

## ST 7

### **Les appropriations méthodologiques d'internet dans la recherche sur des objets politiques**

**Pierre Ratinaud, LERASS, Université de Toulouse, ratinaud@univ-tlse2.fr**

**Nikos Smyrnaios, LERASS, Université de Toulouse, smyrnaios@free.fr**

#### **Une méthode pour articuler analyse des réseaux et des discours sur Twitter autour des controverses politiques : la web sphère de #CharlieHebdo**

Depuis plusieurs années maintenant l'internet est devenu le terrain d'expression privilégié de controverses, c'est à dire de processus pendant lesquels des acteurs interviennent pour développer « des argumentations et des points de vue contradictoires qui les amènent à proposer des versions différentes du monde social et du monde naturel » (Callon, 1986, p. 175). Ces controverses sont particulièrement nombreuses dans le domaine politique où se construisent et se débattent les *problèmes publics* définis comme « l'ensemble des problèmes faisant l'objet d'un traitement, sous quelque forme que ce soit, de la part des autorités publiques et donc susceptibles de faire l'objet d'une ou plusieurs décisions » (Garraud, 1990, p. 27). Dans ce type de situation une multitude d'acteurs peut prendre part à une controverse: autorités, médias, organisations diverses (associations, lobbies, partis politiques) mais aussi simples citoyens sans filiation spécifique.

La particularité de ce type de débat quand il a lieu en ligne, par opposition à d'autres espaces institutionnalisés, est la difficulté à saisir ses limites temporelles, spatiales et sociales. Les controverses politiques émergent souvent de manière soudaine en fonction de l'actualité et leur dynamique et leur étendue sont imprévisibles. Le déroulement des discussions, les discours employés, les relations et les affinités qui s'y développent et les représentations qui s'y expriment dépendent d'une multitude de facteurs qu'il est difficile de prévoir *a priori*. Les controverses politiques qui se déroulent en ligne sont donc loin de présenter les caractéristiques d'un espace public normatif. C'est la raison pour laquelle, afin de pouvoir saisir cette complexité, nous faisons appel à un cadre théorique et méthodologique particulier que nous proposons de présenter ici à partir d'un cas empirique.

#### **Cadre théorique**

Dans le cadre de ce travail nous envisageons les controverses politiques en ligne comme des *web sphères*. Ce concept est à la fois un outil théorique et une proposition méthodologique. Une web sphère est définie par Schneider et Foot (2006) comme un ensemble de ressources numériques librement accessibles, réparties sur différentes pages web ou serveurs internet reliées entre elles par des liens, qui se réfèrent à un événement ou thème spécifique. Il s'agit donc d'un micro-espace public, circonscrit à la fois par une orientation thématique et des limites temporelles. Cette orientation thématique peut être de nature diverse et plus ou moins bien définie (un fait d'actualité, une question sociale ou politique, une controverse scientifique, religieuse etc.), mais elle implique toujours un problème public tel qu'il a été

défini plus haut. Une web sphère est limitée dans le temps mais sa durée de vie n'est pas connue par avance. Elle dépend de l'engagement *ad hoc* des participants qui à son tour est variable et peut subir l'influence d'une multitude de facteurs (par exemple un événement imprévu peut relancer une controverse qui semblait éteinte). L'émergence d'une web sphère peut être le résultat des stratégies (politiques, de communication etc.) ou la conséquence des facteurs institutionnels (élections, votes parlementaires, consultations). Cependant, même quand elle trouve son origine dans des conditions objectives avec des initiateurs facilement identifiables, son évolution n'est pas totalement contrôlable par ces derniers. En effet, la web sphère peut être saisie par des participants spontanés ou organisés, détournée de ses objectifs initiaux, utilisée pour parler d'autre chose, tout en gardant une certaine unité temporelle et une cohérence thématique.

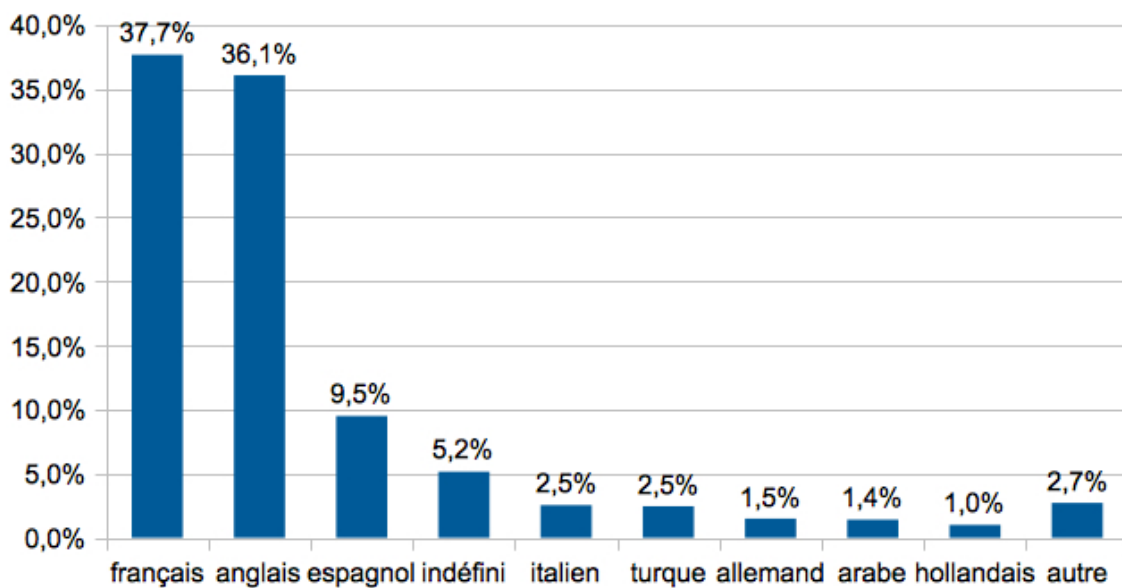
Enfin, la particularité des controverses dans le cadre des web sphères est qu'elles laissent des traces numériques pouvant être collectées et traitées de manière asynchrone. Il s'agit là d'un avantage certain pour le chercheur qui s'intéresse à ce type d'objet lui permettant de disposer d'une hauteur de vue sur des faits sociaux qui se déroulent en ligne qu'il est difficile à atteindre avec des méthodes traditionnelles en sciences sociales (ethnographie, enquête sociologique etc.). Bien entendu, la mise en œuvre des *méthodes digitales* (Rogers, 2009) pour analyser des web sphères pose également une série de problèmes épistémologiques et méthodologiques nouveaux et importants (Rieder, Röhle, 2012). Mais l'un des grands avantages de l'approche des web sphères est qu'elle prend en compte des ressources numériques et des liens, autrement dit des *textes* et des *relations*, et ceci dans un cadre temporel dynamique. A partir de ces trois composantes (textes, relations, temps) il est possible de suivre et d'analyser des controverses et des débats de nature politique à la fois de manière assez surplombante pour saisir les dynamiques globales mais également assez fine pour ne pas les surinterpréter. En effet, une web sphère n'a pas d'existence *per se*. Il n'y a pas d'instance centrale qui déclare sa naissance et définit son périmètre et sa durée. Elle est le résultat d'une observation forcément partielle et partielle. De ce fait il n'y pas d'exhaustivité possible dans son analyse, malgré l'illusion que peuvent constituer les corpus de très grande taille. Les seuils posés par le chercheur ainsi que les angles d'analyse sont décisifs quant aux résultats obtenus, comme on le verra dans l'exemple que nous allons présenter.

### **Méthode et échantillon**

A partir de ce cadre théorique nous avons élaboré une méthode que nous avons appliquée à la web sphère qui a émergé sur Twitter au moment des attaques terroristes qui ont eu lieu à Paris en janvier 2015. L'attaque sanglante de Charlie Hebdo et les événements dramatiques qui ont suivi ont déclenché une vague de réactions énorme en France et dans le monde. Comme c'est la règle désormais, ces réactions ont rapidement envahi les réseaux sociaux numériques et Twitter en particulier. Des millions de tweets ont été écrits à partir de la diffusion des premières informations, le 7 janvier en fin de matinée, et dans les jours suivants en utilisant massivement les *hashtags* #charliehebdo et #jesuischarlie. Comme souvent dans ce genre de situation les internautes se sont d'abord concentrés sur le déroulement des événements, les décrivant quasiment en temps réel, puis immédiatement est venu le temps de l'indignation, de l'hommage, du commentaire et du débat voire de la polémique. Dans cet article nous proposons une analyse d'un échantillon de tweets sur lequel nous appliquons une méthode

innovante (Smyrnaioi, Ratinaud, 2014). Nous mettons en œuvre un protocole permettant de repérer des communautés d'utilisateurs au sein de Twitter – constituées à partir d'interactions en ligne (liens) au sujet de l'attaque de Charlie Hebdo – et, en même temps, de mettre en évidence, de manière dynamique, les principales thématiques mobilisées par chacune de ces communautés (textes) et leur évolution dans le temps<sup>1</sup>.

Dans cette démarche plusieurs difficultés se présentent. Tout d'abord le volume de messages extraordinaire produit sur le sujet. Puis la diversité linguistique des messages due à l'impact mondial de l'événement. Enfin, la densité très forte des échanges entre utilisateurs. Pour cela, nous avons choisi d'effectuer notre analyse à partir d'un échantillon susceptible d'apporter un éclairage sur les ressorts des échanges sur Twitter autour de Charlie Hebdo dans toute leur diversité. Par conséquent, plutôt que de travailler sur le *hashtag* #jesuischarlie, caractérisé par un discours à forte charge émotionnelle et assez homogène dans les premiers jours, nous avons préféré analyser un échantillon à partir de la chaîne de caractères *charliehebdo*, qui fût la première et la plus utilisée. Nous avons donc collecté avec le logiciel DMI-TCAT<sup>2</sup> les messages comportant *charliehebdo* entre le 7 janvier à 13h30 et le 12 janvier à minuit, lendemain de la marche à Paris, ce qui a produit un ensemble de 3,66 millions de tweets<sup>3</sup>. Le graphique 1 rend compte des langues représentées dans cet échantillon.



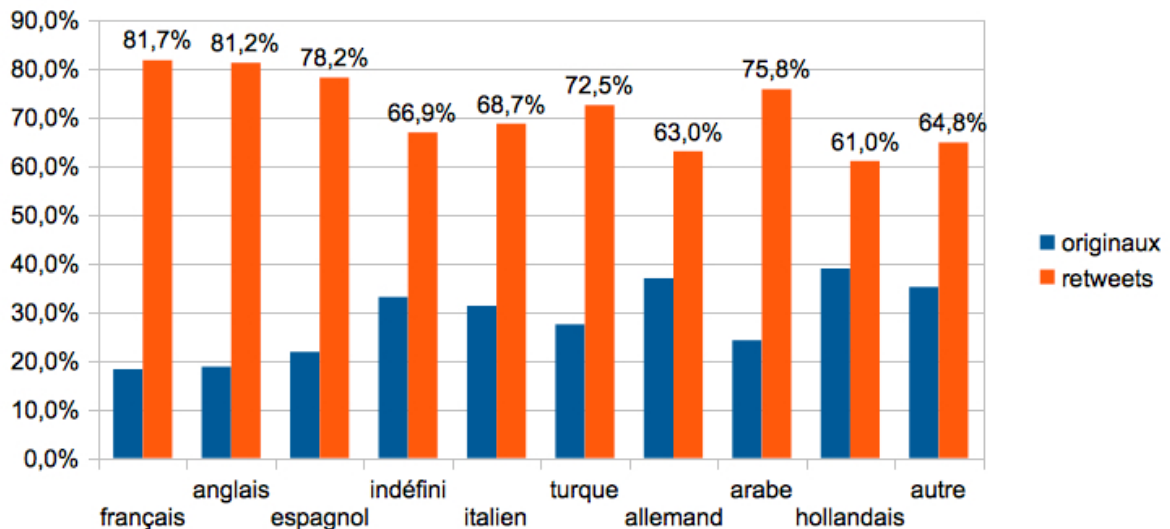
Graphique 1 : proportion des principales langues dans l'échantillon

<sup>1</sup> Une version préliminaire de cette analyse a été publiée dans Smyrnaioi N., Ratinaud P. « Twitter : ce que les graphes nous disent de #CharlieHebdo », *Inaglobal*, 5 février 2015, <http://www.inaglobal.fr/sciences-sociales/article/twitter-ce-que-les-graphes-nous-disent-de-charliehebdo-8099>

<sup>2</sup> Le *Digital Methods Initiative Twitter Capture and Analysis Toolset* est un logiciel libre développé par Erik Bora et Bernhard Rieder (2014), disponible à <https://github.com/digitalmethodsinitiative/dmi-tcat>

<sup>3</sup> En raison du très grand nombre de tweets produits comportant cette chaîne de caractères nous avons dépassé à plusieurs reprises la limite de 1% de tweets autorisés par le REST API de Twitter, ce qui implique un blocage temporaire par le service. De ce fait nous ne disposons pas malheureusement d'une collecte intégrale. Néanmoins nous pensons que l'échantillon, qui représente plus de la moitié de tweets sur la période, est tout de même représentatif de l'ensemble.

Malgré la portée mondiale de l'événement, le français est clairement sur-représenté dans cet échantillon (37% de tweets contre 2% habituellement dans Twitter<sup>4</sup>). Par ailleurs, une part très importante des messages dans chacune de ces langues est constituée de retweets, comme le montre le graphique 2, signe d'échanges très structurés autour des comptes qui concentrent un grand nombre de citations (Rieder, Smyrniaos, 2012).



Graphique 2 : proportion de tweets originaux et de retweets dans les principales langues de l'échantillon.

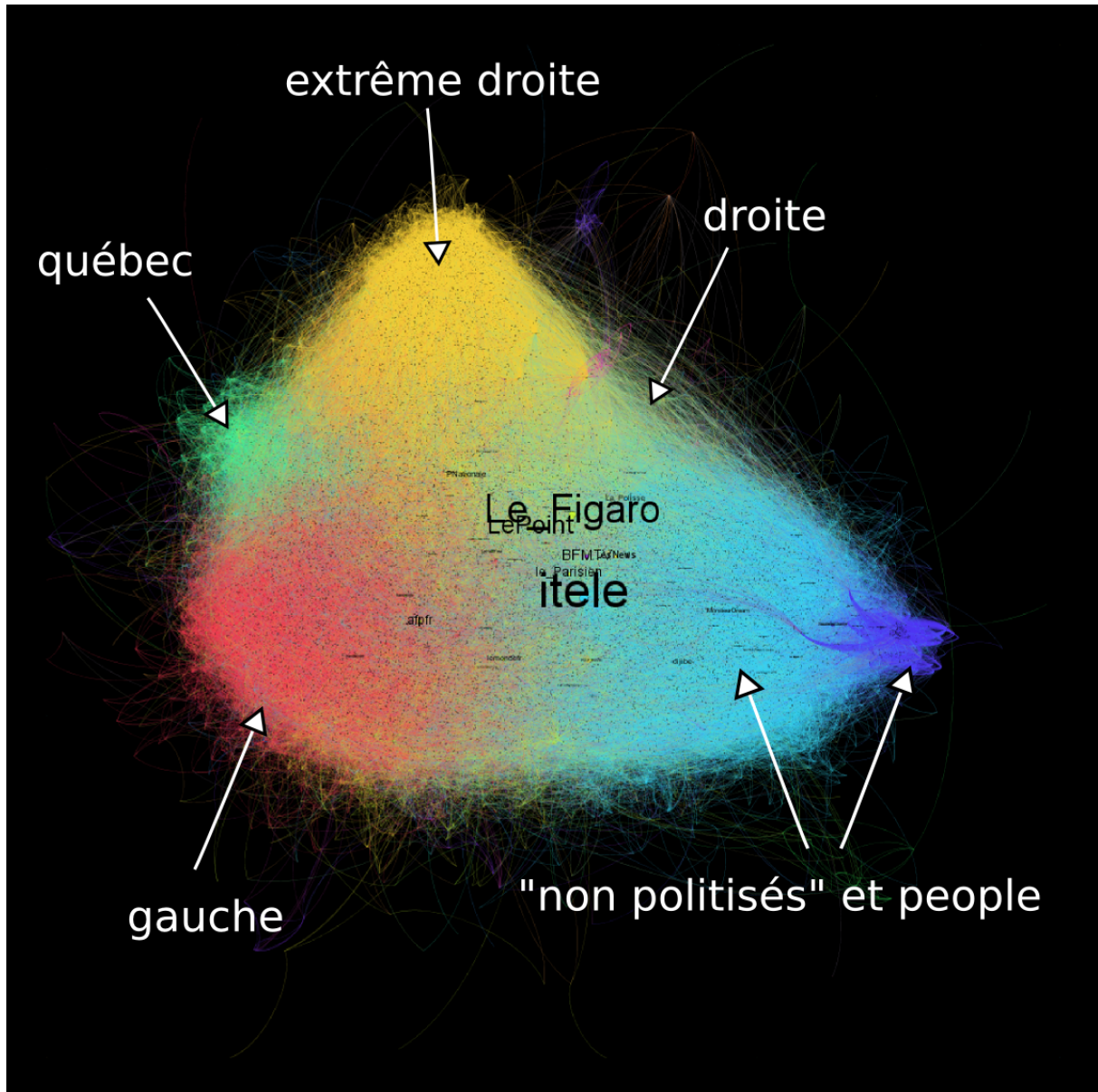
Afin d'adapter l'échantillon aux besoins de notre analyse, nous avons sélectionné les tweets en français (soit 1,38 millions de tweets) desquels nous avons retiré les retweets afin de faciliter le traitement lexicométrique. L'analyse que nous présentons porte donc sur 252 194 tweets originaux en français.

### Les réseaux autour de #charliehebdo

La première étape de notre étude a été une analyse de réseaux. Nous avons construit un graphe des interactions entre utilisateurs à partir des mentions (RT et mentions) de la manière suivante : pour chaque mention présente dans l'échantillon final, nous traitons l'utilisateur qui mentionne et l'utilisateur mentionné comme les sommets d'un graphe. Le nombre de fois où le couple mentionnant/mentionné apparaît (dans cet ordre) donnera l'ordre de grandeur des arêtes qui relient les sommets. Le graphe ainsi obtenu est donc un graphe orienté des relations entre utilisateurs de Twitter qui ont inclus dans leurs messages le terme *charliehebdo* pendant la période étudiée. En limitant l'analyse aux émetteurs ayant au moins produit 10 tweets dans l'échantillon, nous obtenons un graphe reliant 73 726 usagers (représentés par des nœuds) par 426 286 « mentions » (représentés par des liens). L'étape suivante a consisté à déterminer des *clusters* à l'intérieur de ce graphe, c'est-à-dire des ensembles de sommets fortement reliés

<sup>4</sup> Fox Z., « Top 10 Most Popular Languages on Twitter », Mashable, 18 décembre 2013, <http://mashable.com/2013/12/17/twitter-popular-languages/>

entre eux, ce qui revient dans notre exemple à déterminer des groupes d'individus qui ont tendance à se mentionner. Nous avons utilisé ici l'algorithme proposé par Blondel et al. (2008) disponible dans le logiciel Gephi<sup>5</sup> pour obtenir la carte du réseau présentée ci-dessous.



Graphe 1 : Graphe des mentions à partir des tweets des utilisateurs ayant au moins 10 tweets. La taille du texte est proportionnelle au nombre de mentions.

Comme attendu lors d'un tel événement, nous retrouvons au centre du graphe les principaux médias français. *Le Figaro*, *iTélé*, *Le Point*, *BFMTV* et *Le Parisien* présentent le plus grand nombre de mentions dans l'échantillon, mais la plupart des médias français sont présents (*Le Monde*, *AFP*, *France Info*, *L'Obs*, *Le Huffington Post*, *LesNews*, *20 Minutes*, *Charlie Hebdo*

<sup>5</sup> Gephi est un logiciel libre de visualisation des graphes disponible à <http://gephi.org>

lui même, *Canal Plus*, *La Provence*, *TF1*, *Les Echos*...) ainsi que quelques étrangers (*The New Yorker*, *Le Soir* par exemple). Il est à noter que l'algorithme de détection de communautés classe la plupart des médias (à l'exception notable de *iTélé* et de *BFMTV*) dans le même groupe (nœuds de couleur orange comme celui du *Figaro*), signe qu'ils reçoivent de nombreuses mentions par les mêmes comptes d'utilisateurs ou qu'ils se mentionnent entre eux. *iTélé* et *BFMTV* sont associés à deux clusters différents (respectivement bleu ciel et mauve), une caractéristique probablement liée à leur couverture des événements en direct qui a généré du *livetweet*. Proche du centre du graphe on trouve également un groupe de comptes liés à des instances officielles (Ministère de l'intérieur, Police nationale, Gendarmerie, Préfecture de police). Comme dans le cas des médias précités ce sont là des comptes diffusant de l'information brute et des consignes vers le public.

Cependant, un groupe particulier de médias se détache du *cluster* médiatique mentionné précédemment et fait partie d'un ensemble beaucoup plus vaste (couleur rouge, en bas à gauche du graphe). Il s'agit des médias du service public (*France Inter*, *France Culture*, *RFI*, *Radio France*, *France 24*, *FranceTVinfo*), et d'autres de sensibilité de gauche (*Libération*, *L'Humanité*, *Mediapart*, *Politis*, *Reporterre*). Dans cet ensemble on trouve de nombreux comptes de journalistes (Gilles Klein, Michel Mompontet, Cyril Petit, Sylvain Lapoix etc.), mais aussi plusieurs personnalités politiques du Front de gauche comme Jean-Luc Mélenchon, Martine Billard, Eric Cocquerel, et des utilisateurs déclarant clairement leur engagement à gauche. Un peu excentré vers le haut mais lié à ce groupe on trouve même le compte de François Hollande. Sans que ce groupe soit totalement homogène, on peut en déduire qu'il s'agit là d'un réseau d'utilisateurs de sensibilité politique similaire qui s'exprime aussi à travers le choix des médias cités. Ce phénomène d'homologie entre la sensibilité politique des usagers d'un réseau socionumérique et les sites d'information dont ils partagent les liens a été également observé sur Facebook par l'enquête Algopol (2015)<sup>6</sup>. Dans notre étude il est particulièrement flagrant à gauche de l'échiquier politique mais aussi à l'extrême droite.

En effet, aux antipodes du graphe (en haut de couleur jaune) on observe un ensemble d'utilisateurs qui se structure autour des comptes en provenance de tout l'éventail de l'extrême droite. Parmi les plus cités on trouve celui du site *Fdesouche*, proche du Front National, celui de Fabrice Robert, président du Bloc identitaire, le compte officiel du Front National, celui de Marine Le Pen etc. Il est intéressant également d'observer que ce groupe cite régulièrement – au point de les intégrer – des comptes aussi divers que celui du ministre de l'intérieur Bernard Cazeneuve, de l'expert du terrorisme controversé Samuel Laurent, celui de l'Abbé Grosjean mais aussi des médias de droite plutôt classiques comme *Valeurs Actuelles* et *Atlantico*. Les comptes de Nicolas Sarkozy et de l'UMP quant à eux occupent une position frontière entre le groupe en question et le reste du graphe en constituant leur propre (petit) groupe. Le graphe est complété par deux ensembles qui se structurent autour de comptes qu'on pourrait qualifier de *people* ou apolitiques (bleu clair et bleu foncé, en bas à droite du graphe). Les nœuds centraux de ces deux groupes sont des « Youtubeurs » (Norman, Cyprien), des présentateurs et acteurs (Nikos Aliagas, Daphné Burki, Cyril Hanouna, Omar Sy), des journalistes de télévision (Thomas Sotto, Alexandre Delpérier) et des utilisateurs lambda dont les préoccupations habituelles semblent loin de la politique. Ces comptes sont révélateurs du choc qu'a constitué l'attentat poussant à s'exprimer sur le sujet toute une

---

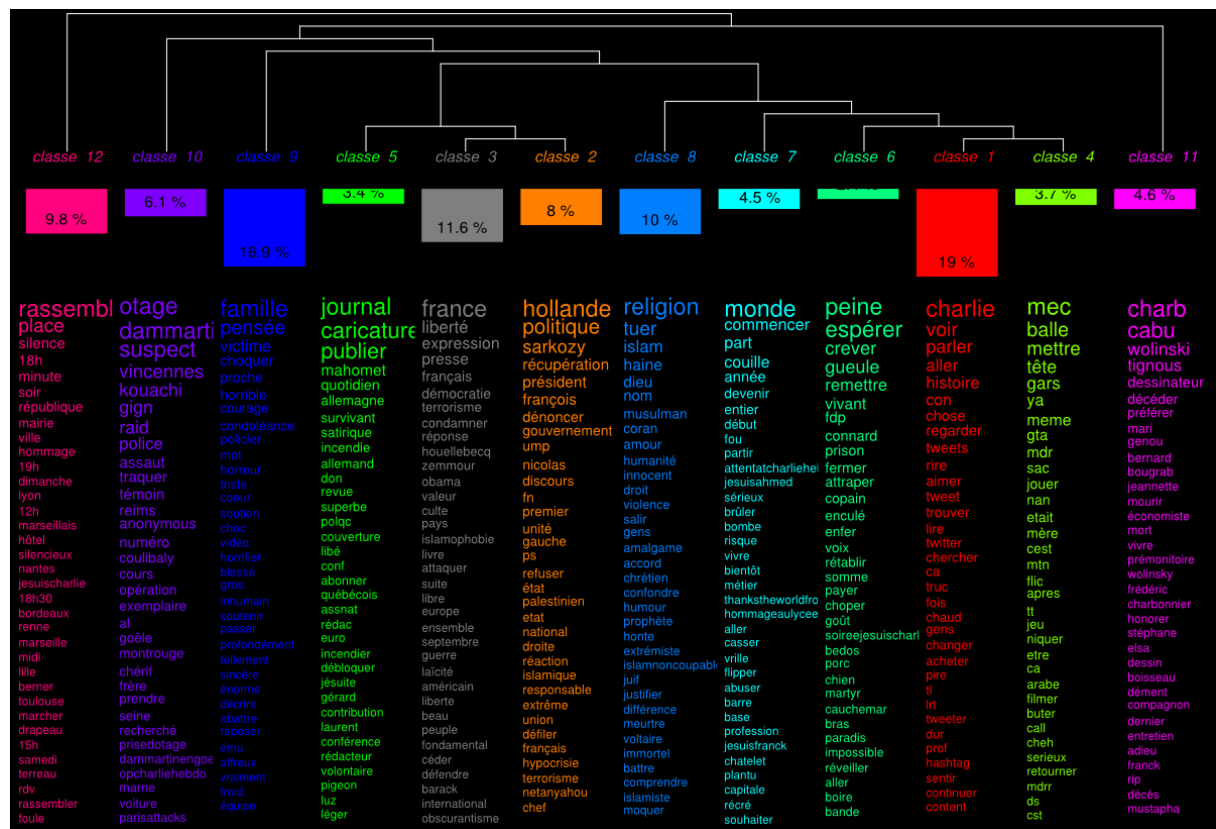
<sup>6</sup> Les premiers résultats du projet de recherche Algopol sont disponible à <http://algopol.huma-num.fr/appresults/>



population très peu politisée. Enfin, les utilisateurs québécois forment un ensemble qui se distingue clairement (vert à gauche du graphe).

### Les discours autour de Charlie Hebdo

Dans un deuxième temps nous avons utilisé le *cluster* d'appartenance des comptes pour marquer chacun des tweets dans l'analyse textométrique. Pour celle-ci, nous avons retiré des tweets les url et les mentions. L'analyse repose sur une classification hiérarchique descendante qui a pour objectif de regrouper les tweets sur un critère de cooccurrence lexicale. Autrement dit, l'analyse met dans une même classe les tweets qui ont tendance à contenir les mêmes mots (Ratinaud & Marchand, 2012). Le dendrogramme 1 rend compte des résultats obtenus en conservant les 8 000 formes pleines<sup>7</sup> les plus fréquentes du corpus (sur un ensemble de 17 818 formes pleines de fréquence supérieure ou égale à trois). Les listes de mots apparaissant sous chaque branche présentent le lexique qui est statistiquement sur-représenté dans les tweets réunis dans chacune des classes. Cette analyse a été réalisée avec IRaMuTeQ<sup>8</sup>.



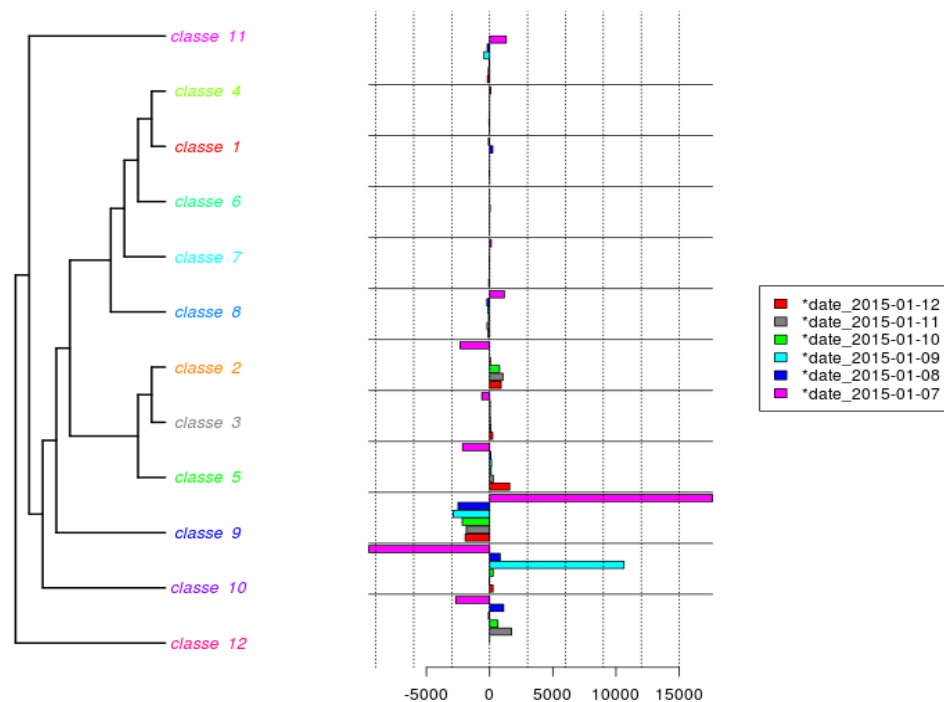
Dendrogramme 1 : résultat d'une classification sur les 8000 premières formes pleines, taille des classes et lexique caractéristique des classes

De gauche à droite dans ce graphique, nous voyons une classe rassemblant les appels aux différentes manifestations, ainsi que les commentaires sur ces événements (classe 12), une

<sup>7</sup> Seules les formes pleines participent à l'analyse. Il s'agit des noms, des adjectifs, des verbes et des adverbes.

<sup>8</sup> IRaMuTeQ est un logiciel libre développé par Pierre Ratinaud (2014) et disponible à <http://www.iramuteq.org>

classe décrivant les prises d'otage du 9 janvier (10) et une classe de discours émotionnel d'empathie et de condoléances en direction des familles des victimes des deux attaques (9). Sur la branche suivante de l'arbre, nous trouvons une classe qui décrit l'histoire du journal Charlie Hebdo (5), une classe sur la liberté d'expression (3) et une classe commentant les réactions politiques et critiquant la « récupération », notamment par François Hollande et Nicolas Sarkozy (2). La classe 8 traite spécifiquement de la question religieuse, avec de nombreux tweets dénonçant le fanatisme et le fait de tuer au nom de dieu, ainsi que des appels au refus des amalgames et des attaques contre la communauté musulmane. Nous trouvons ensuite une classe qui réunit des tweets qui s'indignent de la situation du monde en ce début d'année (7), puis une classe composée d'appels à la capture des responsables, au rétablissement de la peine de mort et d'insultes envers les terroristes (6). La classe 1 est une classe de commentaires très génériques, plutôt sur l'attaque de Charlie Hebdo, et sur les réactions de la twittosphère. Les deux dernières classes réunissent les commentaires sur la vidéo montrant un des terroristes abattre un policier à terre lors de l'attaque du siège du journal satirique (classe 4), et les tweets qui listent les noms des victimes de cette attaque (classe 11). Globalement, en termes de volume, c'est le discours d'empathie qui prédomine, ainsi que les débats sur la question de la liberté d'expression, les amalgames avec l'Islam et la « récupération » politique des événements. Le graphique ci-dessous présente le lien entre les cinq jours de l'échantillon et chacune des classes de discours. Les barres vers la droite indiquent une sur-représentation statistique des tweets émis à la date considérée pour la classe. Les barres vers la gauche indiquent une sous-représentation.



Graphique 3 : Projection des dates sur la classification

Nous pouvons voir que la chronologie des événements structure en partie la répartition des thématiques. La classe des condoléances est sur-représentée le premier jour, alors que,



logiquement, les commentaires sur les prises d'otages sont sur-représentés le 9 janvier jour où celles-ci ont eu lieu. On notera que les classes 7, 6, 4 et 1 ne présentent pas de différences d'un point de vue chronologique. Les tweets traitant de ces questions sont présents dans les mêmes proportions chaque jour de l'indexation.

### Quel genre de discours pour quel groupe ?

Une autre dimension susceptible d'expliquer la structuration de l'analyse est l'appartenance du locuteur à l'un ou l'autre groupes d'utilisateurs identifiés. Le tableau 1 présente les effectifs de tweets provenant de chacune des communautés dans le corpus.

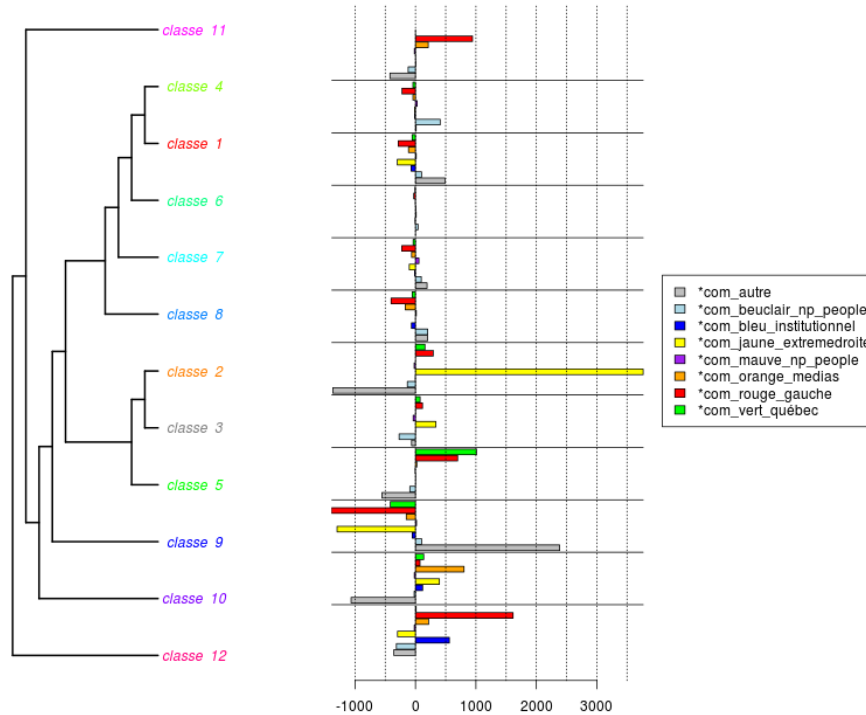
	Effectifs	Pourcentage
*com_autre	148677	58.95 %
*com_beuclair_np_people	31069	12.32 %
*com_bleu_institutionnel	3655	1.45 %
*com_jaune_extremedroite	16067	6.37 %
*com_mauve_np_people	1444	0.57 %
*com_orange_medias	17244	6.84 %
*com_rouge_gauche	28048	11.12 %
*com_vert_québec	5990	2.38 %
	252194	100 %

Tableau 1 : effectifs et proportions de chacune des communautés dans le corpus

Nous pouvons constater que dans le corpus, plus de la moitié des tweets ne sont pas émis par un utilisateur appartenant à l'une des communautés retenues (\*com\_autre). C'est la conséquence de deux phénomènes distincts :

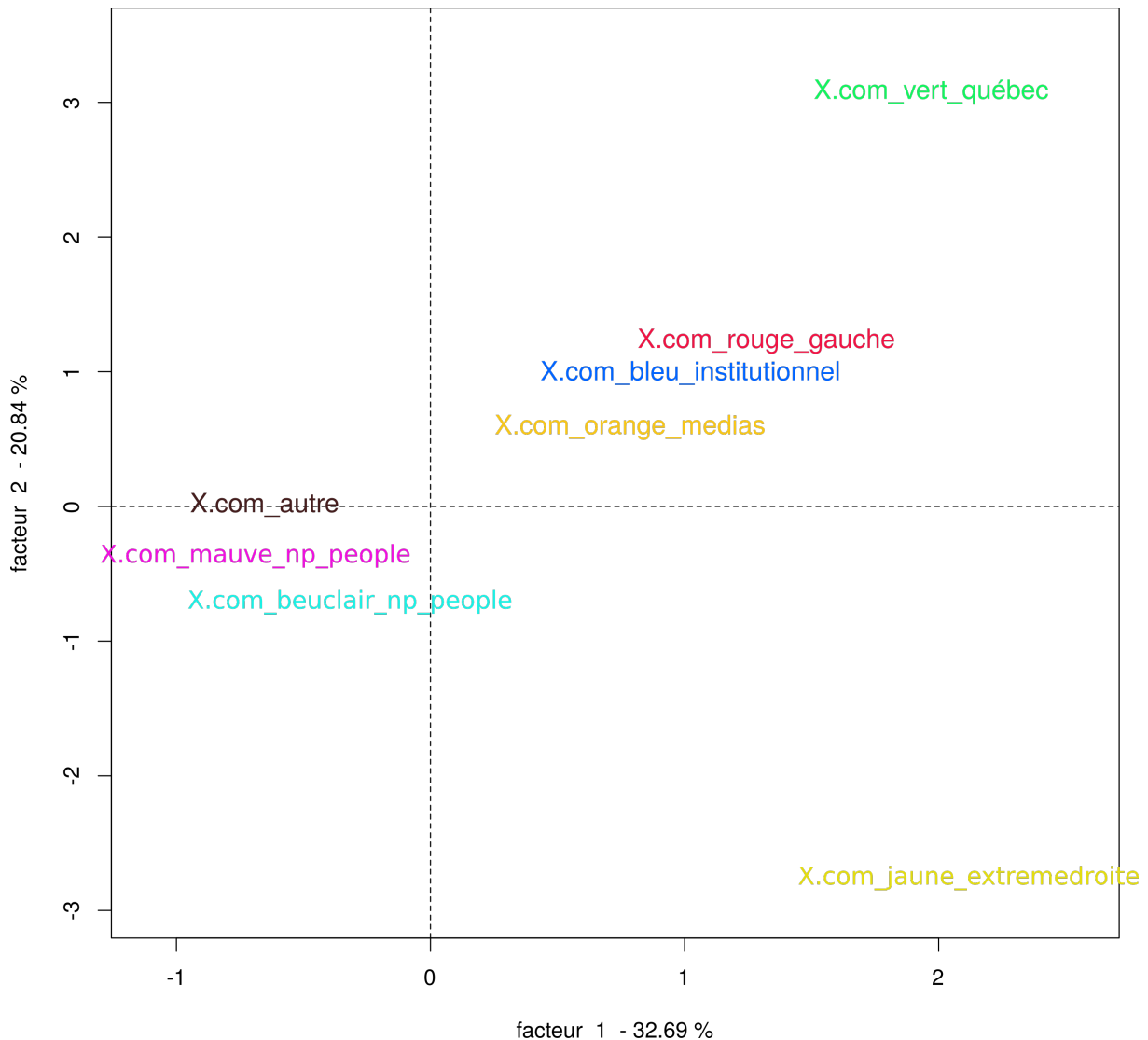
- l'une des raisons pouvant expliquer l'absence d'appartenance à une communauté est le fait de ne pas faire parti de l'échantillon retenu pour l'analyse de graphe. Ceci soit parce que l'utilisateur a produit moins de 10 tweets, soit parce qu'il n'est jamais mentionné par un utilisateur ayant produit au moins 10 tweets.
- L'autre raison est que certains utilisateurs du groupe « autre » faisait effectivement parti de l'analyse de graphe mais ils ont été positionnés dans un *cluster* trop petit pour être retenu.

Le graphique 4 rend compte des liens entre ces groupes et l'analyse lexicale. Il témoigne des tendances générales de chacun des groupes à se saisir des différentes thématiques qui apparaissent dans le corpus.



Graphique 4 : Projection des groupes sur la classification

On trouve, de façon assez logique, le groupe des médias sur-représentée dans la classe qui décrivent les événements du 9 janvier (10). On comprend également que la classe qui décrit ce qu'est Charlie Hebdo (5) est proportionnellement plus portée par le groupe des utilisateurs québécois, le journal étant moins connu de l'autre côté de l'Atlantique, mais aussi dans un moindre degré par le groupe de gauche. On note aussi la très forte sur-représentation du groupe d'extrême droite dans la classe de dénonciation de la « récupération » politique (2), ainsi que dans celle sur la liberté d'expression (3) et sur les commentaires des événements du 9 janvier (10). L'extrême droite, comme la gauche, paraissent sous-représentées dans la classe de discours émotionnel d'empathie (9). A l'inverse le groupe de gauche est fortement sur-représenté dans la classe rassemblant les appels aux différentes manifestations (9), comme les comptes institutionnels dans un moindre degré, et celle qui parle des dessinateurs assassinés de Charlie Habdo (11). La communauté d'extrême droite est donc celle qui semble le plus particulariser un type de discours. L'AFC sur le tableau lexical complet (Lebart et Salem, 1994) de ce corpus partitionné selon les groupes montre le même phénomène.



Graphique 5 : AFC sur le tableau lexical complet (uniquement les formes pleines, effectif minimum à 10)

Le discours porté par la communauté d'extrême droite se différencie de celui de toutes les autres communautés par son positionnement polaire sur les facteurs 1 et 2. Notons que le positionnement de la communauté du Québec est trompeuse sur ce graphique : elle est en faite fortement tirée à l'arrière de ce plan par le facteur 3.

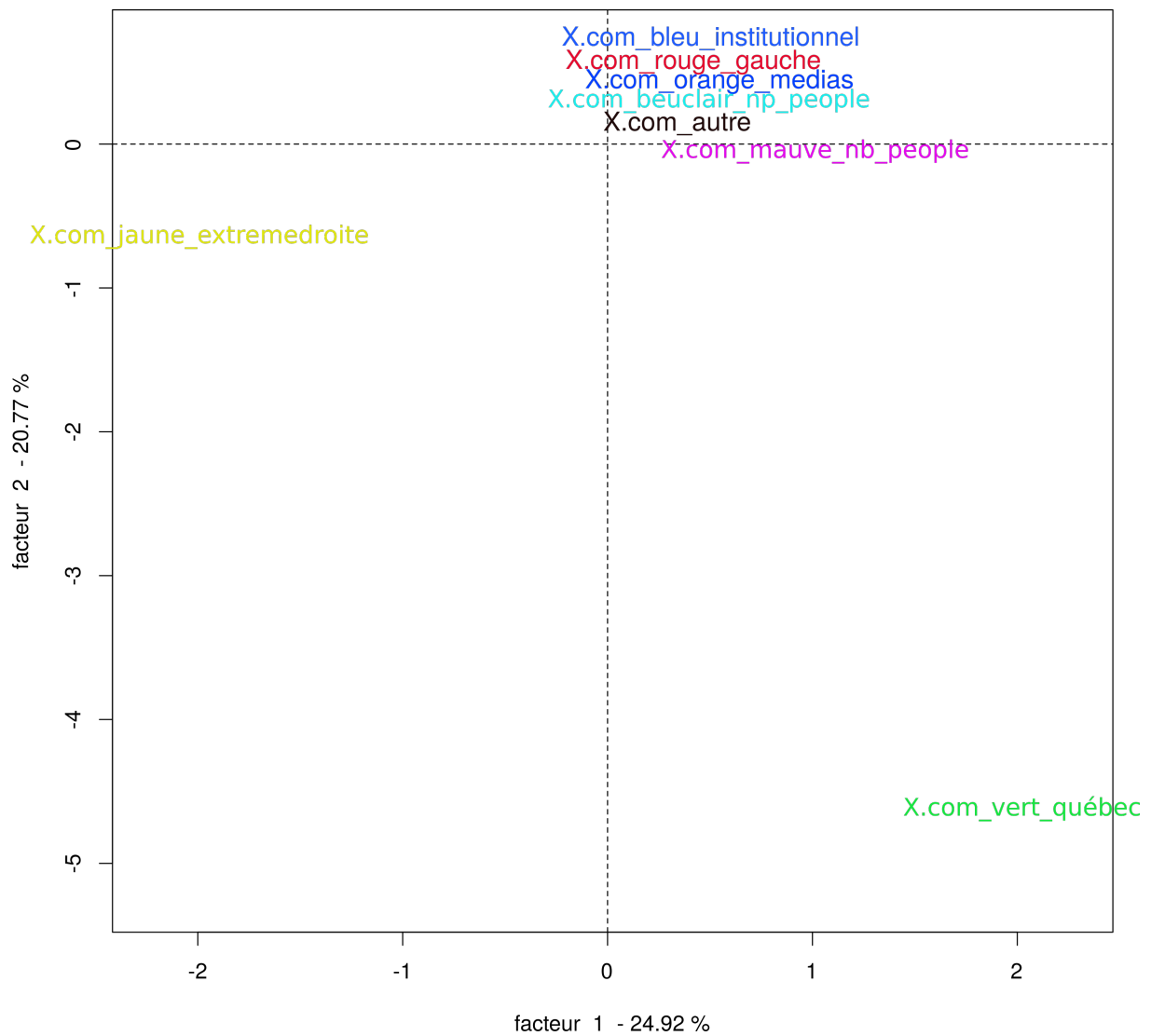
### Classification de deuxième niveau

La plupart des classes produites par la première classification présentent encore un certain niveau d'hétérogénéité. Dans l'objectif d'augmenter la précision de l'analyse et de pointer la particularité du positionnement de l'extrême droite, nous avons soumis la classe dans laquelle les tweets de cette communauté sont les plus sur-représentés à une nouvelle analyse. Pour ce faire nous avons reconstitué un sous-corpus à partir de l'ensemble de tweets appartenant la classe 2.

	Effectifs	Pourcentage
*com_autre	8151	45.16 %
*com_beuclair_np_people	1784	9.88 %
*com_bleu_institutionnel	244	1.35 %
*com_jaune_extremedroite	3122	17.3 %
*com_mauve_np_people	62	0.34 %
*com_orange_medias	1286	7.13 %
*com_rouge_gauche	2725	15.1 %
*com_vert_québec	674	3.73 %
	18048	100 %

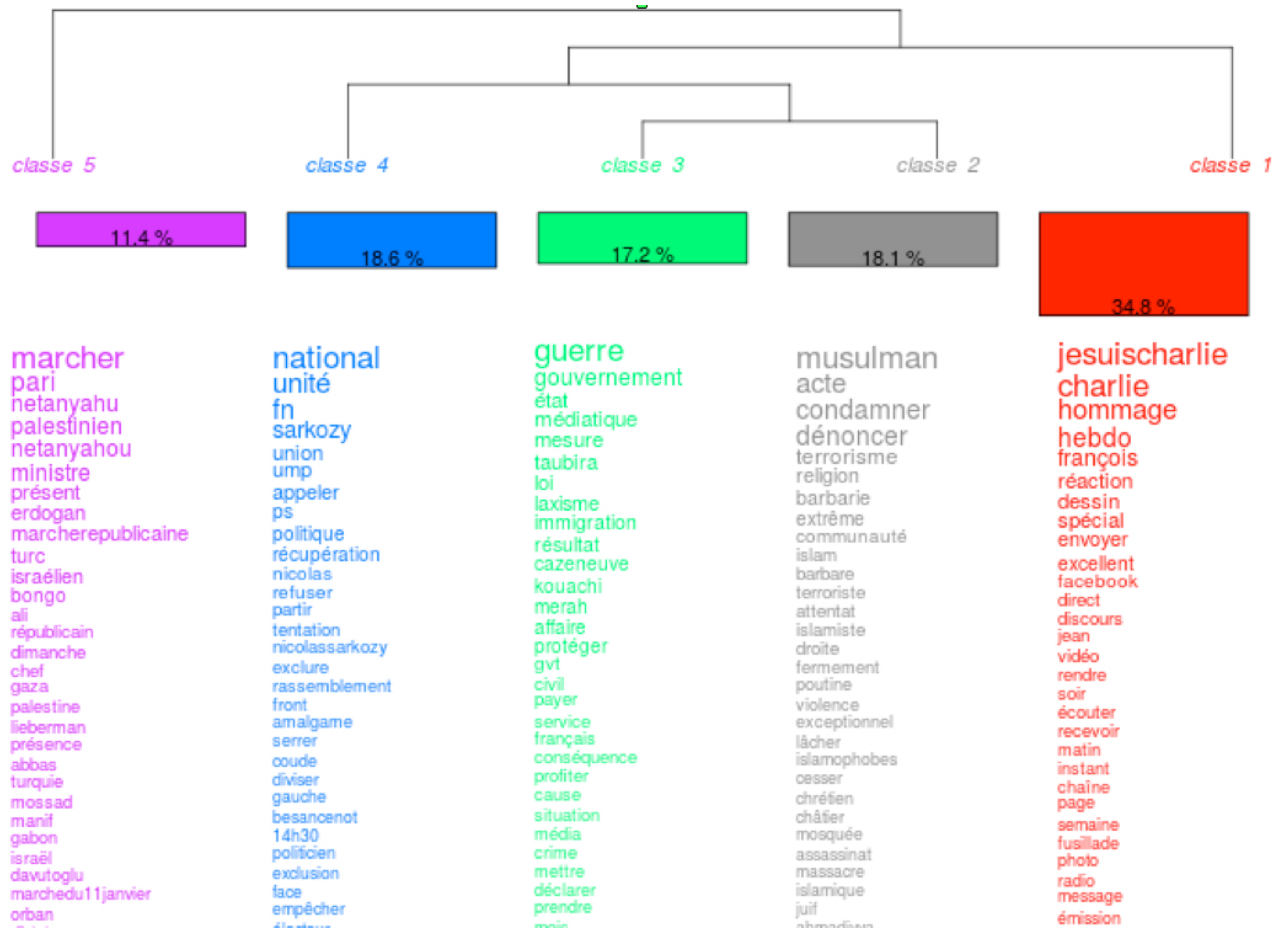
Tableau 2 : détail et répartition des effectifs de la classe 2 (18048 tweets et 273667 occurrences)

Le tableau 2 traduit la sur-représentation de l'extrême droite dans cette classe. Alors que les tweets de cette communauté constituent 6,37 % du corpus, ils forment 17,3 % de cet ensemble. L'AFC sur le tableau lexicale complet semble effectivement pointer une différenciation lexicale de la communauté d'extrême droite dans cette classe.



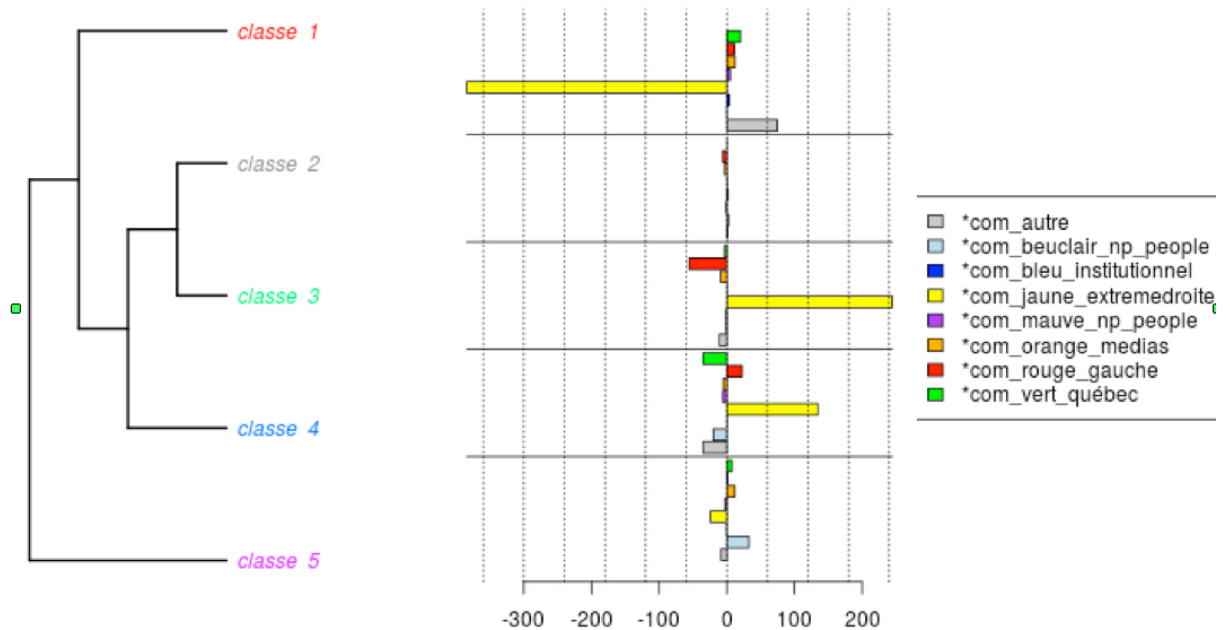
Graphique 6 : AFC sur le tableau lexical complet (uniquement les formes pleines, effectif minimum à 5).

Nous avons soumis ce corpus à une nouvelle classification. Le dendrogramme 2 rend compte des 5 classes obtenues.



Dendrogramme 2 : résultat d'une classification des tweets appartenant à la classe 2 de l'analyse de premier niveau présenté dans le Dendrogramme 1

Nous obtenons ainsi une classe qui se réfère à la marche de Paris du 11 janvier (classe 5), une autre qui porte sur la polémique politique autour du thème de l'unité nationale (4), une classe qui critique le gouvernement pour sa supposée incapacité à gérer une telle crise qui s'apparenterait à une guerre (3), une classe sur Islam et terrorisme (2) et une classe très volumineuse qui diffuse les messages d'hommages aux victimes, notamment ceux provenant des sphères politique et médiatique (1, Je suis Charlie).



Graphique 5 : Projection des communautés sur la classification de deuxième niveau

Nous observons que le groupe d'extrême droite est sur-représenté dans la classe de critique du gouvernement et de son incompétence supposée (3) ainsi que, de manière moins prononcée, dans la classe de polémique autour de l'unité nationale à l'exclusion du Front National (4). De façon finalement assez logique, la gauche est sous-représentée dans la classe 3. Cette analyse de deuxième niveau révèle donc clairement un positionnement particulier de l'extrême droite au niveau de son discours sur Twitter par rapport aux événements de Charlie Hebdo. Ce discours est caractérisé d'une part par une propension à s'opposer à toute unité des forces politiques face au danger du terrorisme, considérée comme de façade, et à critiquer le gouvernement et le Président en mettant en cause leur légitimité et leur compétence. D'autre part le groupe d'extrême droite a tendance à se distancier des hommages unanimes aux victimes. Cette opposition apparaît de manière limpide dans l'analyse factorielle des correspondances qui croise communauté et lexique.

En conclusion, les résultats de notre analyse contredisent, au moins partiellement, l'idée que l'attaque contre Charlie Hebdo a produit un point de vue unanime, au moins pour ce qui concerne notre échantillon. D'abord on s'aperçoit que les échanges sur Twitter à ce sujet ont été structurés fortement par une logique d'homophilie politique. Ensuite, si la condamnation et l'indignation semblent constituer la norme pour la grande majorité des commentaires, il apparaît aussi que le positionnement politique demeure un puissant facteur de différenciation du discours adopté mais aussi des thématiques et des sources privilégiées. Par exemple, les utilisateurs de Twitter proches de l'extrême droite se saisissent de la web sphère qui émerge autour de Charlie Hebdo pour déployer un discours politique d'opposition au gouvernement et à l'unanimité jugé factice. Les utilisateurs appartenant au groupe de gauche à leur tour sont plus prompts à relayer la mobilisation en faveur de Charlie Hebdo et à parler du sujet, c'est à dire du journal, de ses dessinateurs etc. Ces derniers sont aussi plus à même de relayer des médias publics ou engagés.



Bibliographie :

Blondel, V., Guillaume, J.-P., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, (10), p.1008

Borra, E., Rieder, B. (2014). Programmed method: developing a toolset for capturing and analyzing tweets, *Aslib Journal of Information Management*, Vol. 66 Issue 3, p.262-278.

Callon M., (1986). Eléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc, *L'année sociologique*, (36), p.169-208.

Garraud P., (1990). Politiques nationales : l'élaboration de l'agenda. *L'année sociologique*, (40), p.7-41.

Lebart L., Salem A., (1994). *Statistique textuelle*. Paris : Dunod.

Ratinaud, P. (2014). IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires (Version 0.7 alpha 2). Retrieved from <http://www.iramuteq.org>

Ratinaud, P., Marchand, P. (2012). Application de la méthode ALCESTE à de “gros” corpus et stabilité des “mondes lexicaux” : analyse du “CableGate” avec IRaMuTeQ. In *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles*, p.835–844. Liège, Belgique.

Rieder B., Röhle T. (2012). « Digital methods: five challenges », in Berry D.M. (dir.), *Understanding digital humanities*. Houndmills: Palgrave Macmillan, p.67-84.

Rieder B., Smyrniotis N. (2012). Pluralisme et infomédiation sociale de l'actualité: le cas de Twitter. *Réseaux* no 176, p.107-141

Schneider, S. M., Foot, K. A. (2006). « Web Sphere Analysis. An Approach to Studying Online Action », in Hine, C. (dir.), *Virtual Methods. Issues in Social Research on the Internet*, Oxford and New York: Berg, p.157–170.

Smyrniotis, N., Ratinaud, P. (2014). Comment articuler analyse des réseaux et des discours sur Twitter : l'exemple du débat autour du pacte budgétaire européen. *tic&société*, 7(2), p.120-147.