



**HAL**  
open science

## Le vocabulaire des campagnes électorales

Dominique Labbé, Denis Monière

► **To cite this version:**

Dominique Labbé, Denis Monière. Le vocabulaire des campagnes électorales. 14th International Conference on Statistical Analysis of Textual Data, Department of Mathematics and Statistics - University of Roma "la Sapeienza", Jun 2018, Roma, Italie. pp.522 - 540. halshs-01818298

**HAL Id: halshs-01818298**

**<https://shs.hal.science/halshs-01818298>**

Submitted on 18 Jun 2018

**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.

Denis Monière<sup>1</sup>, Dominique Labbé<sup>2</sup>

## Le vocabulaire des campagnes électorales

Iezzi Domenica F., Celardo Livia, Misuraca Michelangelo. *Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*. Roma; UniversItalia, 2018, p. 522-540.

<sup>1</sup> Université de Montréal ([denis.moniere@umontreal.ca](mailto:denis.moniere@umontreal.ca))

<sup>2</sup> PACTE CNRS - Université de Grenoble ([dominique.labbe@umrpacte.fr](mailto:dominique.labbe@umrpacte.fr))

### Abstract

After having done a first presidential term, C. de Gaulle, V. Giscard d'Estaing, F. Mitterrand, J. Chirac and N. Sarkozy were candidates for a second term. In this study, their electoral speeches are compared with their presidential ones drawing attention to the specific nature of the vocabulary used. It would appear that this calculation is mainly biased by grammatical categories and word frequency. We present modifications of the classical formulae which make it possible to neutralize the influence of grammatical categories and, at least partially, that of word frequency. Electoral discourse privileges the verb over the name, as such speech is more personalized than governmental discourse, it focuses on the country and its inhabitants, the rest of the world being pushed into the background. Finally, in recent years, the polemical dimension is becoming predominant.

### Résumé

Après un premier mandat présidentiel, C. de Gaulle, V. Giscard d'Estaing, F. Mitterrand, J. Chirac et N. Sarkozy ont été candidats à un deuxième mandat. On compare leurs discours électoraux avec leurs discours présidentiels à l'aide des spécificités du vocabulaire. Il apparaît que ces spécificités dépendent surtout des catégories grammaticales et des effectifs des mots. On présente des modifications du calcul classique qui permettent de neutraliser l'influence des catégories grammaticales et, au moins partiellement, celle des fréquences. Le discours électoral privilégie le verbe au détriment du nom, il est plus personnalisé que le discours au pouvoir, il se centre sur le pays et ses habitants, le reste du monde passant au second plan. Enfin, ces dernières années, la dimension polémique devient prédominante.

**Keywords:** lexicometry ; political discourse ; French presidential campaigns ; specific vocabulary ; spécificités du vocabulaire.

Manuscrit des auteurs. Toute citation doit être faite à partir de l'ouvrage cité ci-dessus.

## 1. Introduction

Le discours électoral diffère-t-il du discours de gouvernement et en quoi ? La réponse est difficile car il faut neutraliser l'effet des personnalités et des conjonctures pour isoler l'effet sur le discours des choix stratégiques du locuteur. L'idéal serait de pouvoir étudier les mêmes hommes à peu près simultanément dans les deux positions de gouvernant puis de candidat. Le corpus des discours des présidents français depuis 1958 remplit ces deux conditions (présentation du corpus dans Arnold et al 2016). En effet, pour 5 présidents (C. de Gaulle, V. Giscard d'Estaing, F. Mitterrand, J. Chirac et N. Sarkozy), ce corpus contient leurs interventions lorsqu'ils étaient présidents et leurs discours de campagne pour leur réélection. Certes, en 1965, de Gaulle n'a pratiquement pas fait campagne (Labbé 2005), mais ses successeurs ne l'ont pas imité en 1981, 1988, 2002 et 2012 (corpus en annexe).

Pour comparer ces corpus, le calcul des "spécificités" semble l'outil le plus adapté (Lafon 1980 et 1984). Il rapporte le vocabulaire d'un sous-ensemble de textes (sous-corpus) à un corpus de référence. Mais il se heurte à une double difficulté : la spécificité éventuelle d'un vocable est liée à sa catégorie grammaticale et à sa fréquence d'emploi (Labbé, Labbé 1994 ; Monière et al. 2005), comme nous allons le vérifier d'abord avec le cas de Sarkozy en 2012 (Sur cette campagne : Labbé, Monière 2013). Dès lors, la mesure des spécificités doit neutraliser, autant que possible, ces deux inconvénients.

## 2. Les catégories grammaticales du discours électoral

Le discours présidentiel de Sarkozy s'étend de son investiture (16 mai 2007) au 12 février 2012 (annonce de sa candidature). La campagne s'étend jusqu'au soir du second tour (6 mai 2012). Le corpus complet (P) compte 1074 interventions, soit au total 3 221 259 mots avec 21 602 vocables différents. A partir de sa déclaration de candidature, Sarkozy est intervenu 110 fois (sous-corpus E), soit 369 808 mots et un vocabulaire de 8 511 vocables différents. Ces interventions sont d'abord marquées par un net changement de style (tableau1).

Catégories	P-E (Corpus-Sous corpus)	E Sous corpus	(P-E)/P	Indice
Verbes	159.2	169.4	+6.4	+
Futurs	7.0	7.2	+1.6	+
Conditionnels	3.2	2.8	-11.2	-
Présents	82.9	89.3	+7.7	+
Imparfaits	6.4	6.4	-0.2	≈
Passés simple	0.6	0.3	-55.2	-
Participes passés	20.8	23.8	+14.6	+
Participes présents	2.1	2.1	+2.9	≈
Infinitifs	36.3	37.6	+3.6	+
Noms propres	27.9	23.0	-17.3	-
Substantifs	178.4	176.0	-1.3	-
Adjectifs	54.0	46.6	-13.7	-
Adj. participe passé	5.2	4.5	-13.1	-
Pronoms	124.3	132.6	+6.7	+
Pronoms personnels	65.4	69.6	+6.5	+
Déterminants	181.6	182.5	+0.5	+
Articles	131.9	128.1	-2.9	-
Nombres	18.7	20.9	+11.9	+
Possessifs	14.5	17.0	+17.3	+
Démonstratifs	7.6	7.8	+2.7	+
Indéfinis	8.9	8.7	-2.4	-
Adverbes	67.1	68.9	+2.7	+
Prépositions	150.1	145.6	-3.0	-
Coordinations	29.1	25.4	-12.7	-
Subordination	25.9	27.9	+8.0	+

Tableau 1. Densités des catégories grammaticales dans les interventions de Sarkozy lors de la campagne de 2012 comparées à ses interventions comme président 2007-2012 (en ‰)

Dans le discours présidentiel, on rencontre 159 verbes en moyenne pour 1 000 mots ; dans les discours électoraux, cette proportion passe à 169‰, soit une augmentation de +6,4%, ce qui est un écart significatif avec moins de une chance sur 10 000 de se tromper (signe + en dernière colonne). Les lignes suivantes donnent le détail des temps et des modes. Le recul le plus significatif concerne le conditionnel (le discours électoral ne doit pas connaître le doute). En revanche, le participe passé connaît l'augmentation la plus forte (le président sortant peut difficilement éviter de défendre sa gestion).

Les pronoms, les adverbes et les conjonctions de subordination évoluent dans le même sens que les verbes. Ils sont réunis dans le "groupe du verbe". A l'inverse, les substantifs, adjectifs, articles et prépositions suivent la tendance inverse : groupe du nom. Le tableau 3 donne les densités des deux groupes chez les 4 présidents.

Catégories	P-E (Corpus- Sous corpus)	E Sous corpus	(P-E)/E	Indice
Sarkozy (2007-2012)				
Groupe du verbe	376.6	398.9	+5.9	+
Groupe du nom	621.1	599.2	-3.5	-
Giscard d'Estaing (1974-1981)				
Groupe du verbe	351.5	392.5	+11.7	+
Groupe du nom	646.1	604.5	-6.4	-
Mitterrand (1981-1988)				
Groupe du verbe	386.4	427.1	+10.5	+
Groupe du nom	611.0	569.8	-6.7	-
Chirac (1995-2002)				
Groupe du verbe	329.5	333.2	+1.1	+
Groupe du nom	668.8	665.1	-0.6	-

Tableau 3. Densités des groupes du verbe et du nom (en %) dans les discours électoraux (E) comparés aux discours présidentiels (P-E).

Chez tous les présidents en campagne, il se produit une augmentation du groupe du verbe et un recul de celui des noms. Statistiquement, ces mouvements sont significatifs (avec  $\alpha = 1\%$ ). L'écart le plus fort est observé chez Giscard d'Estaing puis chez Mitterrand. Cependant, Chirac tranche sur les autres avec une densité du verbe beaucoup plus faible et une campagne présidentielle presque aussi distancée que ses interventions lors de son premier mandat, marqué par une cohabitation de 5 ans (1997-2002) avec un Premier ministre socialiste (Jospin). Dans son discours électoral, la densité des verbes augmente nettement (+3,6%) mais se trouve en partie compensée par un recul des pronoms, ce qui accentue le caractère dépersonnalisé des propos de Chirac à l'opposé des trois autres.

En conséquence, pour les 4 présidents, les principaux verbes apparaissent en spécificités positives du discours électoral et il ne s'en trouve que quelques-uns en spécificités négatives. Il en est de même pour les pronoms et les adverbes. La situation inverse se constate pour les adjectifs, les substantifs, etc. Autrement dit, si un mot appartient à une catégorie sous-employée dans le sous-corpus (par rapport à sa densité d'utilisation dans le corpus entier), ce vocable a toute chance d'apparaître dans les spécificités négatives (et positives dans le cas inverse). Il est possible de neutraliser ce biais.

### 3. Neutralisation de la catégorie grammaticale

Le calcul standard est le suivant. Soit :

- le corpus de référence ( $P$ ) long de  $N_p$  mots ;
- le sous-corpus  $E$  long  $N_e$  mots dont on recherche les spécificités par rapport à  $P$  ;
- un vocable  $i$  avec  $F_{ip}$  occurrences dans  $P$  et  $F_{ie}$  dans  $E$ . Si sa répartition est uniforme, ce vocable apparaîtra  $E_{ie(u)}$  fois dans le sous-corpus  $E$  :

$$E_{ie(u)} = F_{ip} * U \text{ avec } U = \frac{N_e}{N_p} = \frac{369808}{3223570} = 0.113 \quad (1)$$

La probabilité pour que le vocable  $i$  soit observé  $F_{ie}$  fois dans  $E$  suit une loi hypergéométrique de paramètres  $F_{ip}$ ,  $F_{ie}$ ,  $N_e$ ,  $N_p$  :

$$P(X=F_{ie}) = \frac{\begin{bmatrix} F_{ip} \\ F_{ie} \end{bmatrix} \begin{bmatrix} N_p - F_{ip} \\ N_e - F_{ie} \end{bmatrix}}{\begin{bmatrix} N_p \\ N_e \end{bmatrix}} \quad (2)$$

L'indice de spécificité ( $S$ ) est la somme des probabilités – calculées avec (2) – de survenue des  $J$  valeurs entières de  $X$  variant de 0 à  $F_{ie}$   $\{X=0 ; X= F_{ie}\}$  :

$$S = P(X \leq F_{ie}) = \sum_{j=0}^{j=F_{ie}} P(X=j) \quad (3)$$

Si au seuil  $\alpha$ ,  $F_{ie}$  excède  $E_{ie(u)}$ , le vocable est « spécifique plus » ( $S+$ ) ;  $S-$  dans le cas contraire. Avec ce calcul, la plus grande partie des verbes usuels de Sarkozy apparaissent donc en  $S+$  de sa campagne électorale et la majorité des substantifs en  $S-$ , parce que, dans ses discours électoraux, la première catégorie est privilégiée par rapport au discours de gouvernement où elle est moins utilisée (à l'inverse des substantifs). Pour corriger ce biais, le calcul prend en compte les catégories grammaticales ( $g$ ). La modification est présentée dans : Monière, Labbé, Labbé 2005 ; Mayaffre 2006 et Monière, Labbé 2012.

Soit :  $N_{ge}$  et  $N_{gp}$  le nombre de mots appartenant à la catégorie grammaticale  $G$  respectivement dans le sous-corpus  $E$  et le corpus entier  $P$ . Les formules (1) et (2) deviennent :

$$E_{ie(u)} = F_{ip} * U \text{ avec } U = \frac{N_{ge}}{N_{gp}} \quad (4)$$

$$P(X=F_{ie}) = \frac{\begin{bmatrix} F_{ip} \\ F_{ie} \end{bmatrix} \begin{bmatrix} N_{gp} - F_{ip} \\ N_{ge} - F_{ie} \end{bmatrix}}{\begin{bmatrix} N_{gp} \\ N_{ge} \end{bmatrix}} \quad (5)$$

Les formules (4) et (5) appliquées aux 4 corpus aboutissent à un équilibre relatif, au sein de chaque catégorie, entre les  $S+$  et les  $S-$  (tableau 4). Ces formules neutralisent donc la liaison entre spécificités et densité des catégories grammaticales. Comme indiqué dans Monière & Labbé 2012, cette modification change drastiquement la liste des "mots spécifiques" mais elle laisse subsister la liaison entre spécificité et fréquence.

#### 4. Questions de seuils

Le calcul porte sur une minorité du vocabulaire et il est asymétrique. En effet, avec  $\alpha = 1\%$  :  
 - l'effectif minimal pour être S+ est de 5 occurrences ("seuil de spécificité positive"), toutes dans les discours électoraux ( $E$ ) et à condition que  $E_{ie(u)} < .5$ , ce qui signifie que  $N_{ge} < 0.10N_{gp}$ . Par construction le calcul élimine donc tous les vocables d'effectifs inférieurs à 5. Dans le corpus Sarkozy, cela représente plus de la moitié du vocabulaire (54 % des vocables). Autrement dit, seulement 46% du vocabulaire peut être S+ ;

- le "seuil de spécificité négative" correspond à la situation suivante : un vocable  $i$  absent de  $E$  ( $F_{ie} = 0$ ) alors qu'on en attend au moins 5 ( $E_{ie(u)} \geq 5$ ). En pratique, cela signifie que son effectif dans  $P$  est égal ou supérieur à  $5*1/U$ , soit ici 40. Autrement dit, pour le discours électoral de Sarkozy, 83% du vocabulaire de  $P$  ne peut apparaître en S-.

Dès lors, les vocables dont les effectifs dans  $P$  sont compris entre 5 et 39 peuvent être S+ mais pas S- dans  $E$ . On s'attend donc à ce qu'il y ait plus de vocables S+ que S-.

#### 5. Liaison entre spécificité et fréquence

9 876 vocables apparaissent 5 fois ou plus dans  $P$ . Si ce corpus était homogène (hypothèse nulle  $H_0$ ), une distribution normale des vocables laisserait attendre - avec  $\alpha = 1\%$  - environ 100 vocables spécifiques. Le tableau 4 compare les résultats observés et attendus (avec  $H_0$ ).

	Effectifs ( $F_{ip} \geq 5$ )	$H_0$	S+	S-	Total S
Verbes	1 540	15	176	143	319
Mots à majuscule	1 501	15	112	142	254
Substantifs	4 175	42	455	468	923
Adjectifs	2 065	21	140	115	255
Pronoms	52	1	18	13	31
Adverbes	411	4	20	57	77
Déterminants	72	1	21	12	33
Prépositions & conjonc.	60	1	21	9	30
Total	9 876	100	963	959	1 922

Tableau 4. Effectifs des vocables classés par catégories grammaticales et par spécificités

Il y a donc vingt fois plus de vocables spécifiques que n'en laisse attendre  $H_0$  (répartition homogène des mots entre corpus et sous-corpus). A priori, cela signifie simplement que discours électoral et discours de gouvernement sont fortement contrastés. En fait, ce décalage provient essentiellement des vocables les plus fréquents (tableau 5 et Figure 1).

Classe de fréquence ( <i>P</i> )	Vocables spécifiques de <i>E</i> dans la classe	Total vocables de <i>P</i> dans la classe	Proportion des vocables de <i>P</i> spécifiques de <i>E</i>
5-9	64	2 759	2,3
10-14	68	1 237	5,5
15-19	55	757	7,3
20-29	89	987	9,0
30-49	143	997	14,3
50-99	317	1 054	30,1
100-199	332	799	41,6
200-499	398	686	58,0
500+	473	640	73,9
Total	1 939	9 916	19,6

Tableau 5. Proportion des vocables spécifiques de *E* dans l'ensemble du vocabulaire (*P*) classé par fréquence absolues.

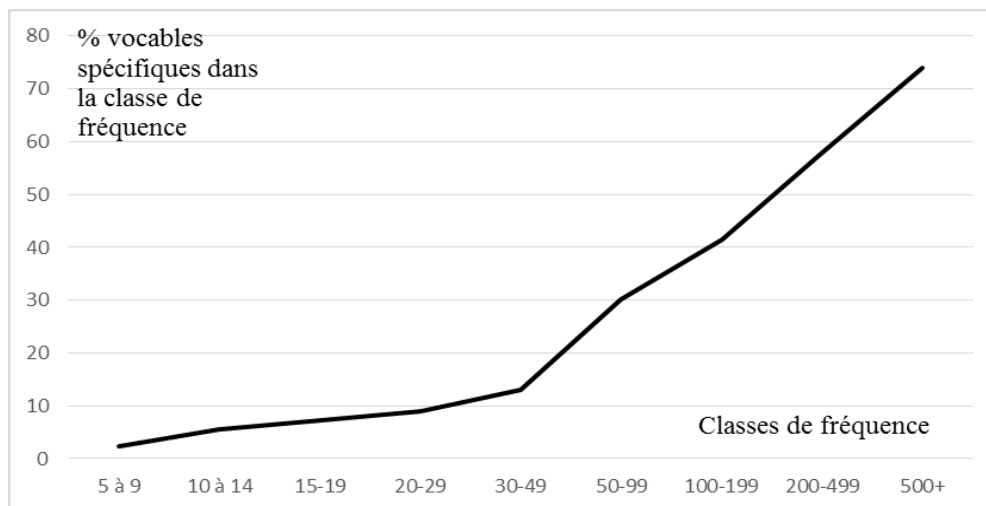


Figure 1. Liaison entre la spécificité et la fréquence

Au-dessus du seuil de spécificité positive (ici 40), la proportion de vocables spécifiques est directement corrélée avec la fréquence : la courbe suit la diagonale du tableau et le coefficient de détermination de *Y* par *X* est égal à 0,997, ce qui indique une liaison rigide et linéaire. Il en est toujours ainsi : plus un vocable donné est fréquent dans un corpus, plus il a de chances d'être "spécifique" à l'une quelconque des parties de ce corpus. Cette dépendance peut être interprétée de deux manières. D'une part, l'essentiel des choix thématiques seraient véhiculés par les vocables les plus fréquents et la variation dans leurs fréquences d'emploi seraient la principale manifestation de ces choix. Cependant, dès que le corpus atteint une certaine longueur, l'observateur se trouve noyé dans des listes qui contiennent la plus grande part du vocabulaire usuel, ce qui en rend l'interprétation difficile. D'autre part et à l'inverse, on peut penser que le raisonnement probabiliste - qui sous-tend ce calcul - doit être adapté à cette liaison manifeste entre spécificité et fréquence.



## 6. Neutralisation de la liaison entre fréquence et spécificité

Les limites des classes de fréquence du tableau 5 et de la figure 1 ont été fixées selon une échelle proche d'une progression géométrique, ce qui assure aux classes des effectifs sinon égaux du moins suffisamment proches et importants. Ceci correspond à une particularité dite "loi de Zipf" - ou "Zipf-Mandelbrot" - selon laquelle le nombre d'occurrences d'un mot dans un texte est lié à son rang dans la distribution des fréquences (Zipf 1935 ; Mandelbrot 1957).

Dès que le corpus atteint une longueur suffisante (au moins un demi-million de mots) et que le sous-corpus est égal à au moins d'un dixième du corpus, on peut découper le vocabulaire en quelques classes de fréquence. Pour un corpus de la dimension de celui de Sarkozy (et des trois autres présidents), trois classes suffisent : vocables "rares" (inférieurs à 100 occurrences) ; "fréquents" (de 100 à moins de 500) ; "très fréquents" (500 et plus). Dans ces trois classes, les vocables sont classés par catégorie grammaticale puis en fonction de leur indice de spécificité et, dans chacune des classes, seuls les plus caractéristiques sont retenus. Le tableau 6 donne les 5% les plus caractéristiques du discours électoral de Sarkozy comparé à son discours présidentiel, pour trois catégories grammaticales.

	<100	100 – 499	500+
Vocables significativement sur-employés :			
Verbes :	voler, cotiser, détester, casser, éduquer, suspendre, démolir	adresser, bénéficier, apprendre, souffrir, supprimer, régulariser	dire, vouloir, parler, vivre, proposer, changer, respecter, défendre
Mots à majuscule	Mélenchon, Le Pen,	François, Polynésie, Hollande, Schengen, TVA	France, Français, Corse
Substantifs	honte, rassemblement, héritier, socialiste, colère, délit, amalgame	jeunesse, souffrance, gauche, destin, erreur, étranger, salaire, outremer	travail, entreprise, droit, république, vie, emploi, ami, enfant, territoire, peuple,
Vocables significativement sous-employés :			
Verbes	admirer, illustrer, expérimenter, inaugurer,	progresser, témoigner, évoquer, marquer, associer	être, devoir, savoir, comprendre, trouver, attendre, remercier, essayer
Mots à majuscule	Bush, Poutine, Roumanie, Qatar	Russie, Inde, Iran, Barroso	Afrique, G20, Méditerranée, Merkel, Paris, Chine
Substantifs	refondation, coalition, scientifique, lycéen	processus, visite, équipe, conférence, planète, gouvernance, alliance,	pays, monsieur, président, état, ministre, politique, gouvernement, question

Tableau 6. Spécificités les plus remarquables du discours électoral de Sarkozy par rapport à son discours présidentiel (par catégories grammaticales en trois classes de fréquence)

Chez Sarkozy, le discours électoral est affaire de volonté, il se centre sur le pays, ses habitants mais aussi l'adversaire – la gauche, Hollande - dont il dénonce les amalgames et les erreurs. Les spécificités négatives indiquent que le discours électoral n'est pas affaire de devoir ou de

connaissance ; il "oublie" le reste du monde et ses dirigeants, les institutions du pays comme le gouvernement et les ministres, etc.

## 7. Conclusions

Lorsqu'un président entre en campagne, il doit descendre dans l'arène et adopter un discours de combat qui se caractérise avant tout par une augmentation de la densité des verbes, une forte personnalisation et un recul de la place accordée aux substantifs et aux adjectifs. La politique internationale et les institutions passent au second plan. Ces caractéristiques se retrouvent dans les discours électoraux des Premiers ministres canadiens (Monière, Labbé 2010). Cependant, en campagne ces derniers insistent sur le "nous" car, dans un système parlementaire, il s'agit de faire élire une majorité de députés, alors que les présidents français privilégient le "je"... Enfin, ces dernières années en Amérique du nord comme en France, la forte présence de la construction négative et la désignation des adversaires (noms propres) soulignent le caractère polémique du discours électoral.

Le calcul des spécificités – tel qu'il est utilisé en analyse des données textuelles – enregistre la catégorie grammaticale du vocable analysé et sa fréquence d'emploi et non pas les choix thématiques du locuteur. La neutralisation de la catégorie grammaticale est aisée si les mots ont été étiquetés. En revanche, l'effet de la fréquence est susceptible de plusieurs interprétations. Toutefois, si l'on souhaite ne pas être enseveli sous les listes produites par le calcul classique, la solution réside dans le classement des vocables en classes de fréquence – délimitée selon une échelle géométrique afin de leur assurer des effectifs sensiblement égaux - et, au sein de chacune de ces classes, dans la sélection des vocables les plus singuliers. A ce prix, les singularités d'un sous-corpus peuvent être identifiées sans avoir à effectuer des tris discutables dans des listes trop longues.

## Remerciements

Les programmes informatiques et les corpus utilisés ont été réalisés avec l'aide de Cyril Labbé (Laboratoire d'Informatique de Grenoble).

Cette communication a bénéficié d'une relecture attentive de : Edouard Arnold (Trinity College), Cyril Labbé (Laboratoire d'Informatique de Grenoble), Jacques Savoy (Université de Neuchâtel) ainsi que des deux relecteurs anonymes du comité de programme des JADT que les auteurs remercient pour leurs remarques très utiles.

Les corpus utilisés sont disponibles en ligne au Centre de linguistique de corpus (Université de Neuchâtel).

## References

Arnold E., Labbé C. & Monière D. (2016). *Parler pour gouverner : Trois études sur le discours présidentiel français*. Grenoble : Laboratoire d'Informatique de Grenoble, 2016.

- Labbé C., Labbé D. (1994). *Que mesure la spécificité du vocabulaire ?* Grenoble : CERAT, décembre 1994. Reproduit dans *Lexicometrica*, 3, 2001.
- Labbé D. (2005). Le général de Gaulle en campagne. In Banks D. (Ed.). *Aspects linguistiques du texte de propagande*. Paris : L'Harmattan, 2005, p. 213-233.
- Labbé D., Monière D. (2010). Quelle est la spécificité des discours électoraux? Le cas de Stephen Harper. *Canadian Journal of Political Science*, 43:1, p. 69–86.
- Labbé D., Monière D. (2013). *La campagne présidentielle de 2012. Votez pour moi !* Paris : l'Harmattan.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, 1, p. 127-165.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève-Paris : Slatkine-Champion.
- Mandelbrot B. (1957). Étude de la loi d'Estoup et de Zipf Fréquences des mots dans le discours. Apostel L et al. *Logique, langage et théorie de l'information*. Paris, PUF, p. 22-53.
- Mayaffre D. (2006). Faut-il pondérer les spécificités lexicales par la composition grammaticale des textes ? Tests logométriques appliqués au discours présidentiel sous la Vème République. In Condé C., Viprey J.-M. *Actes des 8e Journées internationales d'Analyse des données textuelles*. Besançon : Presses universitaires de Franche Comté, II, p. 677-685.
- Monière D., Labbé C., Labbé D. (2005). Les particularités d'un discours politique : les gouvernements minoritaires de Pierre Trudeau et de Paul Martin au Canada. *Corpus*, 4, p.79-104.
- Monière D., Labbé D. (2012). Le vocabulaire caractéristique du Premier ministre du Québec J. Charest comparé à ses prédécesseurs. Dister A. et al. (éds). *Proceedings of the 11th International Conference on Textual Data Statistical Analysis*. Liège : LASLA - SESLA, p.737-751.
- Zipf G. K. (1935). *La psychobiologie du langage*. Paris : CEPL, 1974.