémentés de multiples signes de ponctuation ; le gras et les mots en lettres capitales foisonnent, ainsi que les smileys ou autres abréviations. Tous ces caractères posent évidemment des difficultés spécifiques au traitement des données par les outils utilisés en textométrie, difficultés que l'on rencontre aussi, par exemple, lorsqu'il s'agit d'analyser des SMS (Viennay et Melian, 2004 ; Kobus et al., 2008). Pour l'analyse, ces marques typographiques prennent tout leur sens et ne doivent pas disparaître dans une normalisation des

données. Par manque de place, nous ne pouvons ici détailler les expériences que nous menons actuellement sur un sous-corpus, et qui consistent à comparer les résultats obtenus avec le logiciel Alceste sur deux types de données : les données brutes et les mêmes données soumises à une normalisation purement orthographique (à paraître) ;

- dans la langue utilisée, tant dans les titres, les textes d’appel que les commentaires des pétitions : ainsi, beaucoup de pétitions concernant la Belgique, il n’est pas rare de voir apparaître des textes d’appel rédigés en français et en néerlandais, des commentaires totalement ou partiellement en néerlandais (“trop is te veel”, “TROP DE BRUIT IS TE VEEL !!!”) ou en anglais, mais aussi dans d’autres langues, voire même du latin (“usque tandem Wathelet abutere patientiae nostra?”, dans la pétition contre le nouveau plan de survol de Bruxelles).

4. Vers la caractérisation d’un genre “pétition”

Nous présentons maintenant quelques résultats statistiques liés à l’analyse des textes d’appel des pétitions. Nous explorons l’hypothèse selon laquelle les débuts et fins de textes d’appel des pétitions fournissent quelques éléments caractéristiques des pétitions. Certains de ces éléments sont susceptibles de conduire à une description du genre de la pétition électronique (formes d’adresses, expression d’une demande, remerciements, etc.), d’autres peuvent constituer des indicateurs de la catégorie de la pétition, d’autres enfin sont liés au nombre de signatures des pétitions.

4.1. Des indicateurs de catégorie dans les débuts et fins de textes

Pour étudier les débuts et fins de pétitions, nous constituons des sous-corpus comprenant les 10 premiers mots et les 10 derniers mots de chaque texte, pour chaque catégorie, en passant l’ensemble des textes en minuscules pour éviter de distinguer les mots de début de phrase. La longueur de 10 mots correspond à la longueur typique d’une phrase courte (Évrard, 1990) et semble bien adaptée à nos données (voir matériel supplémentaire sur <http://jadt2016.treecloud.org>). Nous choisissons ce moyen pour cibler principalement la première phrase et la dernière phrase, car la ponctuation, utilisée de manière très variée en fonction des textes, ne peut servir de repère fiable pour repérer les fins de phrase. Nous obtenons alors pour chaque catégorie un document réunissant ces débuts de texte, et un autre réunissant les fins de texte, dont les analyses factorielles construites par le logiciel TXM sont fournies en figure 4.

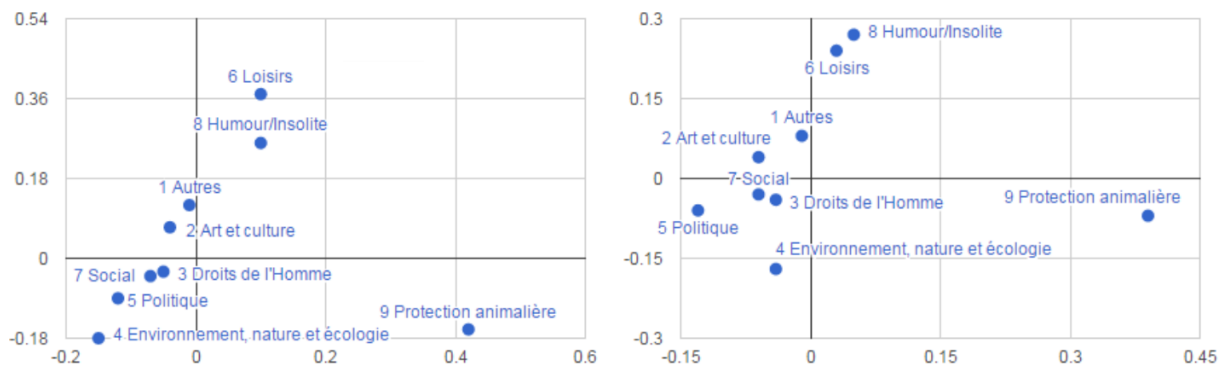


Fig. 4 - Analyses factorielles des correspondances construites par TXM sur les 10 premiers mots (à gauche) et les 10 derniers mots (à droite) des textes de pétitions, réunis par catégorie.

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 animale 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 Damnat 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.