HAL
open science

Le tweet politique efficace comme mème textuel : du profilage à viralité

Julien Longhi

To cite this version:

Julien Longhi. Le tweet politique efficace comme mème textuel : du profilage à viralité. Travaux de Linguistique : Revue Internationale de Linguistique Française, 2016, Les mèmes langagiers. Propagation, figement, déformation, 73 (2016-2). halshs-01483945

HAL Id: halshs-01483945

https://shs.hal.science/halshs-01483945

Submitted on 6 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le tweet politique efficace comme mème textuel : du profilage à viralité

Julien Longhi, Université de Cergy-Pontoise, EA7292 AGORA (julien.longhi@u-cergy.fr)

Résumé :

Dans le cadre d'une recherche sur l'analyse des tweets politiques, cet article s'intéresse à la propension de ces formes à circuler, en se focalisant sur leur efficacité, définie en fonction du nombre de fois où un tweet est « retweeté » ou « favorisé ». Cette efficacité, au regard de la circulation des messages sur internet, permet de saisir leur viralité[1] : nous faisons ainsi l'hypothèse de l'existence, pour le tweet politique, d'une forme de mème plus textuelle que langagière, puisque c'est, tout autant que l'emploi de certains termes, leur mise en syntagme particulière et selon certains points de vue (le profilage) qui en fait le succès et la viralité. C'est donc à l'échelle du tweet dans son ensemble que se résoudrait la question de sa viralité, par la prise en compte de la circulation et de son efficacité.

Abstract :

As part of a research on political tweets, this article focuses on the propensity of these structures to circulate, with a focus on efficiency, defined by the number of times a tweet is "retweeted" or "favorised". This efficiency, according to the traffic on the internet messages, shows their virality: we do the hypothesis of the existence, for political tweets, of textual memes, because more than the use of certain terms, it is their particular implementation in sentences, and according to some views, that makes their success and virality. By taking into account the traffic and the effectiveness of political tweets, we would resolve the issue of the memes 'virality.

Cet article s'intègre dans le cadre de deux projets de recherche (2014 : Humanités numériques et datajournalisme : le cas du lexique politique ; et 2015 : ICI 2.0, « Institutions, communautés et idéologies sur le web et les médias sociaux ») dont un des objectifs est la détection et la caractérisation d'idéologies (selon des critères définis dans Sarfati 2014) sur

[1] On considérera que l'efficacité des tweets, mesurable avec les « retweets » et « favoris », est un bon indiquateur de leur viralité, qui permet de mesurer, en communication ou e-marketing, le taux d'utilisateurs touchés par un message.

Twitter, à partir de la constitution de « règles » linguistiques pour repérer les spécificités formelles de l'idéologie[2].

Une partie de cette recherche nous a conduit à nous interroger plus largement sur la propension des tweets à circuler, en nous focalisant sur leur efficacité, définie en fonction du nombre de fois où un tweet est « retweeté » (c'est-à-dire renvoyé par un twittos à l'ensemble de ses « followers » - twittos qui le suivent) ou mis en favori. L'objectif est de pouvoir corréler des critères linguistiques de la mise en discours avec le pouvoir de circulation (effet « boule de neige » sur le réseau). En effet, malgré une certaine hybridité (Longhi 2014[3]), les tweets politiques partagent des propriétés linguistiques, sémantiques et discursives qui motivent leur circulation (Longhi 2013). La question des mèmes peut donc être posée à propos des tweets politiques, afin de voir dans quelle mesure ils peuvent constituer un point d'entrée vers l'étude du phénomène de viralité, considéré du point de vue linguistique. Après avoir posé les termes de cette question au regard de la littérature existante sur les mèmes et la contagion des idées, nous proposerons, dans le cadre de l'analyse du discours, une perspective sémantique des discours appuyée sur la notion de perception sémantique (Rastier, 2006 : 103). Ceci nous permettra de poser la question des mèmes dans un autre cadre théorique et méthodologique, en la fondant sur le rapport entre les fonds et les formes sémantiques. Nous étudierons alors plus en détail le sous-corpus des tweets les plus efficaces, afin d'en montrer les caractéristiques principales pouvant répondre – moyennant quelques redéfinitions – à la définition d'un même dans le contexte de l'analyse des tweets. Il s'agira donc plus particulièrement de tester l'hypothèse que les tweets porteurs de mèmes linguistiques ont une propension à la circulation supérieure aux autres tweets, ou sont au moins crédités d'un potentiel dialogique qui leur confère une efficacité spécifique. Ce faisant, nous serons amenés à considérer les tweets efficaces comme des mèmes textuels, notion qui invite à revoir le concept de *mème* du point de vue des processus de sémiotisation textuels et discursifs.

---

[2] Djemili, Longhi et al. 2014. Cette recherche s'appuie sur un corpus de tweets politiques, *Politiwteets*, composé de 34273 messages, 205 utilisateurs, soit un corpus de tweets issus de comptes influents. Voir https://repository.ortolang.fr/api/content/comere/v3.1/cmr-polititweets.html pour le téléchargement du corpus et les précisions sur la constitution, notamment au regard du critère d'« influence ». Nous présenterons le corpus en détail au point 2.2.

[3] Nous avions notamment mis en valeur leur statut de « petite phrase » relativement au genre politique, et un usage assez restreint des dimensions technologiques favorisant le référencement et la circulation.

1 La question des mèmes du point de vue de l'analyse du discours

La question de la circulation des discours, du sens, des formes, est centrale pour l'analyse du discours, que cela concerne le dialogisme, la polyphonie, la reprise, la subversion d'éléments formels ou de sens, etc. Pourtant, telle qu'elle a été posée, à cause notamment de la métaphore biologique qui en est à l'origine (voir la présentation de ce numéro), la question des mèmes renvoie à une approche plus cognitiviste du langage. Une définition des différents concepts est donc nécessaire pour appréhender les mèmes du point de vue de l'analyse du discours. En linguistique, la forme est généralement définie comme l'« aspect sous lequel se présente un mot ou un énoncé », et, « par métonymie », comme l'« unité linguistique (morphème ou construction) identifiée par ses traits formels » (TLFi). Cependant, l'analyse du discours mobilise le concept de forme en relation avec certains aspects énonciatifs, argumentatifs, voire idéologiques (Sarfati 1996, Garric et Longhi 2013). La corrélation sémiotique entre forme et visée discursive (Longhi 2011) oriente donc notre conception de ce que pourrait être un mème, en considérant davantage la contribution des formes sur la constitution du réel par les sujets : les formes mobilisées en discours par les sujets parlants sont déterminées par la construction que ces formes contribuent à effectuer lors de la mise en discours, et la mise en discours est vue comme un processus d'appréhension et de constitution du monde social[4].

1.1 Le concept de mème et sa portée pour l'analyse du discours

Antoine Gautier et Gilles Siouffi, dans leur présentation à ce numéro, reviennent sur la naissance du concept de mème et sur sa définition : c'est sur le modèle de l'évolution du vivant, caractérisée par la sélection et la transmission de gènes, que Dawkins a forgé le terme de mème, qu'il définit comme une unité de transmission culturelle, un réplicateur culturel. Pour Dawkins, de la même manière que les gènes se répliquent d'un organisme à l'autre, les mèmes se propageraient d'esprit en esprit par un processus d'imitation ou d'apprentissage. Pour Dan Sperber, à l'inverse (Sperber, 1996), la stricte réplication doit être considérée comme un cas limite de ressemblance maximale plutôt que la norme de la communication : « un processus de communication est fondamentalement un processus de transformation » (Sperber, 1996 : 115). Dan Sperber lie la communication au partage de croyances, explicitable par la prise en compte des représentations des sujets.

[4] « Le langage […] est une activité de constitution plutôt que de représentation, et les rapports entre sens et référent sont de l'odre de l'extension » (Lebas et Cadiot 2003 : 4).

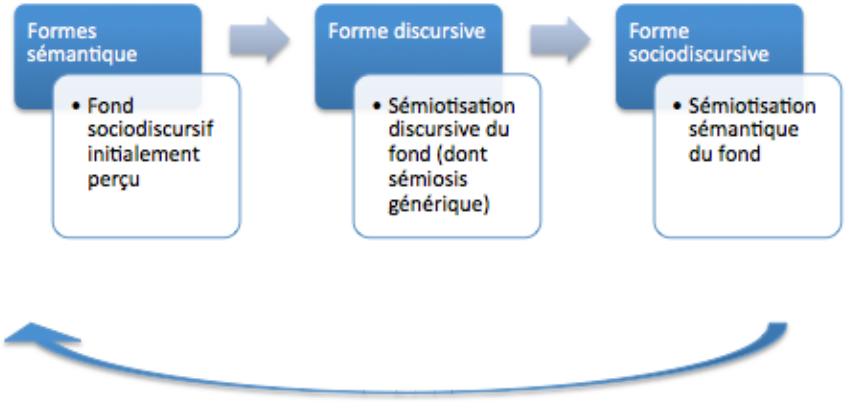
Nous souhaitons dans cet article questionner davantage le liage communication/cognition/culture, en considérant de manière plus précise les formes langagières, et mettre en valeur leur contribution fondamentale dans la construction des représentations des sujets. En effet, « de même que les signes déterminent le réel en vertu d'un système doxique qui en conditionne l'expression, toute intervention critique sur le réel (ainsi constitué) relève prioritairement d'une intervention critique sur la réalité discursive et sémiotique (Sarfati, 1996 IV : 4). Le rapport entre la « réalité » et les représentations doit donc être problématisé, notamment au regard du concept de sens commun, qui conditionne l'appréhension que nous faisons du sens. En lien avec nos travaux en cours sur le concept de forme en analyse du discours (Longhi 2015 et la *Théorie des objets discursifs*), nous considérons que le discours opère comme un champ, à l'intérieur duquel les formes se disposent et se créditent d'un sens. Ceci fait écho à la méthodologie proposée dans la sémantique des textes, et nous retenons de François Rastier la conception générale de la perception sémantique, dans laquelle « l'activité énonciative et interprétative consiste à élaborer des formes, établir des fonds et faire varier les rapports entre fond et forme. […] Ces variations permettent à l'énonciateur de concilier autant qu'il le peut ou le veut les contraintes de la langue, du discours, du genre, de la situation et les rémanences de ce qu'il a déjà dit ou écrit » (2006 : 112). Ce faisant, nous modelons néanmoins de manière différente le processus sémantique, qui met au centre le concept de forme, et son enrichissement dynamique au regard des différentes phases qu'il mobilise (nous y reviendrons au pout 1.2). Cette prise en compte du « fond » se transpose, dans le cadre du discours, en « fond sociodiscursif », qui relève de ce que Georges-Elia Sarfati (1996) a théorisé dans sa reprise du concept de sens commun d'un point de vue linguistique et discursif. Aussi, pour prolonger ce que propose Dan Sperber pour décrire la communication, nous considérons que le sens commun nécessite d'être étudié de manière sémantique et discursive, afin de pouvoir proposer, dans un second temps, des développements en terme de « contagion ».

1.2 La co-construction des formes et des fonds sociodiscursifs : l'action du discours

La question du déploiement des formes renvoie à la question de l'action que font les sujets parlants par les discours. En ce sens, la définition de Victor Rosenthal et Yves-Marie Visetti est importante, car elle situe la question de la perception dans la dynamique qui l'accompagne, et « définit par sa perspective globale les structures et les significations locales

du champ […]. Le Présent du champ pratique intègre donc l'activité dans son dynamisme propre » (2003 : 161).

Comme les auteurs le précisent, dans cette dynamique se co-déterminent formes, valeurs et actions produites par le sujet parlant : « faire sens, pour un percept, c'est d'abord réussir à paraître et se définir dans le champ, en respectant une certaine cohérence globale qui est en jeu dans tout acte de percevoir » (Rosenthal et Visetti, 2003 : 167). Aussi, de la même manière, faire sens pour une forme, c'est se définir dans le système du sens commun, et redéfinir conjointement ce système selon le devenir de l'action de sémiotisation[5] constituée par le discours et l'impact des formes produites :



Restructuration perceptive du système de l'arrière-plan de croyances (Longhi 2015 : 50)

Avec ce schéma, nous souhaitons montrer qu'il y a une co-détermination des fonds et des formes sémantiques : la récurrence, la stabilisation, la circulation, des formes sémantiques, opèrent en effet sur le système du sens commun, qui se réajuste pour les sujets parlants au fil des productions discursives.

Dans cet esprit, nous proposons d'aborder les « mèmes » comme des formes qui connaissent une stabilisation et aussi un pouvoir de circulation spécifiques. L'hypothèse que nous faisons à propos des mèmes est que leur pouvoir de circulation ou de viralité relève à la fois de leur spécificité et originalité, du rapport qu'ils entretiennent avec l'arrière-plan de croyances dans

[5] Par sémiotisation, nous souhaitons souligner la contribution des signes constitutifs des discours dans le processus sémantique.

lequel ils sont mis en discours, et de la manière dont ils peuvent modifier l'écologie initiale de cet arrière-plan.

2 La circulation des mèmes : notions de base

Au regard de la mise en discussion du concept de mème face à une approche sémantique et discursive, nous faisons l'hypothèse que le pouvoir de circulation et de stabilisation de ces formes tient notamment à leur insertion en syntagme, car la brièveté des tweets confère à ces structures une organisation interne spécifique. C'est donc par le concept de « profilage » que nous proposons d'aborder cette question.

2.1 Le profilage

Le concept de profil, considéré comme le résultat d'une opération de profilage, est défini ainsi par Pierre Cadiot et Yves-Marie Visetti (2001 : 130) :

> Ces dynamiques de profilage renvoient pour une part à des frayages déjà enregistrés en lexique et, sous une forme bien plus générique, en grammaire. Mais elles se font aussi par inscription dans des thématiques inédites, qui les reprennent dans leur propre grille, possiblement extrinsèque, soit aux motifs donnés en langue, soit aux normes de profilage lexical déjà attestées.

Il s'agit en fait, dans les processus de mise en discours, de concevoir des parcours de sens et de différenciation qui se fondent sur une certaine mémoire discursive, ou une créativité rendue possible notamment par la brièveté des messages. Les profilages tiennent donc compte de la spécificité des motifs engagés, ainsi que des spécificités de la textualité. Ces profilages font intervenir la différentialité, puisque « dans le profilage, la différentialité est partout, et résulte des processus d'individuation qui entrent alors en jeu » (Cadiot et Visetti, 2001 : 133). Du point de vue de l'émergence du sens, « les fonds sémantiques semblent des suites de points réguliers et comme les formes sont discrétisées par leurs points singuliers, le parcours productif ou interprétatif de ces formes et de ces fonds suppose un rythme, cellule de base de toute action » (Rastier, 2006 : 106). Nous relions ici encore ce rapport fond/forme au concept de profilage, qui permet à certaines formes de se distinguer, de s'individuer, et finalement de pouvoir acquérir un statut de mème : ce statut n'est pas uniquement lié à la forme elle-même

(ce que nous appellerions « mème linguistique »), mais à la mise en discours, dans l'espace textuel du tweet, de cette forme, ce qui nous amène à considérer des « mème textuels ». En effet, considérer l'opération de profilage comme centrale dans l'acquisition d'une certaine viralité pour le tweet, c'est certes tenir compte des formes qui le composent, mais c'est surtout considérer l'interaction des différentes formes présentes dans un même tweet.

Nous postulons donc à ce stade de l'appréhension des mèmes dans le cadre d'une analyse du discours politique numérique (tweets), qu'ils sont des formes spécifiques, stabilisées et circulantes, dont la nature est propice à la viralité ou à la diffusion, et qu'ils mobilisent un rapport fond/forme qui a une portée sur l'économie socio-discursive du discours. Nous allons donc développer cette hypothèse en la confrontant à un corpus spécifique.

2.2 Efficacité, circulation et viralité

Le choix des comptes du corpus *Polititweets* a été guidé par le fait que nous pouvons y mettre en œuvre le critère de l'influence sur Twitter. Pour cela, les étapes suivantes ont été appliquées :

1. Nous sommes parti de 7 personnalités de 6 groupes politiques (les 6 principaux partis de l'échiquier, avec à l'époque un binôme Fillon/Copé en compétition pour la tête de l'UMP, d'où le chiffre de 7) ;
2. Nous avons récupéré toutes les listes créées par des utilisateurs où ils étaient cités : cela a conduit à 7087 listes ;
3. Nous avons sélectionné parmi ces listes celles qui avaient au moins 6 twittos et qui contenaient la chaîne de caractère *politic* dans le nom ou descriptif de la liste : 120 listes ;
4. Sur ces 120 listes nous avons récupéré les 2934 twittos ;
5. Nous avons travaillé par seuils, c'est-à-dire en ne retenant que les comptes présents dans plus de 12 listes : nous arrivons ainsi à 205 twittos politiciens.
6. Sur ces 205 comptes nous avons récupéré les 200 derniers tweets de chacun au 27 mars 2014, soit 34273 tweets.

Cette procédure nous permet d'avoir un ensemble de messages provenant de comptes « politiques », en garantissant que cette étiquette n'est pas projetée sur les données par le

chercheur, mais qu'elle est choisie par les usagers (à l'étape 3) par la dénomination de leur liste avec la chaîne *politic*. Ces comptes n'émanent pas forcément d'individus, mais parfois de partis, d'influenceurs (comptes crédités d'une autorité dans ce domaine, vérifiable par l'audience et les interactions en ligne notamment), ou de comptes satiriques (comme @humourdedroite).

Pour ce travail sur les mèmes, nous avons choisi de nous focaliser sur les tweets ayant été le plus « favorisés » par d'autres utilisateurs, mais aussi « retweetés », c'est-à-dire renvoyés/transférés à des followers (ceux qui suivent quelqu'un sur Twitter, en étant abonnés à ce compte) : pour cela les tweets ont été classés par ordre décroissant en fonction de la colonne « retweetcount » : les 600 premiers ont été sélectionnés, puis reclassés en fonction de la colonne « favoritecount ». Ceci permet ainsi de prendre prioritairement en compte le retweet comme élément caractéristique de la circulation des tweets, et de tenir également compte de l'interaction que les utilisateurs peuvent avoir avec ces messages en leur conférant le statut de « favori » :

tweet_text	screen_name	favoritecount	retweetcount
Toute ma gratitude va à lextraordinaire personnel ...	valtrier	2286	6337
Merci du fond du cœur à tous ceux qui ont envoyé d...	valtrier	1830	3429
"Soyons dignes, soyons patriotes, soyons Français ...	NicolasSarkozy	1632	5711
Avant dentrer à la présidence de la République, je...	fhollande	1401	6806
Que chacun d'entre vous sache combien je suis reco...	NicolasSarkozy	1350	2617
"Vous pourrez compter sur moi à chaque fois quil y...	NicolasSarkozy	614	1832
"Jamais je ne pourrai vous rendre tout ce que vous...	NicolasSarkozy	499	2499
@nicodomenach je pourrais aussi entrer en arrière ...	ramayade	464	2979
Joie daccueillir le père Georges, rayonnant de gén...	fhollande	446	1129
#FF @Elysee pour suivre toutes les informations su...	fhollande	427	993

La base de données des tweets : les deux critères de la viralité des messages

La viralité peut être définie par les trois paramètres résumés par Beauvisage, Beuscart, Couronné et Mellet (2011 : en ligne) : la concentration temporelle de l'attention sur un contenu ; la circulation de ces contenus, en distinguant ce qui relève de la contagion sociale au sens strict et ce qui relève des dispositifs éditoriaux ; les mécanismes de la contagion d'un individu à l'autre. Comme le notent les auteurs, « il existe bien des formes typiques de

focalisation temporelle de l'attention » mais « il est difficile, en revanche, d'identifier la viralité pure, c'est-à-dire la transmission directe d'un contenu entre deux individus, sans autre médiation ». Notre recherche s'attache alors à décrire ce qui peut être vu comme un « point aveugle » dans les recherches en communication numérique : la question de la nature des « contenus » diffusés.

L'objectif de la suite de cet article est donc de saisir les dimensions lexicales, puis textuelles, de ces structures qui circulent et d'établir si on peut leur accorder le statut de mème.

3 Spécificités lexicales des tweets efficaces et repérage des « mèmes » profilés

Pour appréhender ce corpus, nous avons utilisé le logiciel d'analyse de données textuelles Iramuteq, et pour commencer la fonction « analyse des similitudes ».

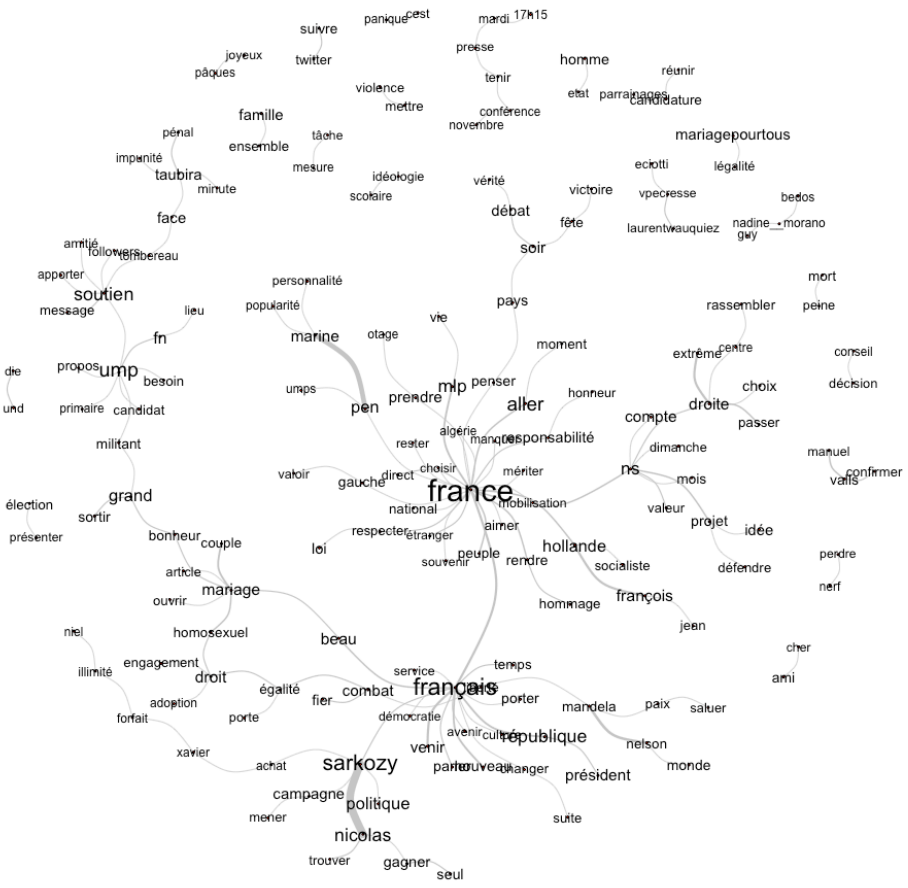
3.1 L'analyse des similitudes

L'analyse des similitudes est une analyse des cooccurrences présentée sous forme graphique. Elle permet de visualiser les réseaux qui constituent les textes, et les associations fréquentes. Pascal Marchand et Pierre Ratinaud (2012 : 688) expliquent ceci à propos de l'ADS :

> L'objectif de l'ADS est d'étudier la proximité et les relations entre les éléments d'un ensemble, sous forme d'arbres maximum : le nombre de liens entre deux items évoluant « comme le carré du nombre de sommets » (Flament & Rouquette, 2003 : 88), l'ADS cherche à réduire le nombre de ces liens pour aboutir à « un graphe connexe et sans cycle » (Degenne & Vergès, 1973 : 473).

D'un point de vue technique, Lucie Loubère et Pierre Ratinaud (2014 : 30) expliquent que cette analyse produit des graphes à partir de la librairie « igraph » de R. Le tableau en entrée est un tableau de présence / absence. La matrice de similitude est calculée à partir de l'un des indices proposés. La plupart des indices proposés sont issus de la librairie « proxy » de R.

Voici la représentation graphique établie à partir de notre échantillon des 600 tweets les plus efficaces du corpus[6], avec cette analyse :



Analyse des similitudes des tweets efficaces

Nous voyons que statistiquement, les termes (lemmes) qui reviennent le plus souvent sont notamment *France*, *français*, *Sarkozy* :

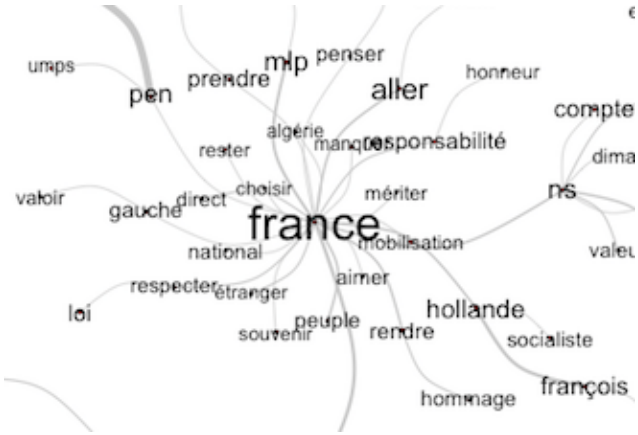
[6] Précisons que le corpus a été constitué de manière à considérer chaque tweet comme unité textuelle, avec un balisage effectué tweet par tweet. Ainsi, le corpus de tweets n'est pas un « sac de mots », mais respecte le statut de chaque élément qui le compose.

france	67	nr		mlp	22	nr
français	42	adj		politique	21	adj
sarkozy	36	nr		ns	21	nr
ump	31	nr		pen	19	nr
aller	27	ver		grand	19	adj
soutien	24	nom		beau	19	adj
république	23	nom		hollande	18	nom
nicolas	23	nr		françois	17	nr

Formes (lemmes[7]) les plus employées dans le sous-corpus de tweets efficaces

Mais l'intérêt de l'analyse des similitudes est aussi de faire ressortir le profilage de ces unités, par la mise en évidence des cooccurrences. Aussi, le graphique précédent fait apparaître les relations suivantes :

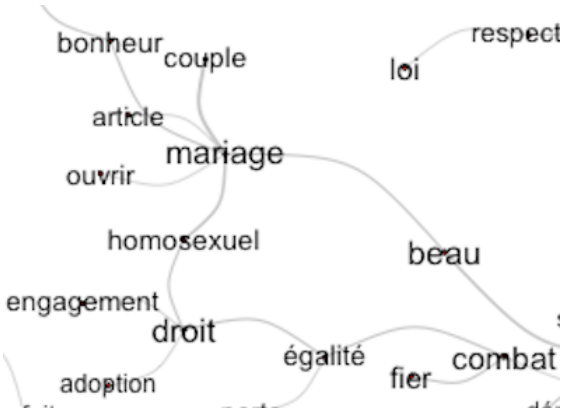
- le lemme *france*, qui est le plus employée, est connecté dans la représentation à des lemmes comme *mériter*, *mobilisation*, *responsabilité*, *choisir* ou *national* :



- le lemme *français* s'accompagne des lemmes *république*, *patrie* et *démocratie*, *service*, *combat* et *porter*, ou encore *temps*, *avenir* et *fier* :

[7] Nous avons opté pour une lemmatisation du corpus, aussi les formes dont il était précédemment questions sont dans cette étude équivalentes à des lemmes.

- autour de *mariage*, on retrouve *homosexuel*, *couple*, *article*, *ouvrir*, *engagement*, *adoption*, *égalité*, *droit*, etc. :



- le lemme *ump* agrège quant à lui les lemmes *primaire*, *candidat*, *militant*, *besoin* et *propos* :

- enfin, autour du lemme *sarkozy*, outre *nicolas*, on trouve *politique*, (*mener*) *campagne*, *gagner* (*seul*), *trouver*, et *achat* :

Ceci nous renseigne sur la manière dont les lemmes plus récurrents, pouvant donc prétendre au statut de mème, acquièrent une certaine viralité, par la manière dont ils sont profilées dans les tweets. Après ce premier repérage, nous pouvons avancer plus concrètement dans l'analyse lexicale.

3.2 Faire émerger des thématiques : la méthode Alceste et les tweets caractéristiques

Le logiciel Iramuteq « permet de décrire des classes lexicales, des profils de spécificités ou même des corpus entiers » (Marchand et Ratinaud, 2012 : 689). Elodie Baril et Bénédicte Garnier (2015 : 20) indiquent que « IRaMuTeQ fournit pour chaque classe des aides à l'interprétation qui permettent à l'utilisateur d'appréhender l'univers lexical de la classe et de lui donner un intitulé/thème ». En effet, pour chaque classe sont édités les profils, et les tables générées présentent les formes/lemmes caractéristiques des classes.

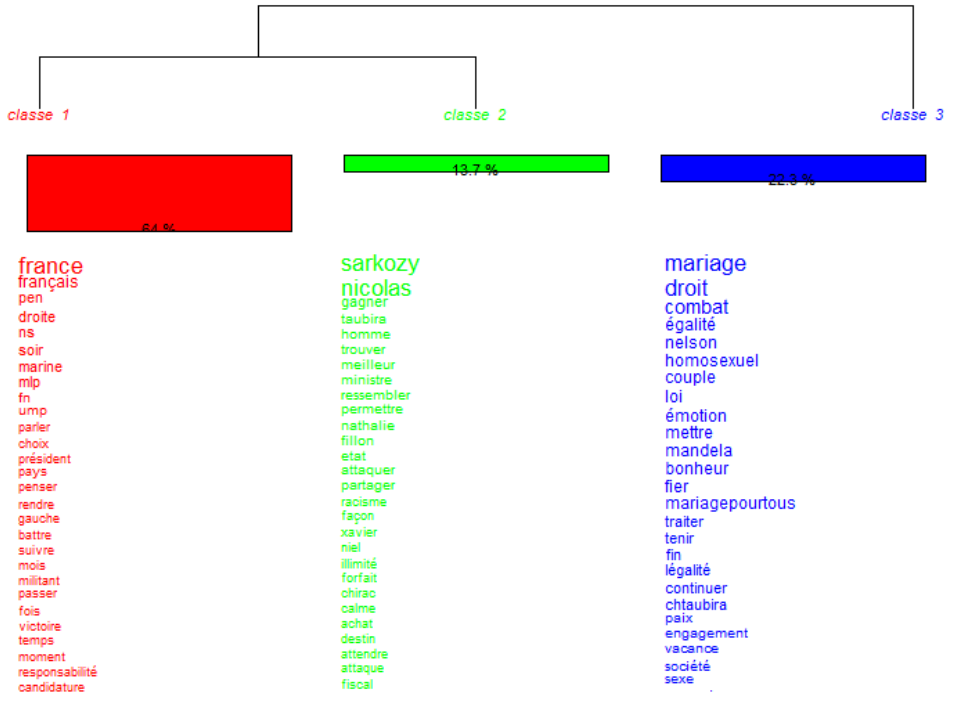
Dans la continuité de l'ADS, la méthode Aleste (développée par Max Reinert), se définit ainsi (Ratinaud et Marchand, 2012 : 837 notamment avec l'usage d'Iramuteq) :

> Une classification hiérarchique descendante : la technique de classification est surement la plus grande originalité de cette méthode. L'algorithme décrit par Reinert (1983) repose sur une série de bi-partitions construite sur la base d'une analyse factorielle des correspondances menée sur un tableau binaire (absence/présence) qui croise les unités textuelles choisies avec les formes pleines sélectionnées. C'est l'adaptation de cet algorithme que nous décrirons. Rappelons que chaque bi-partition est réalisée en trois étapes : 1) Une analyse factorielle des correspondances (A.F.C.) est menée sur le tableau, puis, pour toutes les partitions possibles le long du 1er facteur de l'AFC, l'inertie inter-classe est calculée. Une première coupure intervient pour la partition qui

maximise l'inertie inter-classe. 2) Chaque unité du tableau est permutée d'une classe à l'autre et l'inertie inter-classe est recalculée. Si celle-ci est supérieure à l'inertie inter-classe précédente, la permutation est conservée. Cette partie de l'algorithme boucle jusqu'à ce qu'aucune permutation n'augmente l'inertie inter-classe. 3) Les formes spécifiques d'une classe (au sens du chi2) sont retirées de l'autre.

Ceci permet de faire émerger les thématiques d'un corpus. Les résultats proposés par le logiciel doivent bien sûr être analysés en tant que points de vue sur les corpus, mais permettent de faire ressortir des rapprochements statistiquement mesurables[8].

Dans notre corpus, la méthode Alceste fait ressortir ces trois classes :



Les classes lexicales repérées à partir de la méthode Alceste

Les éléments connaissant une certaine viralité seraient donc ceux qui entrent dans les thématiques suivantes : 1) France/français, 2) sarkozy, 3) mariage/combat/droit[9].

---

[8] Alceste, *à partir d'un corpus, effectue une première analyse détaillée de son vocabulaire, et constitue le dictionnaire des mots ainsi que de leur racine, avec leur fréquence. Ensuite, par fractionnements successifs, il découpe le texte en segments homogènes contenant un nombre suffisant de mots, et procède alors à une classification de ces segments en repérant les oppositions les plus fortes. Cette méthode permet d'extraire des classes de sens, constituées par les mots et les phrases les plus significatifs, les classes obtenues représentent les idées et les thèmes dominants du corpus.* (https://fr.wikipedia.org/wiki/Alceste_(logiciel))

Afin de retourner aux données sur la base de cette hypothèse, nous pouvons utiliser l'outil « Segments de textes caractéristiques » (ST caractéristiques). Un Texte se définit comme suit : « l'unité de base est appelée « Texte ». Un texte peut représenter un entretien, un article, un livre ou tout autre type de documents. Un corpus peut contenir un ou plusieurs textes (mais au minimum un) » (Loubère et Ratinaud 2014 : 4). Nous utilisons la classification simple sur texte, dans laquelle les textes sont conservés dans leur intégralité, ce qui permettra à la classification de regrouper les Textes les plus proches :

> Classification simple sur segment de texte : les **Textes** dans leur intégralité, la classification permet ainsi de regrouper les **Textes** les plus proches.
> Taille rst1 et rst2 (regroupement de segments de textes): 12 et 14 (ces unités représentent le nombre de formes actives regroupées par ligne).
> Liste les ST caractéristiques de la classe. Score absolu: les ST sont classées en fonction de la somme des chi2 de liaison à la classe des formes actives qu'ils contiennent.

Concrètement (Loubère et Ratinaud, 2014 : 8), « les segments de texte sont construits à partir d'un critère de taille et de ponctuation. Iramuteq cherche le meilleur ratio taille/ponctuation (par ordre de priorité, les ".", "?" "!" en premier, puis en second ";" et les ":" en troisième la virgule et en dernier l'espace). L'objectif est d'avoir des segments de tailles homogènes en respectant le plus possible la structure du langage ».

Nous pouvons donc recenser les segments (tweets) les plus caractéristiques de chaque classe, les termes en gras indiquant les formes caractéristiques de ces classes[10].

---

[9] Ceci indique que le contexte joue bien sûr un rôle. Néanmoins, tous les événements ou actualités n'ont pas le même statut, et ne deviennent pas des sujets populaires. Aussi, comme indiqué à propos du concept de mème chez Sperber, nous considérons que le matériau langagier contribue à faire émerger une thématique forte et une actualité. Nous ne pouvons développer dans l'espace de cet article, mais notons simplement que le rapport entre l'actualité d'un thème, et sa viralité, est à questionner et ne va pas de soi.

[10] Les * correspondent aux balises dans Iramuteq : la « force » correspond à la viralité (force 1 pour les 200 tweets les plus efficaces, 2 pour les 200 suivants, et 3 pour les 200 suivants, soit 600 tweets), selon le tri effectué. Ces forces sont distinctes des scores obtenus dans chacune des classes, et nous privilégions, pour la présentation des résultats, l'ordre selon un score décroissant. Notons aussi que pour ces exemples, Iramuteq propose un « nettoyage » des données, qui enlève notamment les signes tels que le # ou @. Pour cette phase de l'analyse, cela n'est pas problématique.

[1] Classe 1 France/français

a) ** *force_3 *tweet_27**

score : 54.63

ceci n est pas ma **rentrée** politique le **jour** où je **reprendrai** la parole ce sera pour **parler** aux **français** de la **France**

b) ** *force_1 *tweet_41**

score : 53.37

je **penserai** à vous dimanche à nous à notre **pays** nous serons au **rendez_vous** de la **france** forte au **rendez_vous** de la **victoire ns**

c) ** *force_2 *tweet_110**

score : 51.67

je **rends** possible l union nationale et la **vraie mobilisation** des **français** au service de la **France**

[2] Classe 2 sarkozy

a) ** *force_1 *tweet_186**

score : 407.05

comme je l ai toujours dit ceux qui **attaquent** le bilan que **nicolas sarkozy** et **francois fillon** ont en **partage** me **trouveront** sur leur **chemin**

b) ** *force_2 *tweet_167**

score : 392.63

d8tv campagneintime on nous avait menti sur **nicolas sarkozy** lhomme est **calme** et serein **face** au **destin** la définition de l **homme** d **etat**

c) ** *force_3 *tweet_10**

score : 386.68

xavier niel vient de faire plus avec son **forfait illimité** pour le pouvoir d **achat** des français que **nicolas sarkozy** en 5 ans christophegen

d) ** *force_2 *tweet_117**

score : 367.23

la **campagne** se **termine** à 00h maintenant c est vous qui êtes les **meilleurs** ambassadeurs pour faire **gagner nicolas sarkozy** dimanche

e) ** *force_2 *tweet_69**

score : 345.70

si **nicolas sarkozy** veut **trouver** les exilés **fiscaux** ne devrait il pas commencer par **consulter** son carnet d **adresses**

f) ** *force_3 *tweet_198**

score : 344.29

nicolas sarkozy est et de loin le mieux placé pour faire **gagner** nos idées en 2017 **préparons ensemble** et maintenant un projet courageux

g) ** *force_3 *tweet_87**

score : 311.34

avec **nicolas sarkozy** nous sommes convenus **ensemble** que **face** au péril il appartenait à notre **famille politique** de faire bloc

[3] Classe 3 mariage/combat/droit

a) ** *force_1 *tweet_39**

score : 211.51

je suis **fier** d être le pm qui **porte** le projet **mariage** pour tous et **adoption** c est un **combat** juste pour l **égalité** des **droits** 1/2

b) ** *force_2 *tweet_68**

score : 209.25

fin du 15 août réaffirmation de mon **engagement** pour le **droit** au **mariage** et à l **adoption** pour les **couples homosexuels** pas négociable

c) ** *force_2 *tweet_19**

score : 196.56

je **salue** avec **émotion** la **mémoire** de **nelson mandela citoyen** du monde figure universelle du **combat** pour l **égalité** la **liberté** et la **paix**

d) ** *force_2 *tweet_123**

score : 156.52

vote solennel je suis **fier** de cette **majorité** qui **porte** la **loi** sur le **mariage** pour tous **belle** et **grande** avancée pour l **égalité**

e) ** *force_1 *tweet_50**

score : 152.10

un **grand** pas sur le **chemin** de **légalité** le **mariage** est désormais **ouvert** aux **couples** de même **sexe** tous mes voeux de **bonheur**

f) ** *force_1 *tweet_176**

score : 149.70

obama notre **combat** pour l **égalité** ne sera pas terminé tant que nos frères et soeurs **homosexuels** n auront pas l **égalité** des **droits**

g) ** *force_1 *tweet_15**

score : 144.92

ça y est j ai débouché le champagne rosé 25 ans de **combat** ça le valait bien quel **bonheur** et bientôt les **mariages mariagepourtous**

h) ** *force_2 *tweet_1**

score : 142.59

le sénat vient d adopter l **article** qui **ouvre** le **mariage** aux **couples homosexuels belle nuit** à tous toutes lemariageprtous

Nous allons à présent considérer la manière dont ces résultats nous permettent de répondre aux interrogations initiales sur la question des mèmes dans les tweets politiques.

3.3 De la forme comme mème linguistique au tweet comme mème textuel

Les analyses statistiques, de cooccurrences, et les tweets emblématiques, nous renseignent sur l'efficacité de certaines tournures/formes, et donc de leur viralité (au regard de la définition de la viralité présentée précédemment). Ces résultats mettent en avant certaines formes (plus fréquentes et statistiquement caractéristiques de thématiques) mais surtout pointent selon nous la nécessité de considérer finalement le tweet comme unité circulante, et donc créditée en tant qu'objet textuel d'une certaines viralité. Il faut donc considérer la contribution conjointe de différents éléments linguistiques (des formes statistiquement significatives), et la manière dont ces éléments interagissent et contribuent à profiler les termes récurrents vers une certaine appréhension spécifique des thématiques : le rapport du politique aux *français*, l'insertion du lemme *sarkozy* dans des tweets liés au contexte politique de son parti et aussi national, et la question du *mariage* dans le contexte du mariage pour tous. A titre d'exemple, nous avons recherché les segments répétés avec la fonction correspondante dans *Lexico 3*, et nous avons pu identifier les structures « de la France » et « en France » comme étant spécifiques :

```
arler aux Français de la France . " Ma mission , r
lence avec un amour de la France encore plus grand
ider à la destinée de la France fut un honneur imr
ons au rendez-vous de la France forte, au rendez -
2013 , le sauveur de la France s'appelle Mamadou
ire , à la hauteur de la France , de son Histoire
quand FH2012 parle de la France En route pour DPDA
et à notre vision de la France . Débat ouvert Ceu
rançais au service de la France Pour comprendre pc
ié le gouvernement de la France . Je mesure sa cor
choisir le destin de la France ! 22avril FH2012 :
```

Segments « de la France »

En retournant au corpus, nous voyons que lorsque *France* est employée comme complément du nom, ceci se fait dans une structure axiologiquement positive, comme avec *amour de la France*, *sauveur de la France*, *hauteur de la France, destinée de la France* ou *service de la France*. A chaque fois, *amour*, *sauveur*, *hauteur*, ou *service*, profilent *France* en l'orientant vers un point de vue positif et donc une vision favorable.

Ceci est très différent dans la structure « en France » :

```
'bengallerey : Restez en France ! ( Vu ligne 9 ) : 
rappes : La seule loi en France est la loi français
obligation à rester en France . Pourquoi mon choi
mieux les musulmans en France que les chrétiens n
génération sont nés en France donc pleinement fra
rs La carte des noirs en France : Elles sont pas be
va être la révolution en France la CEJ donne l'auto
étrangers qui vivent en France subviennent à leurs
priorité au logement en France est délirant PP3tv 
. Vivement la gauche en France ! Leonarda A lire a
```

Segments « en France »

Ces structures questionnent l'appartenance de certains Français à la France, comme dans *noirs en France* ou *musulmans en France*, ou rapportent des éléments polémiques qui faisaient débat au moment de l'envoi du tweet (*gauche en France* dans le cadre de l'affaire Leonarda). Bien sûr, cette première analyse devra être approfondie par un travail systématique autour du lemme *France*, mais ces deux rapides analyses indiquent qu'en fonction de la nature du profilage (comme complément du nom qui oriente vers une axiologie positive, ou comme localisation qui oriente vers une axiologie polémique ou négative), le tweet qui comporte ces éléments connaîtra des suites différentes en fonction des communautés de discours dans lesquelles il s'intégrera.

Ceci permet de reprendre la question du rapport entre les fonds et les formes sémantiques, et de le relier concrètement à la question des mèmes : les formes constitutives des tweets peuvent acquérir le statut de mème linguistique en considérant leur fréquence et leur capacité à agréger d'autres formes autour d'elles. Ces interactions avec d'autres formes contribuent à conférer aux tweets (grâce à leur brièveté notamment) le statut de mème textuel. En fonction des rapports entre les différentes formes, et notamment selon les types de profilages produits (voir l'exemple *de la France*/ *en France*), ces tweets acquièrent une certaine viralité relativement au fond sémantique sur lequel ils se déploient (tel tweet deviendra viral au regard du « patriotisme » qu'il véhicule, tel autre pour son « universalisme », etc.).

Conclusion

Finalement, ces résultats interrogent le concept de mème : plus que l'existence de mèmes linguistiques dont le caractère efficace, figé, circulant, serait la clé d'analyse de la problématique posée, nous pourrions avancer qu'il y a peut-être des « degrés » dans l'identification de structures linguistiques qui s'apparenteraient aux mèmes. En ce qui

concerne notre objet d'étude, c'est plutôt le tweet qui apparaît, par la mise en discours et le profilage de certaines formes, comme un mème, en tant qu'unité textuelle saillante et stabilisée par la récurrence de certaines formes.

Dans la tentative que nous avons menée de relier la question des mèmes linguistiques avec l'analyse d'un corpus de tweets politiques, nous avons dû prendre en considération certaines spécificités de notre cadre méthodologique, et de l'objet d'étude. La spécificité de Twitter, et la méthodologie employée[11], ont probablement complexifié la caractérisation de cet objet. Mais ce questionnement a permis d'interroger le corpus d'une manière originale, et surtout il nous a amené à formuler l'hypothèse de l'existence, pour le tweet politique, d'une forme de mème plus textuelle que langagière, puisque c'est, tout autant que l'emploi de certains termes, leur mise en syntagme particulière et selon certains points de vue, qui en fait le succès et donc la viralité. La spécificité de ces mèmes tiendrait à l'adéquation d'un jeu des formes sémantiques sur les fonds mis en discours, et sur la récurrence, la stabilisation, et le profilage, de certaines formes. Concrètement en effet, c'est par la récurrence d'une forme dans plusieurs unités textuelles (tweets), par l'association privilégiée de cette forme avec d'autres, et donc par la constitution de thématiques, que se propagent des éléments qui peuvent acquérir le statut de mème, tel que ce concept est proposé par Dan Sperber. C'est donc à l'échelle du tweet que se résoudrait la question de la viralité des mèmes, par la prise en compte de la circulation et de l'efficacité, d'un tweet politique. Mais c'est aussi au regard de l'ensemble du corpus que se détermine cette viralité, les dimensions locales (tweet) et globales (discours accessibles par les corpus) se co-déterminant réciproquement. Notamment, le profilage des termes saillants, et leur orientation vers certaines thématiques ou topiques (comme la valorisation de la France, ou le caractère polémique de l'accès à la France) contribuent à rendre les tweets efficaces. Bien sûr, ces analyses lexicales et textuelles pourront dans la suite des recherches être reliées aux instances de production des messages, comme cela est davantage fait dans les recherches en communication numérique notamment.

[11] Nous n'avons pas abordé l'analyse pas un terme ou un syntagme préalablement identifiée comme ayant connu un certain succès, mais directement par le corpus, selon une approche « émergentiste » des aspects saillants. Soit préciser, soit enlever. Le lecteur ne comprend pas, sinon.

Références bibliographiques

BARIL E. et GARNIER B., 2015, *Utilisation d'un outil de statistiques textuelles*. URL : http://www.iramuteq.org/documentation/fichiers/Pas%20a%20Pas%20IRAMUTEQ_0.7alpha2.pdf.

BEAUVISAGE T., BEUSCART J.S., COURONNE T et MELLET K., 2011, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », *Tracés. Revue de Sciences humaines* [En ligne], 21 URL : http://traces.revues.org/5194.

CADIOT P. et VISETTI Y.-M., 2001, *Pour une théorie des formes sémantiques*, Paris, PUF, coll. « Formes sémiotiques ».

DJEMILI S., LONGHI J., MARINICA C., KOTZINOS D. et SARFATI G.-E., 2014, « What does Twitter have to say about ideology ? », Konvens 2014 - Workshop proceedings vol. 1 (*NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication / Social Media – Pre-conference workshop at Konvens2014)*, Germany (2014), p. 16-25.

GARRIC, N. et LONGHI J., 2013, « Atteindre l'interdiscours par la circulation des discours et du sens », *Langage et société*, 144, p.65-83.

LEBAS, F. et CADIOT, P., 2003, « La constitution extrinsèque du réfèrent : présentation », *Langages*, 37, p.3-8.

LONGHI, J., MARINICA, C., BORZIC, B. et ALKHOULI, A., 2014, *Polititweets, corpus de tweets provenant de comptes politiques influents*. In CHANIER T. (éd.) Banque de corpus CoMeRe. Ortolang.fr : Nancy. [cmr-polititweets- tei-v1]. Permalien : https://repository.ortolang.fr/api/content/comere/v3.1/cmr-polititweets.html.

LONGHI J., 2011, *Visées discursives et dynamiques du sens commun*, Paris, L'Harmattan.

LONGHI J., 2013, « Essai de caractérisation du tweet politique », *L'Information grammaticale*, 136, p. 25-32.

LONGHI J., 2014, « L'hybridation du discours institutionnel à l'épreuve du numérique : renouvellement et reconfiguration de la parole institutionnelle », in LONGHI J. et SARFATI G.-E. (éd.), *Les discours institutionnels en confrontation. Contribution à l'analyse des discours institutionnels et politiques*, Paris, L'Harmattan, p. 167-188.

LONGHI J., 2015, La théorie des objets discursifs : concepts, méthodes, contribution, Mémoire d'HDR, Université de Cergy-Pontoise.

MARCHAND P. et RATINAUD P., 2012, « L'analyse de similitude appliquée aux corpus textuels : les primaires socialistes pour l'élection présidentielle française (septembre-octobre

2011) », *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles - JADT 2012*, Liège, p. 687-699.

LOUBERE L. et RATINAUD P., 2014, *Documentation IRaMuTeQ 0.6 alpha 3 version 0.1*. URL : http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf.

RASTIER F., 2006, « Formes sémantiques et textualité », *Langages*, 163, p. 99-114.

RASTIER F., 2011, *La Mesure et le grain. Sémantique de corpus*, Paris, Honoré Champion.

RATINAUD P. et MARCHAND P., 2012, « Application de la méthode ALCESTE aux « gros » corpus et stabilité des « mondes lexicaux » : analyse du « CableGate » avec IRAMUTEQ », *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles - JADT 2012*, Liège, p. 835-844.

ROSENTHAL V. et VISETTI Y.-M., 2003, *Köhler*, Paris, Les Belles Lettres.

SARFATI G.-E., 1996, *La sémantique : de l'énonciation au sens commun. Éléments d'une pragmatique topique*. *Texto !* décembre 2004. URL : http://www.revue-texto.net/Inedits/Sarfati/Sarfati_Semantique.html.

SARFATI G.-E., 2014, « L'emprise du sens : Note sur les conditions théoriques et les enjeux de l'analyse du discours institutionnel », in LONGHI J. et SARFATI G.-E. (éds.), *Les discours institutionnels en confrontation, Contribution à l'analyse des discours institutionnels et politiques*, Paris, L'Harmattan, Collection Espaces Discursifs, p. 13–46.

SPERBER D., 1996, *La contagion des idées*, Paris, Odile Jacob.

TLFi : Trésor de la langue française informatisé.