



HAL
open science

Entre distinctivité et acceptabilité: les contenus des sites Web de partis politiques

Pascal Marchand, Pierre Ratinaud

► To cite this version:

Pascal Marchand, Pierre Ratinaud. Entre distinctivité et acceptabilité: les contenus des sites Web de partis politiques. Réseaux: communication, technologie, société, 2017, n° 204 (4), pp.71-95. 10.3917/res.204.0071 . halshs-02161211

HAL Id: halshs-02161211

<https://shs.hal.science/halshs-02161211>

Submitted on 20 Jun 2019

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.

ENTRE DISTINCTIVITÉ ET ACCEPTABILITÉ

Les contenus des sites Web de partis politiques

Pascal Marchand et Pierre Ratinaud

La Découverte | « Réseaux »

2017/4 n° 204 | pages 71 à 95

ISSN 0751-7971

ISBN 9782707197177

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-reseaux-2017-4-page-71.htm>

Distribution électronique Cairn.info pour La Découverte.

© La Découverte. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

ENTRE DISTINCTIVITÉ ET ACCEPTABILITÉ

Les contenus des sites Web de partis politiques ¹

Pascal MARCHAND
Pierre RATINAUD

1. Ce travail a été réalisé dans le cadre du LABEX SMS portant la référence ANR-11-LABX-0066.

Étudier les sites Web des partis politiques ne se résume pas à en analyser les contenus et programmes ; c'est pénétrer un univers de débats. Le propre du débat est d'introduire un échange conflictuel et d'ouvrir un espace de comparaison : chacun doit marquer sa différence à l'égard de ses adversaires pour ne pas risquer d'être perçu comme partageant avec lui des positions (Papastamou, 1993). La légitimité d'un parti politique repose donc sur sa capacité à générer et alimenter le conflit politique, quitte à développer une posture polémique (Marchand et Ratinaud, 2014), par ailleurs recherchée et exploitée par les dispositifs médiatiques. Le concept de « distinctivité », développé par Henri Tajfel (1974), marque cette motivation pour les membres engagés dans un groupe à se différencier des autres groupes : « À la limite, les membres d'un groupe n'ont rien d'autre en commun que le fait de ne pas partager les caractéristiques du hors-groupe. [...] L'identité sociale repose [alors] carrément sur les épaules du hors-groupe » (Allen *et al.*, 1983 cité par Lorenzi-Cioldi et Dafflon, 1999, pp. 136-137). Pour reprendre une déclaration fameuse de Tajfel, « nous sommes ce que nous sommes parce qu'ils ne sont pas ce que nous sommes » (Tajfel, 1979, p. 183).

Dans le même temps, une différenciation poussée à l'extrême représente un risque d'incomparabilité vis-à-vis de tous les autres, et par conséquent de disqualification possible dans la compétition électorale. Il faut donc être acceptable dans le jeu politique, et il s'agit de « convaincre de ses similitudes tout en démontrant ses différences » (Mange et Marchand, 2011, p. 87).

Les sites, qui reprennent les interventions dans les meetings ou les médias plus traditionnels, participent à dessiner les « territoires informationnels spécifiques » (Marty *et al.*, 2012) sur lesquels s'élaborent les stratégies discursives qui devront mobiliser à la fois des lieux communs, permettant l'identification d'un discours politique acceptable dans le contexte donné, et des spécificités, permettant l'identification des appartenances politiques et des positionnements partisans.

Les discours vont donc se construire dans une tension entre les attitudes paradigmatique et syntagmatique (Marchand, 2010), autrement dit entre

des stratégies référentielle (« parler de quelque chose ») et socio-régulative (« parler à quelqu'un ») qui vont amener à mobiliser des marqueurs socio-langagiers (Scherer et Giles, 1979).

On peut distinguer différents types ou niveaux de marqueurs, renvoyant par exemple davantage au positionnement sur l'objet (marqueurs plus lexicaux) ou au statut social (marqueurs plus syntaxiques). Nous nous intéressons ici aux marqueurs lexicaux. Ces derniers peuvent parfois sembler peu structurés et stables (Laver et Trudgill, 1979), en tout cas moins que la syntaxe, voire même la phonologie ou la gestuelle (*les mots mentent, la voix et les gestes, non*). Mais lorsqu'on entend dire que « les mots ont une importance » et lorsqu'on est prêt à débattre sur leur usage, on reconnaît bien qu'il y a une structure régulière du lexique. Mais cette régularité ne tient qu'à la valeur sociale des mots, tributaire à la fois de l'histoire, qui a figé des formules (Krieg-Planque, 2009), et du contexte, qui impose des usages.

On peut donc s'intéresser au lexique sans pour autant recourir à une sémantique interprétative, mais en posant que la distinctivité linguistique (ou psycholinguistique pour Bourhis et Giles, 1977) investit le choix des mots (voir aussi Brennan et Clark, 1996). Jessica Mange (2016) fait une revue de question particulièrement pertinente sur ces processus, à l'intersection de la psychologie sociale et des sciences de la communication. On retiendra, par exemple, que l'emploi différencié des termes « immigrants illégaux », « demandeurs d'asile » et « réfugiés » dans les médias australiens produit des effets distincts sur les attitudes des récepteurs à l'égard des groupes de migrants (Augoustinos et Quinn, 2003). En France, Rodolphe Ghiglione (1985) avait déjà montré que, pour parler de l'avortement, le choix d'un déterminant défini (« le ») ou indéfini (« un ») et d'un vocabulaire spécifique (« bébé » vs « fœtus ») dépend de l'attitude du locuteur envers l'objet (« défavorable » vs « favorable ») et génère des réponses attitudeles (Mikolajczak et Bilewicz, 2014).

De tels effets ont été observés sur des thématiques aussi différentes que le SIDA (Mange *et al.*, 2004), la seconde guerre du Golfe (Mange, Lepastourel et Georget, 2009), l'homosexualité masculine (Mange et Lepastourel, 2013) ou encore le « mariage pour tous » (Sanrey, Testé et Mange, 2016).

On rejoint donc Ghiglione (1985) lorsqu'il posait que « le proposant, porteur d'une attitude préalable, inscrit dans son énoncé des traces de cette attitude (plus ou moins volontairement...), et ce faisant inscrit l'interprétant dans un

jeu d'influence » (p. 243). Il avait d'ailleurs montré que les mêmes textes étaient mieux reconnus par des militants s'ils respectaient les marquages morpho-syntaxiques des personnalités politiques que s'ils ne les respectaient pas (Ghiglione, 1989).

La « stabilisation » du lexique est alors renvoyée à celle de la structure sociale : l'organisation normative des mots permet d'investir des lexiques pour montrer notre appartenance groupale dans un contexte qui rend saillante et importante la catégorisation sociale (Ratinaud et Marchand, 2015).

Max Reinert (1993) évoquait la notion de « mondes lexicaux » pour établir un lien statistique entre un texte et ses conditions de production. Le tableau lexical qu'il construit croise des énoncés, qui renvoient au « point de vue » du sujet énonciateur, et des lexèmes, qui renvoient à l'objet référentiel. Un « monde lexical » était alors défini comme la trace d'un lieu référentiel et l'indice d'une forme de cohérence liée à l'activité spécifique du sujet énonciateur. Il s'agit donc bien de s'inscrire dans une approche multidimensionnelle de la co(n)textualité telle que la définit Mayaffre (2014).

Nous reprendrons ici la méthode de Reinert (1983), telle qu'elle a été implémentée dans le logiciel libre d'analyse textométrique *Iramuteq* (Ratinaud, 2014), pour tenter d'observer comment les sites des partis politiques investissent le lexique pour faire jouer les similarités et distinctivités qui définissent leur territoire sur la carte politique à la veille d'échéances électorales.

DES SITES AUX LEXIQUES

Constitution des corpus

Les corpus que nous nous proposons d'analyser sont issus des sites Web des partis politiques français. Il est tout d'abord à noter qu'en 2015, lors du recueil des données, l'UMP a changé de nom pour devenir « Les Républicains ». Ce changement a eu pour conséquence une refonte du site Web du parti et la disparition des archives du site précédent. Nous n'avons donc pas pu introduire ce parti dans notre échantillon. Le corpus se limitera donc aux sites Web de trois partis politiques : le *Parti socialiste* (PS), *Europe Écologie Les Verts* (EELV) et le *Front National* (FN)².

2. Les données ont été extraites à l'aide de l'outil *Gromoteur* (Gerdes, 2014).

De façon à homogénéiser les corpus, nous avons fait le choix de ne retenir que les documents relevant de la catégorie « actualités ». Cette catégorie, commune aux trois sites Web, est la seule qui présente ici un intérêt. Toutes les autres entrées proposées sur ces sites conduisent à des formulaires d'adhésion, à des descriptions formelles de l'organisation interne des partis ou à des outils « participatifs ». Pour le *Front National*, nous avons également retenu la catégorie « projet ». Le tableau 1 décrit succinctement les trois corpus concernés.

Tableau 1. Description des trois corpus³

	FN	PS	EELV
textes	3 267	5 419	2 405
segments de texte	24 945	41 328	25 948
occurrences	897 173	1 487 840	924 289
nombre de formes	34 383	33 255	33 479
hapax ³	13 728	11 673	13 066
nombre de classes	14	14	15
dates	07/2011-05/2015	12/2006-05/2015	02/2011-10/2015

Source : auteurs.

À ce stade, les stratégies d'analyse et de constitution du corpus peuvent être multiples. L'homogénéité de genre et la temporalité commune des textes sont ici un argument qui irait dans le sens de la constitution d'un seul corpus regroupant les trois sites Web et qui pourrait être soumis à une classification. Nous opérerons toutefois un choix différent et le traitement retenu fera plutôt intervenir trois corpus : un pour chaque site. Ce choix est légitimé à nos yeux par deux dimensions. D'une part, l'homogénéité des textes est ici toute relative. Le « nettoyage » des pages Web proposé par *Gromoteur* ne permet pas toujours d'extraire de ces pages uniquement le texte que le chercheur veut analyser. Il reste souvent des traces du formatage particulier des pages sur chacun des sites. Ces traces, récurrentes sur les pages de chaque site, sont autant de « bruits » qui peuvent potentiellement orienter l'analyse automatique. Par ailleurs, l'objectif même de communication assigné à ces outils conduit à noter une forte présence de noms propres dans les textes. Nous verrons que la fréquence d'apparition des patronymes des acteurs des différents partis, notamment dans les phases électorales, mène systématiquement à la constitution d'une ou plusieurs classes marquées par ces noms. D'autre part, l'objectif

3. Les hapax sont les mots qui n'apparaissent qu'une seule fois dans un corpus.

de notre analyse est de comparer les thématiques abordées par chacun de ces partis. Une classification menée sur l'ensemble des textes sera alors moins précise que la comparaison de classification sur chacun des textes.

La démarche adoptée pour générer ces données rend les corpus tout à fait comparables, même si le site Web du PS couvre presque deux quinquennats alors que les deux autres (EELV et FN) ne proposent des documents que de 2011 à 2015.

Stratégie d'analyse

Une comparaison des discours produits sur les sites Web des partis politiques avec les débats à l'Assemblée nationale a déjà été proposée (Ratinaud et Marchand, 2016), et nous reprendrons ici la stratégie utilisée pour comparer les partis politiques entre eux. La première étape est une classification descendante hiérarchique (CDH), telle qu'elle a été définie par Max Reinert (Reinert, 1983 ; Ratinaud et Marchand, 2012), et qui sera menée sur chacun des trois sites. Cette étape doit mettre en évidence les thématiques abordées dans les corpus.

L'analyse utilisée comprend trois phases. Dans un premier temps, les textes sont découpés en segments d'une quarantaine de mots. Ce découpage tient compte de la ponctuation et essaie donc de conserver la division naturelle des textes. Dans un deuxième temps, le lexique est *lemmatisé* (les verbes sont mis à l'infinitif, les adjectifs au masculin singulier et les substantifs au singulier) et une distinction est faite entre les mots dits « pleins » (les verbes, les adverbes, les substantifs et les adjectifs) et les mots outils (pronoms, connecteurs...). Seuls les mots pleins participeront à l'analyse. La troisième phase est la classification proprement dite.

L'originalité de l'approche présente tient à ce que nous proposerons de comparer plusieurs classifications, calculées sur des corpus indépendants, et nous adopterons deux approches différentes : une analyse factorielle des correspondances sur le tableau lexical regroupant l'ensemble des classes et une analyse en termes de distance lexicale.

CLASSIFICATIONS ET DISTRIBUTIONS LEXICALES

Notre première approche est très classique, mais rarement utilisée pour comparer des résultats de classifications indépendantes : il s'agit d'une analyse factorielle des correspondances (AFC) sur le tableau lexical entier (Lebart et

Salem, 1994) regroupant les trois classifications. Il faut donc préalablement appliquer la classification (CHD) aux trois sites Web.

Les classifications lexicales des trois sites Web (PS, EELV, FN)

Cette étape est indispensable à notre perspective comparative, mais ne pourra pas être approfondie ici.

Il s'agit d'une classification hiérarchique descendante (CHD) conçue comme une série de bipartitions (découpage d'une matrice en deux classes) sur la base d'une analyse factorielle des correspondances (AFC). Autrement dit, la matrice croise le lexique (lemmatisé et étiqueté) et les segments de textes. Les unités classées par cette analyse sont les segments de textes et ils sont donc réunis dans des classes à partir d'un critère de cooccurrence de mots pleins. Les classes obtenues sont des regroupements de segments de texte qui ont tendance à contenir le même vocabulaire⁴.

Les résultats se présentent sous la forme de classes de discours qui sont décrites à partir du lexique qui est statistiquement surreprésenté dans les segments réunis dans les classes.

Chacune des classes lexicales pourrait faire l'objet d'une description fine par le lexique qui la définit, les segments caractéristiques qui la représentent⁵ et les variables extratextuelles (métadonnées) qui l'expliquent. Nous n'en livrerons que les grandes lignes pour faciliter la compréhension des étapes qui suivront.

Analyse du site Web du Parti socialiste

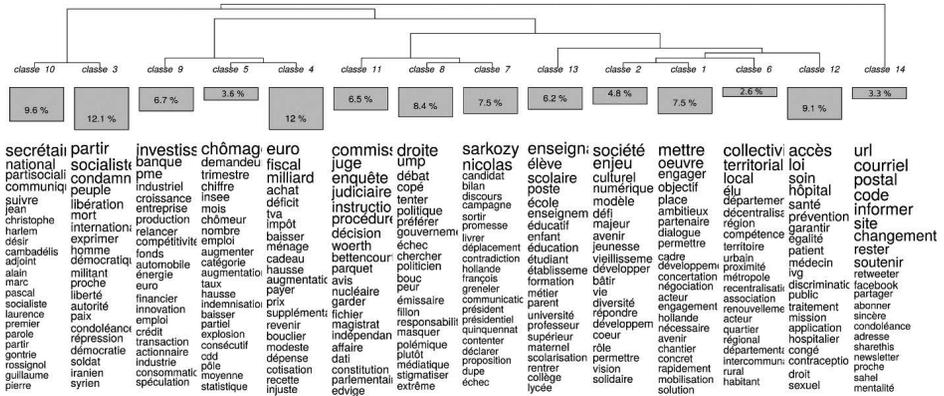
Le dendrogramme⁶ 1 rend compte des résultats de la classification (CDH) menée sur le corpus du *Parti socialiste* :

4. Autrement dit, la matrice présente des suites de 0 ou de 1 selon qu'une forme lexicale est présente ou absente d'un segment de texte. Cette matrice laisse donc apparaître des profils (suites de 0/1) pour les formes lexicales comme pour les segments de texte. On regroupe les profils semblables et on oppose les profils différents.

5. Les segments de texte caractéristiques sont les segments qui maximisent la somme des χ^2 de liaison des formes pleines du profil. Ce sont les segments les plus « lourds » de la classe et donc les plus représentatifs du point de vue statistique.

6. Un dendrogramme est un diagramme qui permet de rendre compte de l'enchaînement des découpages dans les classifications hiérarchiques.

Figure 1. Dendrogramme, taille des classes et lexique caractéristique des classes (par χ^2 décroissant) pour le site du Parti socialiste



Source : auteurs.

La première classe qui se distingue dans cette analyse est la classe 14, à droite sur le dendrogramme. Elle est caractérisée par un lexique « technique » assez typique des corpus extraits de pages Web. Il s’agit des mots de l’Internet et des réseaux socio-numériques que l’on retrouve sur la plupart des sites. C’est le signe manifeste que le nettoyage des textes n’est pas parfait. Cette classe sera éliminée dans les traitements suivants. Les classes 10 et 3 forment la seconde branche de l’arbre (à gauche du dendrogramme). La classe 10 regroupe des segments de texte marqués par la présence des noms des acteurs du parti. Les entêtes des communiqués de presse sont caractéristiques de cette classe, mais on trouvera également des textes relatifs à des résultats électoraux :

« **communiqué de presse du secrétariat national à l international suivre parti-socialiste le premier secrétaire du Parti socialiste harlem désir** ainsi que le **secrétaire national à l international et à l europe jean christophe cambadélis** » (segment de texte le plus caractéristique de la classe 10)

La classe 3 est marquée par un lexique qui semble renvoyer aux hommages et aux condamnations liés à des événements internationaux.

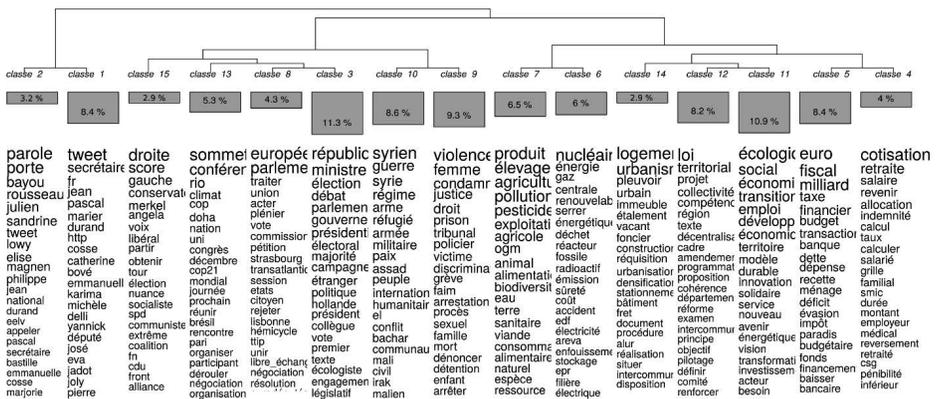
En poursuivant la lecture du dendrogramme vers la droite, nous trouvons trois classes relevant de la dimension économique. La première (classe 9) est

7. Le χ^2 est un test statistique qui permet de mesurer un lien de dépendance ou d’indépendance entre les modalités de deux variables nominales.

composée du lexique de la finance et de la production industrielle, la seconde (classe 5) traite spécifiquement de la question du chômage et de l'emploi, enfin la troisième (classe 4) aborde les questions budgétaires et fiscales. Les classes 11, 8 et 7 regroupent les dépêches sur l'adversaire politique du PS : la droite. La classe 11 souligne la mise en avant des affaires judiciaires impliquant des membres de l'UMP. La classe 8 cible le parti et la classe 7 Nicolas Sarkozy. Le dendrogramme présente ensuite une classe sur l'éducation au sens large (classe 13), une classe traitant des questions sociétales mises en avant par François Hollande pendant sa campagne présidentielle (classe 2), puis une classe relative aux modalités de concertation et de mise en œuvre des réformes sur ces questions (classe 1). Les segments regroupés dans ces deux classes (1 et 2) sont significativement issus de textes mis en ligne à partir de 2012. Les deux classes suivantes abordent les collectivités territoriales (classe 6) et la politique de santé (classe 12).

Analyse du site Web d'Europe Écologie Les Verts

Figure 2. Dendrogramme, taille des classes et lexique caractéristique des classes (par chi2 décroissant) pour le site d'Europe Écologie Les Verts



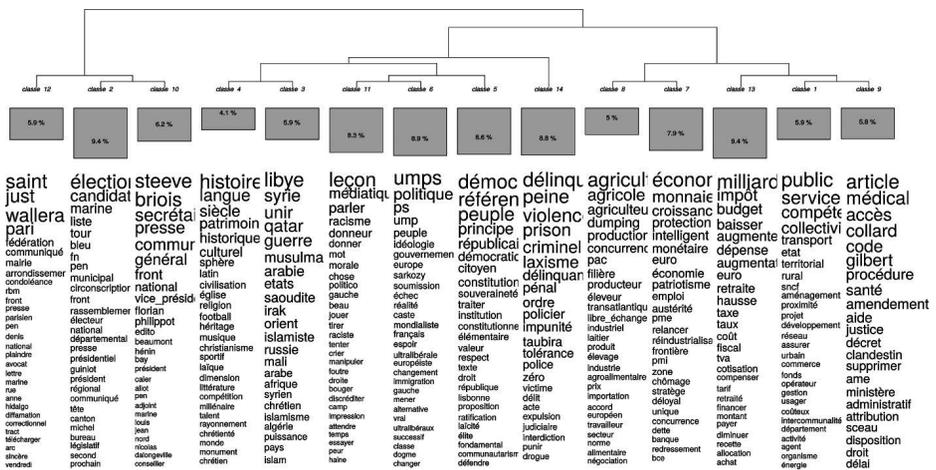
Source : auteurs.

Comme dans l'analyse précédente, les premières classes à se distinguer listent des noms de membres du parti et le vocabulaire relevant des réseaux socio-numériques (classes 2 et 1). Nous trouvons ensuite une classe commentant des résultats électoraux (classe 15), une classe qui porte sur les conférences internationales liées à l'écologie (classe 13), une classe sur le parlement européen (classe 8) et une classe sur le gouvernement (classe 3). La branche suivante

du dendrogramme fait apparaître une classe sur le conflit syrien (classe 10) et une classe sur la justice qui semble plus particulièrement dénoncer des injustices (classe 9). Viennent ensuite une classe sur l'agriculture (classe 7) et une classe sur les énergies (classe 8). La dernière branche du dendrogramme présente une classe sur le logement (classe 14), une classe sur l'aménagement du territoire (classe 12), une classe sur la transition écologique au sens large (classe 11). Enfin, les deux dernières classes contiennent le vocabulaire économique : la classe 5 est caractérisée par le lexique lié au budget et à la fiscalité et la classe 4 par celui lié au système de protection sociale.

Analyse du site Web du Front National

Figure 3. Dendrogramme, taille des classes et lexique caractéristique des classes (par chi2 décroissant) pour le site du Front National



Source : auteurs.

Ici encore, cette classification commence par différencier un groupe de classes dans lesquelles les noms propres sont surreprésentés : deux classes listent des acteurs du parti (classes 12 et 10) et une rend compte de résultats électoraux (classe 2). Le dendrogramme fait ensuite apparaître une classe sur l'histoire et la culture (classe 4), une classe sur le Moyen-Orient et l'islam (classe 3), une classe sur les médias (classe 11), une classe sur l'« UMPS », une classe sur la souveraineté nationale et l'Europe (classe 5) et une classe sur la justice et la délinquance (classe 14). La dernière branche du dendrogramme présente

une classe sur l'agriculture (classe 8), une classe sur l'économie et l'industrie (classe 7), une classe sur le budget et la fiscalité (classe 13), une classe sur les services publics (classe 1) et une classe sur la santé et la justice (classe 9).

Synthèse de la première étape

Ce premier parcours des classifications fait apparaître quelques convergences entre les discours des trois partis. Certaines thématiques traversent clairement tous les corpus (la finance, le budget, l'industrie, l'agriculture, les services publics par exemple), quand d'autres semblent spécifier fortement les partis : l'éducation et la santé pour le *Parti socialiste*, l'écologie pour *Europe Écologie Les Verts* et la délinquance, l'immigration et la nation pour le *Front National*. Ainsi posées, ces thématiques peuvent être interprétées selon les termes d'acceptabilité (marqueurs communs de pertinence et de crédibilité pour le jeu politique) et de distinctivité (marqueurs spécifiques d'identité et de comparaison) et on pourrait être tenté d'en rester là. Mais ces traits apparaissent un peu grossiers et les analyses suivantes permettent d'affiner ces pistes interprétatives.

L'Analyse factorielle des correspondances (AFC)

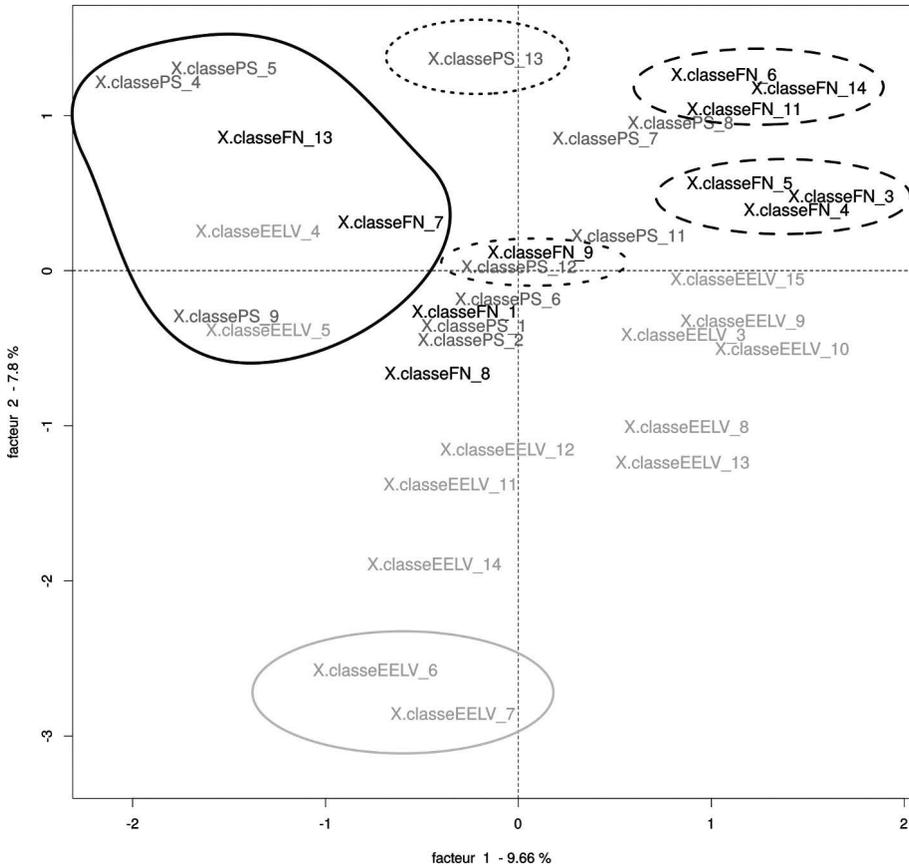
Pour réaliser cette analyse, nous avons éliminé de chacune des classifications précédentes les classes caractérisées par des noms propres ou par le lexique d'Internet (classe 10, 3 et 14 pour le *Parti socialiste*, classes 2 et 1 pour *Europe Écologie Les Verts* et classes 12, 2 et 10 pour le *Front National*).

Les classes restantes déterminent alors les colonnes du tableau tandis que le lexique général en constitue les lignes. L'AFC menée sur ce tableau permet de repérer l'organisation des colonnes (les classes) sur les plans factoriels.

La figure 4 représente le plan factoriel des facteurs 1 et 2 obtenu sur le tableau lexical entier, *i.e.* regroupant toutes les autres classes.

Nous pouvons percevoir un effet massif permettant d'entrevoir des rapprochements et des singularités. La zone entourée en trait plein sur la gauche du graphique signale le regroupement de toutes les classes traitant du budget, de la fiscalité, de l'industrie et de l'État providence. Ces thématiques sont non seulement communes à toutes les analyses, mais elles semblent circonscrites à un lexique suffisamment unifié pour rapprocher les discours des partis sur ces dimensions. C'est également le cas du discours sur la santé, symbolisé par la classe 12 du *Parti socialiste* et la classe 9 du *Front National*.

Figure 4. Plan factoriel de l’AFC sur le tableau lexical entier réunissant l’ensemble des classes



Cette analyse met également en évidence les îlots de spécificité de chaque parti. La classe 13 du *Parti socialiste* qui traite d’éducation se distingue sur le facteur 2 et est isolée de toutes les autres. Au sud du facteur 2, nous trouvons deux classes très caractéristiques d’*Europe Écologie Les Verts* sur l’agriculture (classe 7) et les énergies (classe 6). Cette analyse met à distance le discours des *Verts* sur l’agriculture, pensée prioritairement dans son rapport à l’écologie (pollution, pesticide, OGM, biodiversité...), de celui du *Front National*, plus économique (*dumping*, production, concurrence, libre échange...).

Un ensemble de classes du *Front National* semble également se distinguer : les classes 6 et 11 portent le discours sur les médias et sur les partis politiques

de gouvernement (UMP et PS). En cela, elles partagent du lexique avec les classes du PS centrées sur l'UMP et Nicolas Sarkozy (classes 7 et 8). La classe 14 du FN (en haut à droite) est plus polaire sur le facteur 1, c'est la classe sur la délinquance, que l'on ne retrouve pas dans les autres analyses. Les classes 5, 4 et 3, semblent également très caractéristiques du FN, elles portent sur les dimensions historique, culturelle et religieuse de l'identité française (classe 4), sur le Moyen-Orient et l'islam (classe 3) et sur la souveraineté nationale (classe 5).

Cette AFC confirme les impressions issues du parcours des classes et précise les thématiques sur lesquelles s'établissent les convergences et les distinctions. Si cette analyse permet une représentation graphique rapide des relations entre les classes, son interprétation est toutefois délicate, puisque les deux premiers facteurs résument un nuage à 34 dimensions. Par ailleurs, dans ce domaine de recherche, la triangulation statistique est toujours souhaitable et une confirmation par les distances intertextuelle peut s'avérer nécessaire.

L'ANALYSE ARBORÉE DE LA DISTANCE LEXICALE

Notre seconde analyse utilise l'indice de distance intertextuelle de Dominique Labbé (Labbé et Monière, 2000 ; Labbé et Labbé, 2003). Il s'agit de l'un des indices disponibles dans la littérature pour estimer dans quelle mesure deux textes sont proches ou différents du point de vue du lexique qui les compose. La distance entre les textes A et B, notée $D_{(a,b)}$ est obtenue par le calcul suivant :

$$D_{(a,b)} = \frac{\sum_{v \in A, v \in B} |F_{ia} - E_{ia(v)}|}{N_a + N'_b}$$

Nous supposons ici que le texte B a une taille supérieure au texte A. Dans ce calcul, les fréquences du lexique du texte B sont ramenées à ce qu'elles seraient si l'étendue du texte B était équivalente à celle du texte A (N_a). La fréquence attendue d'une forme i de B dans le texte A, notée $E_{ia(u)}$ dans la formule, est alors égale à la fréquence de la forme i dans B (F_{ib}) multiplié par le ratio de taille entre A et B :

$$E_{ia(u)} = F_{ib} * U_{(a,b)} \text{ avec } U_{(a,b)} = U_a / U_b$$

On note N'_b l'étendue du texte B' , qui est la réduction du lexique de B en fonction de la taille de A . L'indice apprécie alors la distance entre la fréquence des formes dans A (F_{ia}) et celle dans B' . Ce calcul est appliqué à toutes les formes des textes aux exceptions suivantes :

- on ne considère qu'une seule fois les formes communes aux deux textes ;
- on applique le calcul à l'ensemble des formes de A , mais seules les formes de B ayant une fréquence attendue dans A supérieure ou égale à 1 seront considérées.

Soulignons enfin que les auteurs déconseillent l'emploi de cet indice lorsque la différence de taille entre A et B dépasse un facteur 10 (Labbé et Labbé, 2003). Nous avons pleinement conscience que notre utilisation de cet indice s'éloigne souvent des préconisations des auteurs. Par exemple, nous n'appliquerons pas ce calcul à toutes les formes des classes, mais uniquement aux formes pleines, qui sont les seules à avoir structuré la classification.

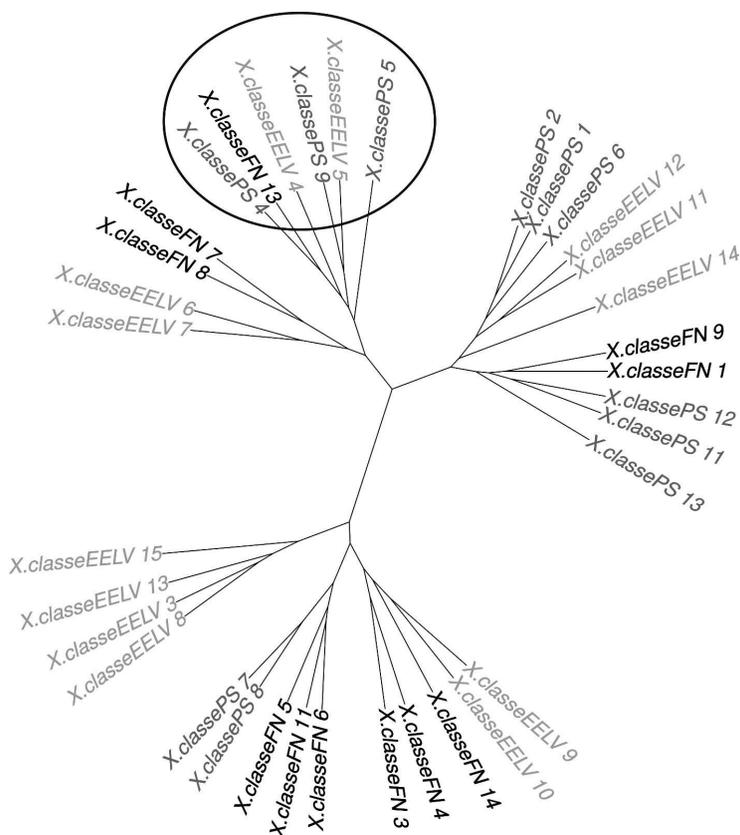
Après avoir calculé, avec l'indice de Dominique Labbé, les distances entre chacune des classes, nous avons soumis la matrice à une classification hiérarchique descendante avec la méthode de Ward afin d'en obtenir une représentation arborée (figure 5).

Sur ce graphique, la distance entre deux classes est représentée par leur distance dans l'arbre. Par exemple, les classes 7 et 8 du PS (en bas de l'arbre) sont plus proches entre elles qu'elles ne le sont des classes 5, 11 et 6 du FN.

Cette analyse vient conforter et renforcer la précédente. Nous retrouvons, entourées en noir, les classes sur le budget, la finance, l'industrie et l'État providence. Mais cette thématique est le seul exemple dans lequel la source du texte n'est pas fortement différenciatrice.

Le rapprochement des classes 7 et 8 du FN et des classes 6 et 7 d'EELV se fait certainement sur le lexique lié au domaine de l'agriculture et d'une partie du lexique économique. Pour autant, sur ces dimensions, les discours se distinguent, comme nous l'avons précédemment indiqué. Il est intéressant de constater que le discours sur l'agriculture d'EELV (classe 7) se rapproche davantage de leur discours sur les énergies (classe 6) que de celui du FN sur l'agriculture (classe 8).

Figure 5. Analyse arborée des distances entre les classes (indice de Labbé)



Source : auteurs.

On retrouve le même phénomène sur le discours du PS à propos de la santé (classe 12), qui se retrouve plus proche du discours du PS sur la justice (classe 11) que du discours du FN sur la santé (classe 9) ou sur le service public (classe 1).

La thématique de la justice traitée par EELV (classe 9) se rapproche plus de leur traitement de la situation syrienne (classe 10) que du discours du FN sur la délinquance (classe 14) et même sur les questions liées au Moyen-Orient (classe 3).

Paradoxalement, cette analyse souligne également la distinction entre les discours : sur chaque ensemble de thématiques, les classes se rapprochent en fonction du parti qui les a produites.

DISCUSSION : L'IDENTITÉ LEXICALE

Au terme de cette analyse, une alternative semble pouvoir émerger à deux théories de sens commun. La première voudrait que tous les politiques disent peu ou prou la même chose. La seconde voudrait que les partis politiques se distinguent sur des thématiques spécifiques.

La classification textométrique que nous proposons évite une focalisation souvent excessive, soit sur les fréquences des formes et ce qu'elles sont censées dire de leur emploi, soit sur les thématiques prétendument privilégiées par les individus et les groupes politiques, pour mettre en évidence, non seulement les organisations lexicales internes des partis politiques (*Front National, Parti socialiste, Europe Écologie Les Verts*), mais également les convergences et divergences des thématiques entre les partis.

Elle montre assez clairement que les thématiques sont les premières régulatrices du lexique. De ce point de vue, le discours politique sur le Web ne se distingue pas de la littérature (Brunet, 2004). La proximité lexicale est d'abord le fait de l'objet du texte, avant d'être celui de l'auteur.

Mais si la thématique était seule responsable des regroupements, toutes les classes sur le même thème convergeraient. Ce n'est pas le cas : les thématiques sont liées entre elles dans des cohérences propres aux partis. Lorsqu'ils abordent les mêmes thématiques, les partis le font donc différemment. Et, surtout, ils se montrent capables d'établir des liens spécifiques entre des thématiques différentes.

L'identité d'un parti repose alors autant sur les thématiques qu'il privilégie que sur les liens lexicaux qu'il établit entre diverses thématiques. L'agriculture, par exemple, sert au FN à renforcer sa préconisation d'un protectionnisme économique et d'une défiance par rapport à l'Europe, là où elle est mobilisée par EELV dans un rapport strict à l'écologie.

Nombre de commentateurs font référence à une stratégie de « dédramatisation » ou de « respectabilité » du *Front National* depuis quelques années, qui

l'amène à s'éloigner des polémiques sur base d'antisémitisme ou d'antisio-nisme, de racisme ou de racialisme, de révisionnisme ou de négationnisme, qui étaient fréquentes et explicites chez les fondateurs du mouvement. Depuis une trentaine d'années, la rhétorique du FN mobilisait volontiers son opposition à l'*établissement* et à la *bande des quatre*, et se jouait sur les terrains de l'immigration, de l'insécurité, des minorités, des races assimilées aux relations familiales, des « détails » de l'histoire de France, des complots et de la corruption des élites... Récemment, le FN a progressivement atténué l'influence des éléments les plus radicaux et amoindri ses liens avec des militants nationaux-catholiques, pour « déringardiser » et « déconfessionnaliser » le parti et présenter une offre politique plus sociale, plus conforme au fond idéal républicain, et se rapprocher d'un répertoire plus proche des partis de gouvernement (Camus et Lebourg, 2015 ; Dézé, 2015).

Les résultats présents confirment effectivement une telle évolution du FN sur sa distinctivité lexicale vis-à-vis des autres partis. Le parti d'extrême droite a su investir d'autres domaines, politique, social et surtout économique, qui ne peuvent pas se satisfaire de simples effets rhétoriques, prenant parfois la forme de « dérapages » médiatiquement exploités. L'économie ne peut pas se contenter d'implicite, de non-dit, de « ils », de « on » et de « j'me comprends », volontiers teintés de dérision (Bonnafous, 2001). Aller sur le terrain de l'économie oblige à convoquer un discours plus objectivé, plus formel, voire même académique, qui a ses règles et son lexique.

La nécessité, pour le FN comme pour les autres partis, de mettre en mots une stratégie économique au service d'un programme acceptable pour la pensée dominante ouvre la possibilité d'une comparaison, au détriment de la distinctivité. Chaque parti est amené à identifier les thématiques communes, qui assurent l'acceptabilité d'un parti de gouvernement, et y créer son propre discours en établissant des liens originaux, que les autres partis ne font pas, entre des lexiques différents.

C'est à partir de ces dynamiques de lexicalisation que les partis s'approprient les registres thématiques pour faire jouer, au travers des marquages socio-langagiers (Scherer et Giles, 1979), des rapports d'alliance et d'opposition intra- et intergroupes. Chaque groupe politique, selon ses intérêts et ses programmes, peut ainsi contextualiser un objet avec d'autres notions jusqu'à pouvoir lui faire jouer un rôle de conditionnement évaluatif (Beauvois, 2011, notamment pp. 70 et suivantes) : en juxtaposant des thématiques, un groupe

va pouvoir créer des effets de sens qui vont avoir une fonction de renforcement identitaire pour le lectorat visé, avec lequel s'établit un contrat particulier (Burguet, 1999) qui vient stabiliser les « mondes lexicaux » (Reinert, 1993).

Cette stabilisation trouvera son expression dans les discours et débats de campagne, où chaque porte-parole, dans chacune de ses interventions, est en position de mobiliser le lexique et surtout les relations lexicales construites en amont. La mise en mots des choses traduit ainsi les relations entre les gens et c'est cette capacité à reconnaître le groupe derrière la structure lexicale qui justifie la notion d'identité lexicale.

L'étape ultime du débat d'entre-deux-tours de l'élection présidentielle a ainsi pu être qualifiée de « conquête lexicale » (Dupuy et Marchand, 2009 ; 2016) pour décrire les stratégies mises en place par chaque duelliste pour occuper l'espace lexical, imposer l'organisation de son vocabulaire sur des thématiques données et dans le plus grand nombre de séquences du débat⁸.

En termes d'analyse de la communication politique, il s'agit d'aller au-delà du prisme réducteur des petites phrases assassines et des moments perçus comme des instants clés pour aborder les relations entre les agencements lexicaux et les stratégies discursives construites par les duellistes et qui sont une manifestation du travail lexical effectué en amont.

CONCLUSION : DE NOUVEAUX HORIZONS MÉTHODOLOGIQUES

Les progrès actuels de la lexicométrie ne reposent pas uniquement sur de nouvelles méthodes. Celles que nous mobilisons ici ne sont pas récentes : la mesure des distances lexicales par l'indice de Labbé a une quinzaine d'années et l'algorithme de classification proposé par Reinert a été publié il y a 35 ans. L'analyse des sites Web politiques fait également partie du paysage de la littérature (voir par exemple Greffët, 2001 ; Bonhomme et Stalder, 2006).

L'originalité de ce travail est davantage de présenter quelques-unes des nouvelles possibilités offertes en matière d'analyses textuelles, qu'il s'agisse de

8. Le débat de l'entre-deux-tours de l'élection présidentielle de 2017, que les commentateurs s'accordent à évaluer comme un échec de la candidate du FN, peut alors être interprété comme le retour en force de la distinctivité au détriment de l'acceptabilité.

la constitution des corpus ou des stratégies d'analyse. Ici, l'étude des sites est exhaustive : l'intégralité des textes publiés sur ces supports par les partis politiques est utilisée. L'extraction et le nettoyage de ces milliers de pages ont été simplifiés et accélérés par l'outil *Gromoteur*. Mais la production de tels corpus n'aurait aucun sens si l'on ne pouvait pas les analyser. C'est l'un des effets les plus évidents du développement des outils et des méthodes de ces dix dernières années : la textométrie s'est adaptée au volume des textes disponibles aujourd'hui.

D'autres adaptations, peut-être moins directement perceptibles, sont également susceptibles d'avoir des conséquences sur les pratiques de recherche. Les outils permettent, de façon de plus en plus simplifiée, de manipuler les corpus, de les segmenter et de les regrouper. Ils facilitent l'exportation des résultats qu'ils produisent dans des formats qui autorisent leur utilisation avec d'autres outils. Il devient alors aisé d'appliquer et de comparer différentes méthodes d'analyses sur les mêmes collections. La manipulation des textes de grandes dimensions réclame cette possibilité de triangulation analytique, qui n'est que la conséquence de la prise en compte des limites propres aux différentes stratégies.

Mais même si la simplification de la mise en place de ces procédures doit se poursuivre, dans un mouvement qui contribue également à leur enseignement, leur utilisation concrète dans la littérature demeure délicate. C'est en fait l'ensemble de la chaîne de production de la connaissance qui doit s'adapter à ces nouvelles possibilités. Les développements des méthodes de visualisation des données textuelles nécessiteraient par exemple des supports qui permettent *a minima* de publier des graphiques en couleur. Lorsque cela est possible, et dans une visée falsificationniste, les corpus utilisés devraient être systématiquement partagés pour faciliter les contre-expertises, les méta-analyses, les comparaisons et la formation aux outils.

La conséquence la plus louable, à nos yeux, de ces développements est illustrée dans l'étude présente. Tous les commentateurs de la vie politique ont une « idée », une « image » des discours portés par les partis. Mais les sciences humaines et sociales ont toujours eu la volonté de substituer les représentations naïves de la vie sociale par des faits. Certes, aucune méthode ne peut se prévaloir d'exposer une vérité indépassable et l'approche textométrique contient également une phase d'interprétation des résultats qui mobilisera forcément des catégories et des modes interprétatifs empreints d'une part de

subjectivité des chercheurs. Pour autant, il convient de souligner que cette démarche, qui établit formellement les relations entre textes et co(n)textes (Mayaffre, 2014) limite fortement cette part et guide l'interprétation vers la lecture de faits statistiques liés à la distribution du lexique dans les textes. Les résultats qu'elle propose sont par ailleurs reproductibles : les pages analysées sont toujours disponibles en ligne (et elles le seront sûrement encore longtemps, grâce aux outils d'archivage du Web comme la *Wayback Machine*⁹ par exemple). Il est alors envisageable de reconstituer ce corpus et de reproduire à l'identique les analyses, ou d'en appliquer d'autres, pour une co-construction critique et raisonnée des effets de sens.

9. <https://archive.org/web/>

 RÉFÉRENCES

- ALLEN V.L., WILDER D. A., ATKINSON M.L. (1983), « Multiple group membership and social identity », in T.R. Sarbin, K.E. Scheibe (Eds.), *Studies in Social Identity* (pp. 92-115). New York, Praeger.
- AUGOUSTINOS M., QUINN C. (2003), « Social categorization and attitudinal evaluations: Illegal immigrants, refugees, or asylum seekers? », *Nouvelle Revue de psychologie sociale*, vol. 2, n° 1, pp. 29-37.
- BEAUVOIS J.-L. (2011), *Les influences sournoises. Précis des manipulations ordinaires*. Paris, François Bourin.
- BONHOMME, M., STALDER, P. (2006), « Analyse sémiolinguistique des pages d'accueil des sites politiques suisses sur Internet », *Mots. Les langages du politique*, n° 80, pp. 11-23.
- BONNAFOUS S. (2001), « L'arme de la dérision chez Jean-Marie Le Pen », *Hermès*, n° 29, pp. 53-63.
- BOURHIS R.Y., GILES H. (1977), « The language of intergroup distinctiveness. Language, ethnicity, and intergroup relations », in H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119-136), London, Academic Press.
- BRENNAN S.E., CLARK H.H. (1996), « Conceptual pacts and lexical choice in conversation », *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, n° 6, pp. 1482-1493.
- BRUNET E. (2004), « Où l'on mesure la distance entre les distances ? », *Revue Texto*. En ligne : <http://www.revue-texto.net/index.php?id=645>
- BURGUET A. (1999), « Le lecteur, le texte et le contexte », *Psychologie française*, vol. 44, n° 1, pp. 65-73.
- CAMUS J.-Y., LEBOURG N. (2015), *Les droites extrêmes en Europe*. Paris, Seuil.
- DÉZÉ A. (2015), « La "dédiabolisation". Une nouvelle stratégie ? », in S. Crépon, A. Dézé, N. Mayer (dir.), *Les faux-semblants du Front national : sociologie d'un parti politique*. Paris, Presses de Sciences Po.
- DUPUY P.-O., MARCHAND P. (2009), « Débat de l'entre-deux tours 2007 : la conquête de l'espace lexical », *Mots. Les langages du politique*, n° 89, pp. 105-117.
- DUPUY P.-O., MARCHAND P. (2016), « Les débats de l'entre-deux-tours de l'élection présidentielle française (1974-2012) au prisme des stratégies discursives : du monopole du cœur à la doxa économique-comptable », *Mots. Les langages du politique*, n° 112, pp. 69-80.
- GERDES, K. (2014), « Corpus collection and analysis for the linguistic layman: The Gromoteur », in E. Née, M. Valette, J.-M. Daube, S. Fleury (dir.), *JADT 2014*.

Journées internationales d'analyse statistique des données textuelles (pp. 261-269). Paris.

GHIGLIONE, R. (1985), « L'enfant paraît, un enfant disparaît », *Psychologie française*, vol. 30, n° 1, pp. 59-68.

GHIGLIONE R. (1989), *Je vous ai compris ou l'analyse des discours politiques*. Paris, Armand Colin.

GREFFET F. (2001), « Les Partis politiques français sur le Web », in D. Andolfatto, F. Greffé, L. Olivier (dir.), *Les Partis politiques. Quelles perspectives ?*, Paris, L'Harmattan, pp. 161-178.

KRIEG-PLANQUE A. (2009), *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Besançon, Presses universitaires de Franche-Comté.

LABBE D., MONIÈRE D. (2000), « La connexion intertextuelle. Application au discours gouvernemental québécois », in M. Rajman, J.-C. Chappelier (dir.), *Actes des 5^{es} Journées internationales d'analyse statistique des données textuelles*, Lausanne, EPLF, pp. 85-94.

LABBÉ, C., LABBÉ D. (2003), « La distance intertextuelle », *Corpus*, 2, pp. 95-117.

LAVER J., TRUDGILL P. (1979), « Phonetic and linguistic markers in speech », in K.R. Scherer, H. Giles (Eds.), *Social Markers in Speech*, Cambridge, Cambridge University Press, pp. 1-32.

LEBART L., SALEM A. (1994), *Statistiques textuelles*. Paris, Dunod.

LORENZI-CIOLDI F., DAFFLON A.-C. (1999), « Rapports entre groupes et identité sociale », in J.-L. Beauvois, N. Dubois, W. Doise (dir.), *La construction sociale de la personne* (pp. 131-145). Grenoble : Presses universitaires de Grenoble.

MANGE J. (2016), *Amorçage de rapports : intergroupes et conséquences*. Document de synthèse pour l'obtention d'HDR, Université de Caen.

MANGE J., LEPASTOUREL, N. (2013), « Gender Effect and Prejudice: When a Salient Female Norm Moderates Male Negative Attitudes Toward Homosexuals », *Journal of Homosexuality*, vol. 60, n° 7, pp. 1035-1053.

MANGE J., LEPASTOUREL N., GEORGET, P. (2009), « Is your language a social clue? Lexical markers and social identity », *Journal of Language and Social Psychology*, vol. 28, n° 4, pp. 364-380.

MANGE J., MARCHAND P. (2011), « Convaincre de ses similitudes tout en démontrant ses différences : un exemple d'analyse automatique des débats internes au PS », in L. Bagnet, T. Guilbert (dir.), *Discours en contextes*, Paris, PUF, coll. « Curapp », pp. 87-101.

MANGE J., MARCHAND P., GAFFIE B. (2004), « Positionnement politique et traitement de l'information », *Revue internationale de psychologie sociale*, vol. 17, n° 4, pp. 23-49.

- MARCHAND P. (2010), « Attitudes towards language: new perspectives », *International Review of Social Psychology*, vol. 23, n° 3-4, pp. 125-153.
- MARCHAND P., RATINAUD P. (2014), « Postures sociodiscursives et conquête du pouvoir », *Recherches en communication*, n° 41, pp. 117-139.
- MARTY E., REBILLARD F., POUCHOT S., LAFOUGE T. (2012), « Diversité et concentration de l'information sur le Web. Une analyse à grande échelle des sites d'actualité français », *Réseaux*, n° 176, pp. 27-72.
- MAYAFFRE D. (2014), « Plaidoyer en faveur de l'Analyse de Données co(n)Textuelles. Parcours cooccurrentiels dans le discours présidentiel français (1958-2014) », in *Actes des JADT 2014*, Paris, Inalco-Sorbonne nouvelle, pp. 15-32.
- MIKOŁAJCZAK M., BILEWICZ M. (2015), « Foetus or child? Abortion discourse and attributions of humanness », *British Journal of Social Psychology*, vol. 54, n° 3, pp. 500-518.
- PAPASTAMOU S. (1993), « Valeur stratégique et émergence de la psychologisation comme résistance à l'innovation », in J.-L. Beauvois, R.-V. Joule, J.-M. Monteil (dir.), *Perspectives cognitives et conduites sociales, IV* (pp. 175-194). Fribourg, Cousset-DelVal.
- RATINAUD P. (2014), « IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires (Version 0.7 alpha 2) [Windows, GNU/Linux, Mac OS X] ». En ligne : <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 11^{es} Journées internationales d'Analyse statistique des Données Textuelles (JADT)*, 13-15 juin, pp. 835-844.
- RATINAUD P., MARCHAND P. (2015), « Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014) », *Mots. Les langages du politique*, 108, pp. 57-77.
- RATINAUD P., MARCHAND P. (2016), « Quelques méthodes pour l'étude des relations entre classifications lexicales de corpus hétérogènes : application aux débats à l'assemblée nationale et aux sites Web de partis politiques », in D. Mayaffre, C. Poudat, L. Vanni, V. Magri, P. Follette (Eds.), *Statistical Analysis of Textual Data* (pp. 193-202). Nice.
- REINERT M. (1983), « Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte », *Les Cahiers de l'analyse des données*, vol. 8, n° 2, pp. 187-198.
- REINERT M. (1993), « Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars », *Langage et société*, n° 66, pp. 5-39.

SANREY C., TESTE B., MANGE J. (2016), « “Homosexual Marriage” or “Marriage for All”? Social Lexical Markers’ Effects on Persuasion », *Journal of Language and Social Psychology*, vol. 35, n° 5, pp. 471-490.

SCHERER, K.R., GILES, H. (Eds.) (1979). *Social markers in speech*. Cambridge: Cambridge University Press/Éditions de la Maison des Sciences de l’Homme.

TAJFEL H. (1974), « Social identity and intergroup behaviour », *Social Science Information*, n° 13, pp. 65-93.

TAJFEL H. (1979), « Individuals and groups in social psychology », *British Journal of Social Psychology*, n° 18, pp. 183-190.