



HAL
open science

Radioscopies de la campagne présidentielle 2012. 9. La course de fond des candidats à l'élection présidentielle (jusqu'au 14 avril).

Dominique Labbé, Denis Monière

► To cite this version:

Dominique Labbé, Denis Monière. Radioscopies de la campagne présidentielle 2012. 9. La course de fond des candidats à l'élection présidentielle (jusqu'au 14 avril).. 2012. halshs-00717928

HAL Id: halshs-00717928

<https://shs.hal.science/halshs-00717928>

Submitted on 14 Jul 2012

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.



Radioscopies de la campagne présidentielle 2012 (IX)

La course de fond des candidats à l'élection présidentielle

(jusqu'au 14 avril)

Dominique Labbé, UMR Pacte, Sciences Po Grenoble (Dominique.Labbe@sciencespo-grenoble.fr)

Denis Monière, Université de Montréal (denis.moniere@umontreal.ca)

A partir du 10 février 2012, le site www.trielec2012.fr a publié régulièrement des "radioscopies" de la communication des principaux candidats à l'élection présidentielle française de 2012. A part des corrections de forme, ceci est la neuvième note. Elle a été mise en ligne le 20 avril.

Résumé

Les thèmes et le style des principaux candidats ont-ils changé depuis leur entrée en lice ? L'analyse conduit à trois conclusions principales.

Premièrement, les principaux candidats et leurs équipes ont privilégié la simplicité. A l'exception de M. Le Pen, ils ont adopté, dans leur discours et leurs entretiens, un vocabulaire peu diversifié et assez général, réservant le vocabulaire spécialisé aux communiqués.

Deuxièmement, tous les candidats se sont tenus aux thèmes qu'ils s'étaient fixés au départ, sans y apporter de modifications importantes en cours de campagne. F. Bayrou a présenté son programme fin janvier et l'a développé sans y apporter de modifications ou de nouveauté majeure. F. Hollande a développé pendant six mois, les thèmes présentés lors de la convention d'investiture d'octobre 2011, accentuant simplement les critiques contre N. Sarkozy en fin de campagne. N. Sarkozy a profité des vœux que le président adresse en début d'année aux différents corps constitués – puis de la crise de l'Euro – pour esquisser son programme pour un second quinquennat. En dehors de son discours de Villepinte, il n'a rien apporté de neuf à cette thématique.

Troisièmement, F. Hollande est le seul à présenter des fluctuations stylistiques importantes, comme s'il avait eu du mal à trouver son style de communication ou comme si ses rédacteurs avaient changé à plusieurs reprises. Les autres font preuve d'une stabilité remarquable qui suggère que leur communication a été confiée à un très petit nombre de collaborateurs sous une direction unique.

La campagne électorale a souvent été comparée à une course de fond ou à un marathon. En filant cette métaphore, les candidats sont des coureurs et, grâce aux intentions de votes, on a su, à tout moment, qui faisait la course "en tête" et qui était "devancé", qui progressait et qui reculait...

De la même manière, on peut également, grâce à la lexicométrie, mesurer leurs performances respectives, détecter les moments où ils répètent des choses déjà dites et ceux pendant lesquels ils trouvent une sorte de "deuxième souffle", avec de nouveaux thèmes ou de nouvelles formules.

A une semaine du premier tour, il est possible de retracer cette "course" des principaux candidats puisqu'on dispose de la quasi-totalité des messages émis pendant leurs campagnes (annexe 1). Deux mesures sont proposées pour résumer leurs choix stylistiques (première partie). Puis, on localise les ruptures dans leur communication, ce qui permet de distinguer d'éventuelles périodes (partie II).

I. DEUX INDICES

L'apparition d'un vocabulaire nouveau signale qu'on aborde une question, un thème neuf qui n'a pas été traité auparavant, c'est-à-dire une *rupture thématique* dans le corpus (pour le détail du calcul, voir annexe 2).

La spécialisation du vocabulaire mesure la capacité d'un auteur à réserver l'emploi de certains mots au traitement d'un thème particulier. Par exemple, un indice de 100 pour mille mots signifie que dans n'importe quel bloc de 1000 mots prélevé dans le corpus, 100 mots sont dédiés au traitement du (ou des) thème(s) traités dans le passage considéré et donc 900 appartiennent au vocabulaire que l'orateur utilise quel que soit le thème abordé. Un vocabulaire trop spécialisé risque de faire passer l'orateur pour un pédant. A l'inverse l'utilisation d'un vocabulaire trop général risque de faire passer l'orateur pour un homme peu sérieux ou peu au fait de ses dossiers.

En-dessous de 100 vocables spécialisés pour 1000 mots, la diversité est faible et signale une certaine propension aux généralités.

Entre 100 et 200‰, la diversité moyenne est moyenne ;

Au-dessus de 200‰ : forte propension à utiliser un vocabulaire adapté à chaque thème traité.

La diversité du vocabulaire traduit la capacité plus ou moins grande d'un auteur à varier ses expressions pour éviter les répétitions des mêmes mots à des intervalles rapprochés. Cette dimension *stylistique* est mesurée par le nombre de vocables différents employés dans de courts empan de mots dont la taille sera conservée fixe pour pouvoir comparer les résultats obtenus sur plusieurs auteurs. Par exemple, un indice de diversité du vocabulaire égal à 1 500 signifie que, dans le corpus considéré, on trouve en moyenne 1 500 vocables différents dans tous les extraits possibles longs de 10 000 mots qui peuvent être tirés de ce corpus.

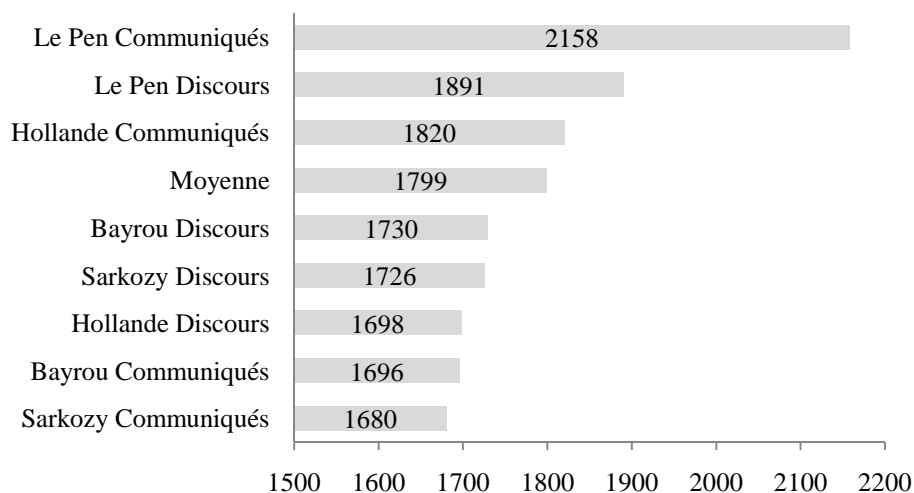
Une forte diversité signale généralement des questions auxquelles le locuteur a longuement réfléchi et pour laquelle il pense détenir la "bonne solution". Une faible diversité est symptomatique de la situation inverse. Elle peut aussi être le résultat d'un choix conscient : exposer les choses de manière simplifiées pour être bien compris. Enfin, pour un même orateur, la diversité est toujours supérieure dans un texte écrit (ou récité) par rapport à celle de ses interventions orales spontanées.

Naturellement, ces indices ne peuvent être calculés que sur des populations de textes relativement importantes. J.-L. Mélenchon et le Front de Gauche n'ont pas mis en ligne le minimum de textes qui auraient permis ces calculs qui sont donc effectués sur les quatre autres candidats. Le tableau 1 ci-dessous résume les résultats obtenus sur ces corpus.

Tableau 1. Diversité et spécialisation du vocabulaire dans les différents corpus

Corpus	Nombre de mots	Diversité (pour 10 000 mots)	Spécialisation (‰)
Discours :			
Bayrou	205 149	1517	0
Hollande	307 721	1797	0
Le Pen	82 035	1891	152
Sarkozy	248 907	1726	31
Communiqués			
Bayrou & Modem	112 371	1705	171
Hollande & PS	158 642	1820	194
Le Pen & FN	45 718	2158	150
Sarkozy & UMP	194 996	1680	210

Graphique 1. La diversité moyenne du vocabulaire (nombre de vocables différents pour 10 000 mots)



Lecture : les communiqués de M. Le Pen contiennent en moyenne 2 158 vocables

différents pour 10 000 mots. Durant cette campagne, c'est la plus grande « richesse ». A l'opposé, cette diversité n'est que de 1 680 vocables différents pour 10000 dans les communiqués de N. Sarkozy. C'est le vocabulaire le plus « pauvre ».

Ces données révèlent des choix de communication assez différents. N. Sarkozy et l'UMP ont fait le choix de la plus grande simplicité du vocabulaire. Mais, à quelques pourcents près, c'est aussi celui qui est fait par F. Bayrou, dans ses communiqués et ses discours, et par F. Hollande, pour ses discours. En revanche, le vocabulaire des communiqués du PS et de F. Hollande est de 8% plus divers que celui de l'UMP, ce qui est déjà important. M. Le Pen et surtout le FN se caractérisent par un choix opposé : le vocabulaire de leurs communiqués est de 28% plus divers que celui de N. Sarkozy et de l'UMP.

Trois des quatre principaux candidats – F. Hollande, F. Bayrou et N. Sarkozy ont fait le choix d'une spécialisation nulle ou négligeable pour leur discours et le choix inverse – spécialisation moyenne ou forte – pour leurs communiqués.

Cela peut s'expliquer de la manière suivante :

- les discours sont fait pour être entendus, à chaque fois par un public différent, hétérogène et peu défini : on s'en tient donc à des généralités, en couvrant des sujets assez larges et divers ;

- les communiqués sont ciblés. Les destinataires sont les médias et, aussi souvent, une clientèle particulière. Le sujet est retreint, la plupart du temps sur l'actualité. Ces caractéristiques obligent à spécifier le vocabulaire.

Cependant, il ne s'agit pas d'une loi générale. Au moins M. Le Pen semble y échapper : ses communiqués ne contiennent pas plus de vocabulaire spécialisé que ses discours.

II. LES RUPTURES THEMATIQUES ET STYLISTIQUES DANS LA CAMPAGNE DE CHAQUE CANDIDAT

Les candidats sont examinés par ordre alphabétique et selon deux optiques (les entretiens et les discours d'une part ; les communiqués d'autre part). On recherche dans ces deux corpus, d'éventuelles ruptures thématiques puis les fluctuations stylistiques selon une méthode présentée en annexe 2.

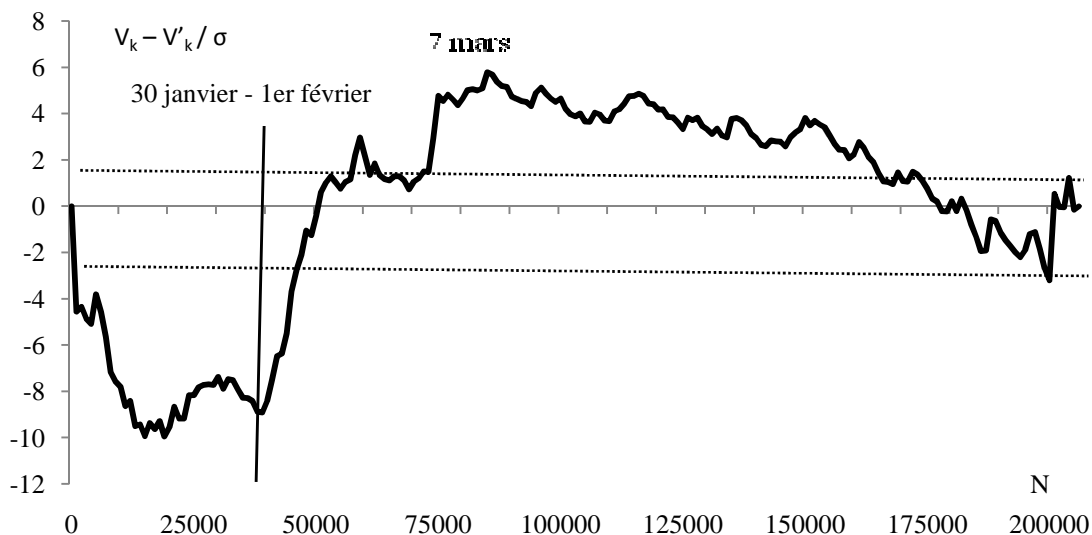
François Bayrou

Les entretiens et les discours sont classés d'un côté, les communiqués de l'autre. Ces documents sont rangés par ordre chronologique.

Les discours et les entretiens de F. Bayrou

Le corpus comporte tous les textes mis en ligne par F. Bayrou du 7 décembre 2011 au 14 avril 2012. Le graphique 2 ci-dessous retrace l'accroissement du vocabulaire durant toute cette période et localise deux périodes principales.

Graphique 2. Croissance du vocabulaire dans les discours et entretiens de F. Bayrou depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Comme expliqué en annexe 2, ce graphique décrit les fluctuations marginales et non la tendance moyenne de la croissance du vocabulaire. Par exemple, le creux au début ne signifie pas une baisse absolue mais une croissance du vocabulaire moins rapide que la tendance moyenne sur l'ensemble de la campagne. Cela permet de localiser les points de rupture majeurs : ici la fin janvier puis le début mars.

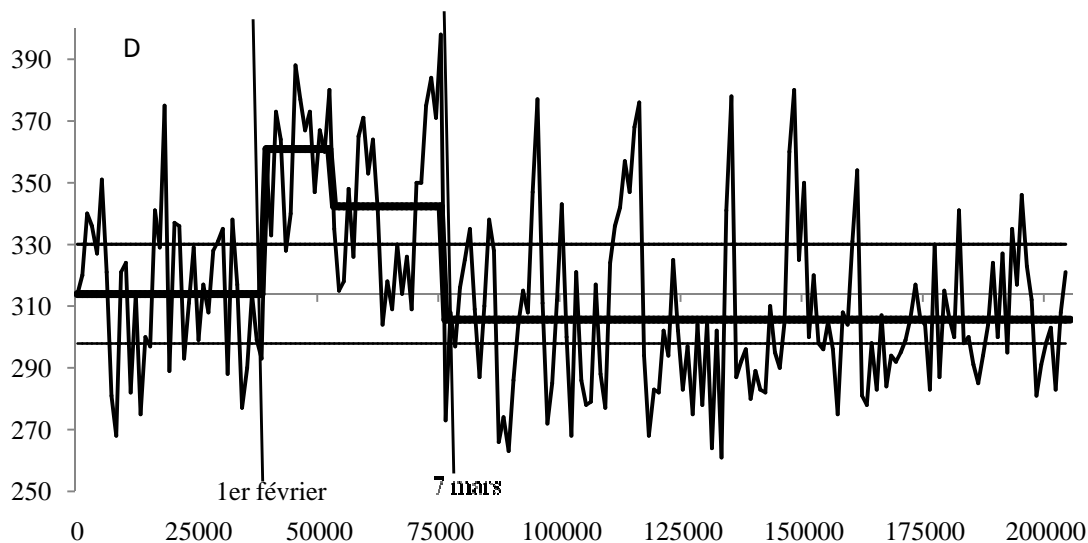
De sa déclaration de candidature à la fin janvier, F. Bayrou expose sa démarche en termes assez généraux. Une rupture se produit donc à la fin janvier.

Quels sont les discours qui produisent cette rupture ?

Le 30 janvier, F. Bayrou tient un meeting à Clermont-Ferrand pendant lequel il présente quelques-unes des propositions qu'il officialise le lendemain dans une conférence de presse à Paris. Durant tout le mois de février et jusqu'au début mars, il détaille ces propositions et apporte quelques compléments. La moralisation de la vie publique est le dernier thème à apparaître (25 février). Puis F. Bayrou multiplie les entretiens, les discours de meeting, en déclinant son programme avec des exemples particuliers – souvent empruntés à la région qu'il visite - mais sans apporter de thèmes nouveaux, ce qui explique la pente négative de la courbe.

Cette périodisation est-elle vérifiée du point de vue stylistique (graphique 3) ?

Graphique 3. Évolution de la diversité du vocabulaire dans les discours et entretiens de F. Bayrou depuis le début de sa campagne. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques (trait gras).



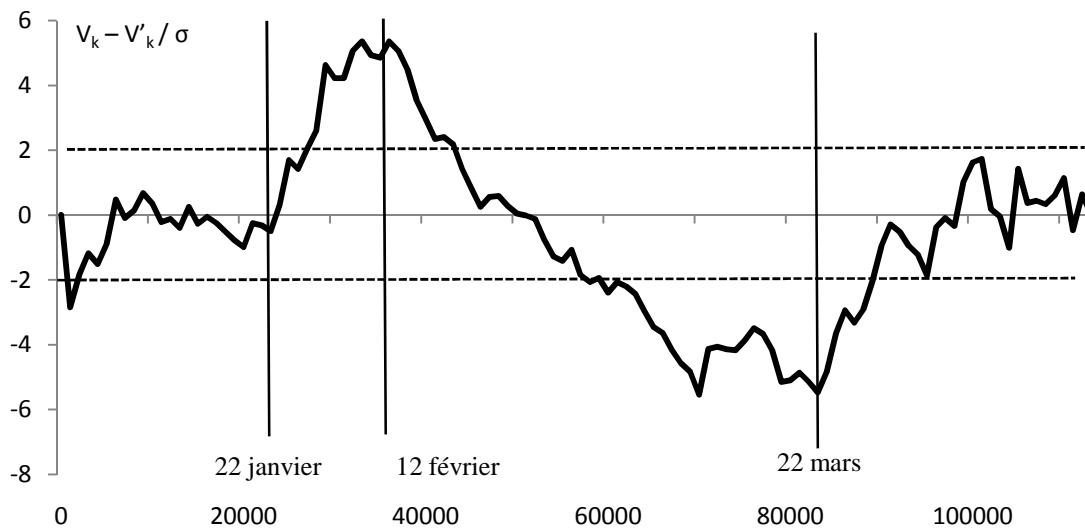
Le trait gras représente le meilleur ajustement possible de la courbe observée (trait maigre) d'apparence chaotique. L'ajustement indique que le mois de février est la seule période où la diversité est nettement plus importante que la moyenne (assez basse). Le moment d'innovation thématique est aussi celui d'une certaine qualité stylistique qui disparaît durant la période suivante dont la moyenne est encore un peu plus basse.

Devant le durcissement de la campagne et sa stagnation dans les intentions de vote, F. Bayrou a donc choisi de ne pas changer de thématique, de se répéter et de faire simple. Ces choix se sont-ils répercutés sur les communiqués émis par lui-même et par le Modem ?

Les communiqués de F. Bayrou et du Modem

Les communiqués émis par F. Bayrou et par le Modem depuis le 1^{er} janvier, sont rangés par ordre chronologique et soumis aux mêmes analyses (accroissement du vocabulaire (graphique 4) puis fluctuations stylistiques (graphique 5) ;

Graphique 4. Croissance du vocabulaire dans les communiqués de F. Bayrou et du Modem depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)

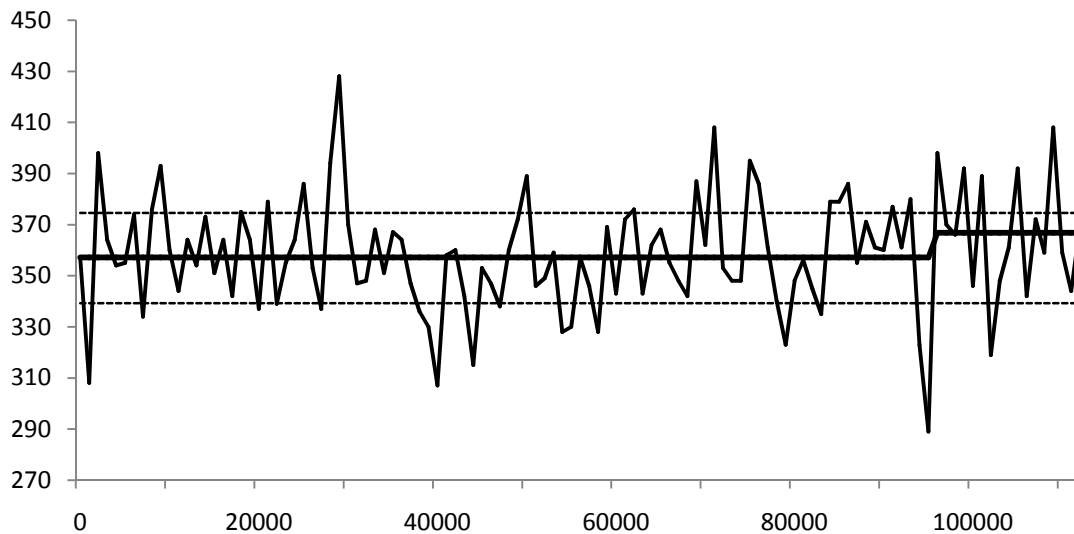


On retrouve dans les communiqués, la même périodisation que dans les discours, mais avec des variations de beaucoup plus grande ampleur. Les communiqués seraient-ils faits pour amplifier les nouveautés ? Ou bien le genre serait-il plus sensible aux variations marginales ? Si ces explications étaient les bonnes, on retrouvera ces mêmes caractéristiques dans les communiqués des autres candidats...

Le discours de Villepinte puis les événements de Toulouse ont donc eu un écho important dans les messages que F. Bayrou et son équipe ont destinés à la presse et aux internautes, alors que ces mêmes événements ne semblent pas en avoir eu sur les discours de meeting, les émissions et les entretiens accordés par le candidat (voir également Monière & Labbé g).

Il faut signaler qu'en fin de période, le Modem cesse quasiment de produire des communiqués. Ceux placés sur le site du candidat se contentent de citer et de commenter les interventions de F. Bayrou. Cet alignement progressif rejaillit sur les caractéristiques stylistiques de ces communiqués (graphique 5).

Graphique 5. Évolution de la diversité du vocabulaire dans les communiqués de F. Bayrou et du Modem depuis le 1^{er} janvier. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques (trait gras).



La lecture de ce graphique est la même que celle du graphique 3 ci-dessus.

Cette composante est marquée par une quasi-stabilité durant toute la période, avec juste une très légère augmentation durant les trois dernières semaines. Deux explications sont envisageables (et seront également examinées pour les trois autres) :

- Rédacteur unique ou petite équipe très soudée sous une direction unique ?
- Unité de genre (le communiqué de presse).

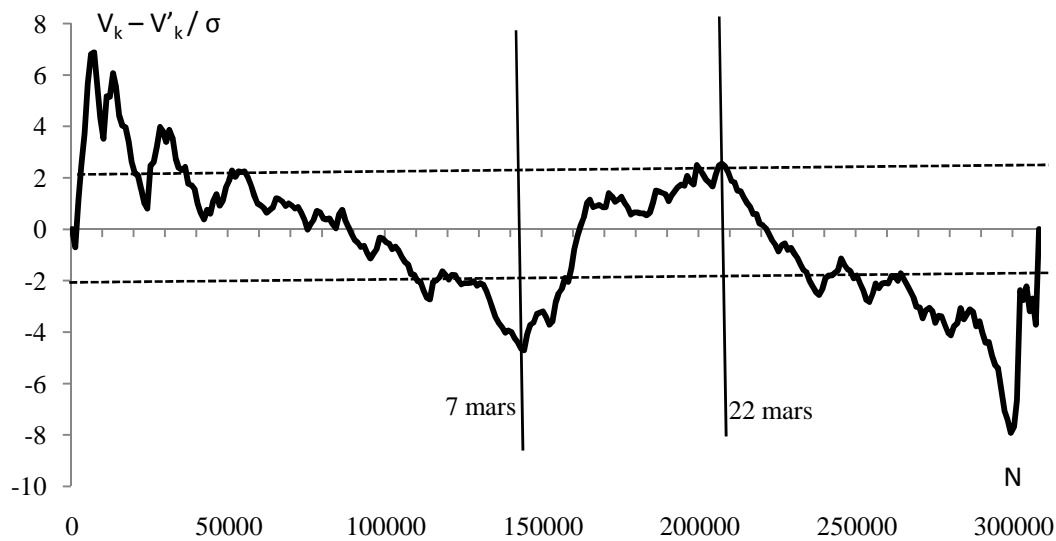
F. Hollande

La principale critique lancée contre F. Hollande par l'UMP - la versatilité voire l'inconstance - se vérifie-t-elle ?

Les entretiens et les discours de F. Hollande

L'analyse des entretiens et des discours couvre la période allant de la convention d'investiture (22 octobre 2011) et jusqu'au 14 avril 2012 (graphique 6). Pour la lecture de ce graphe, voir sous le graphique 2 ci-dessus.

Graphique 6. Croissance du vocabulaire dans les discours et entretiens de F. Hollande depuis le 22 octobre 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Plus de 8 points sur 10 se trouvent dans l'intervalle de fluctuation normale et la plupart des autres ne s'en écartent guère. Il n'y a donc pas eu de ruptures majeures au long de ces six mois. Le candidat socialiste a développé les mêmes thèmes depuis son investiture.

Au début, un fort accroissement se produit dans deux discours successifs, notamment celui prononcé le 22 octobre 2011 devant la convention d'investiture.

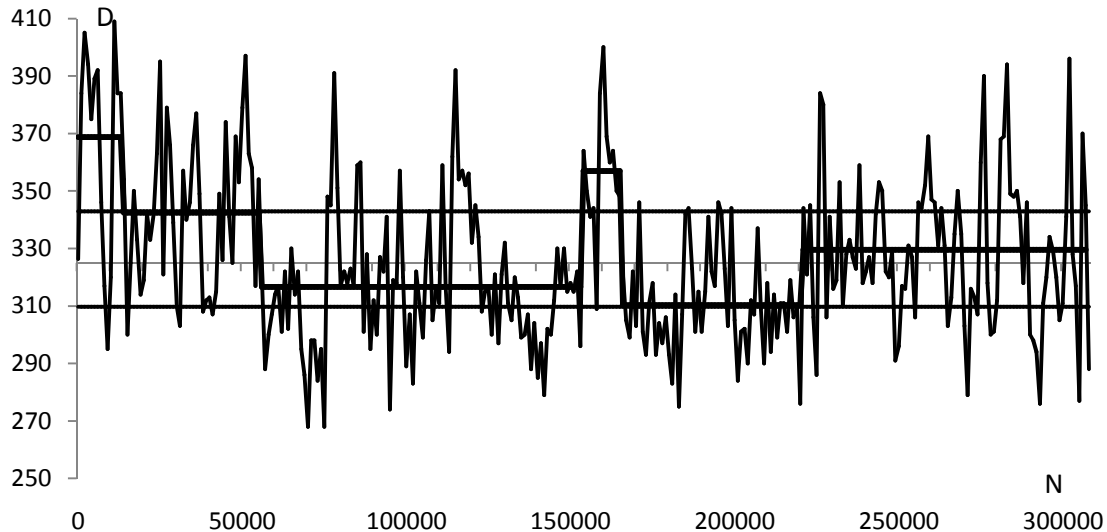
Ces deux discours semblent avoir épuisé la créativité de F. Hollande... Même le discours du Bourget (22 janvier 2012) – présenté comme fondateur par ses partisans – n'apporte en fait rien de neuf. Puis, à partir du 5-7 mars, et pendant une quinzaine, F. Hollande multiplie les discours et les entretiens, avec à chaque fois, une touche de nouveauté. Citons notamment :

- matinale d'Europe 1 (7 mars) suivant une émission télévision de N. Sarkozy ;
- un entretien télévisé le soir sur France 2 (le même soir) ;
- un discours aux Français d'outremer (10 mars) et un autre sur la défense nationale (11 mars)

Durant cette quinzaine, F. Hollande se montre plus offensif et met en cause la personnalité de son rival. Il annonce une nouvelle tranche d'impôts sur le revenu des plus riches. Mais cette créativité est malgré tout assez limitée car, le 22 mars à la fin de cette période, l'accroissement du vocabulaire n'est pas parvenu au dessus de la marge de fluctuation normale. Puis la répétition reprend.

Il est donc certain que F. Hollande a défini sa thématique de campagne dès le mois d'octobre et qu'il s'y est tenu depuis lors, se contentant, durant le dernier mois d'accentuer ses critiques envers le "bilan" et la personnalité de N. Sarkozy.

Graphique 7. Évolution de la diversité du vocabulaire dans les discours et entretiens de F. Hollande depuis le 22 octobre. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques (trait plein gras)



L'automate a localisé plusieurs ruptures délimitant six périodes (trait gras plein).

Deux encoches vers le haut sont très significatives :

- au tout début : un style très soigné et "écrit" correspondant à ses deux premiers discours déjà mentionnés ;

- mi mars : riposte au discours de N. Sarkozy prononcé à Villepinte.

La relative variabilité du style est caractéristiques de :

- corpus hétérogène : la proportion entre les discours rédigés à l'avance et ceux (plus) improvisés serait assez différente selon les moments. Les parties basses comporteraient un peu plus d'improvisations (entretiens télévisés par exemple). Les parties hautes, plus d'écrits, récités ou lus par l'orateur.

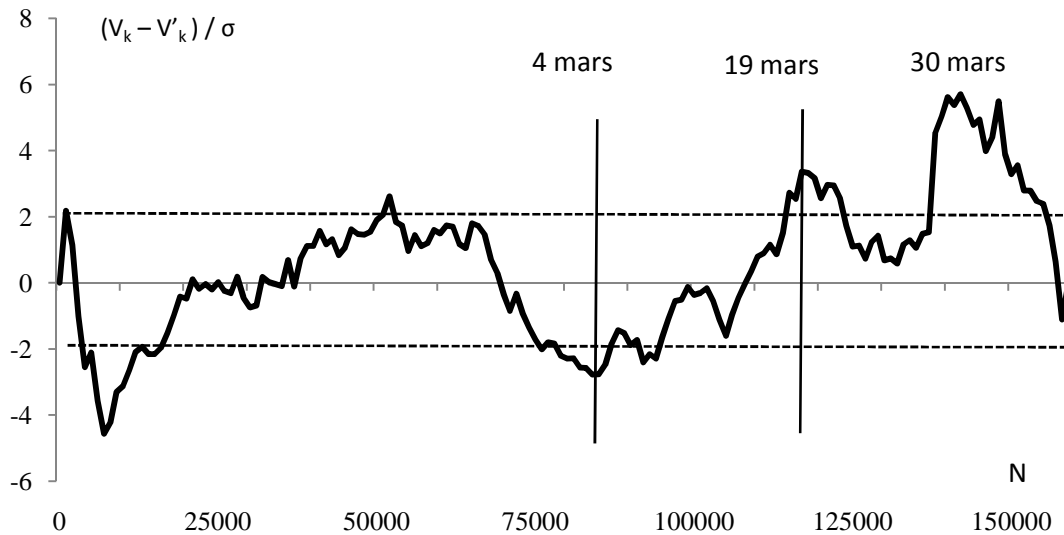
- plusieurs rédacteurs avec des pratiques diverses (ou une plume qui ne trouve pas exactement le style du patron)...

L'UMP n'a cessé de répéter que F. Hollande était ondoyant, versatile ou obscur. La variabilité formelle de ses prestations a pu donner consistance à ces reproches alors même que ces fluctuations stylistiques recouvraient une grande stabilité thématique. Ces caractéristiques se retrouvent-elles dans les communiqués émis par F. Hollande et son parti ?

Les communiqués

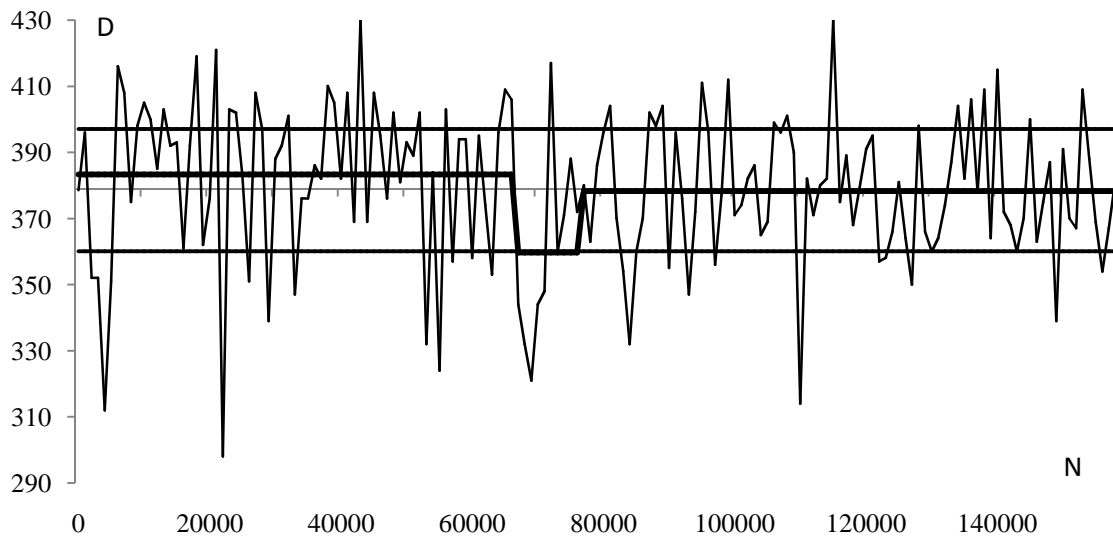
Les communiqués émis depuis le 1^{er} janvier sont rangés par ordre chronologique et soumis aux mêmes analyses (graphiques 8 et 9).

Graphique 8. Croissance du vocabulaire dans les communiqués de F. Hollande et du PS depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Comme pour les discours, le profil de l'accroissement du vocabulaire dans les communiqués montre une campagne à thématique unique du moins jusqu'au 19 mars. Une très brève période de créativité se situe entre le 22 et le 30 mars. L'examen des communiqués de cette période montre qu'il s'agit essentiellement de la critique de N. Sarkozy sur son bilan, son style que l'on qualifie d'agressif et sur le plan de la sécurité. Puis on retombe dans la répétition. Cette stabilité relative se retrouve dans la diversité du vocabulaire des communiqués (graphique 9).

Graphique 9. Évolution de la diversité du vocabulaire dans les communiqués de F. Hollande et du PS depuis le 1^{er} janvier. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques (trait plein gras)



Une douzaine de communiqués, situés au début mars, sont écrits dans un style nettement plus pauvre que la moyenne. Ils correspondent à un point bas de la courbe d'accroissement du vocabulaire (graphique... ci-dessous). Pour le reste, le style est aussi uniforme que la thématique.

Cette uniformité est d'autant plus remarquable que :

- F. Hollande de son côté ne semble pas être parvenu à une unité de style comparable (graphique).

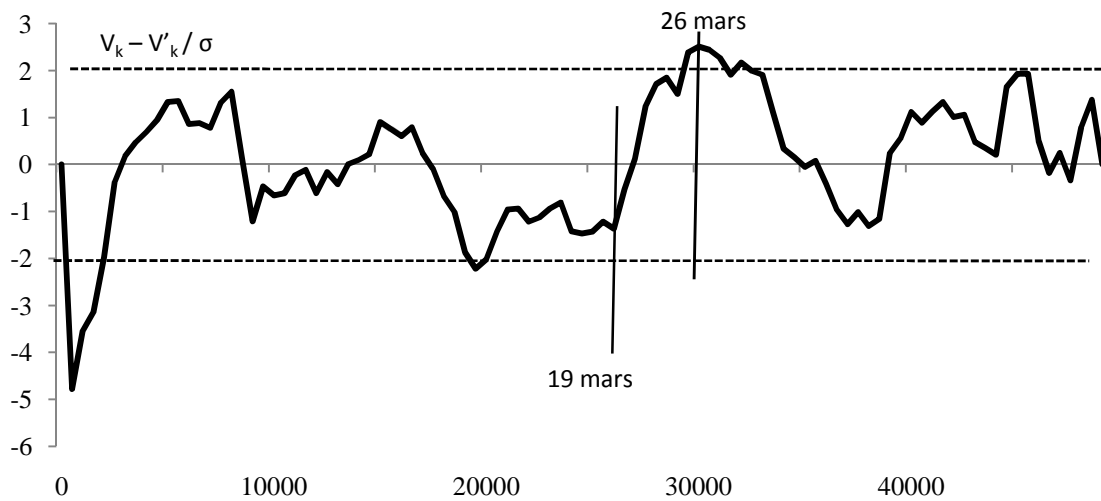
- ces communiqués sont signés par plus d'une douzaine de responsables socialistes différents. On ne peut écarter l'hypothèse d'une équipe de rédacteurs beaucoup plus réduite, voire celle d'un superviseur unique.

M. Le Pen

M. Le Pen n'a communiqué que de 21 discours et rien depuis la fin mars.

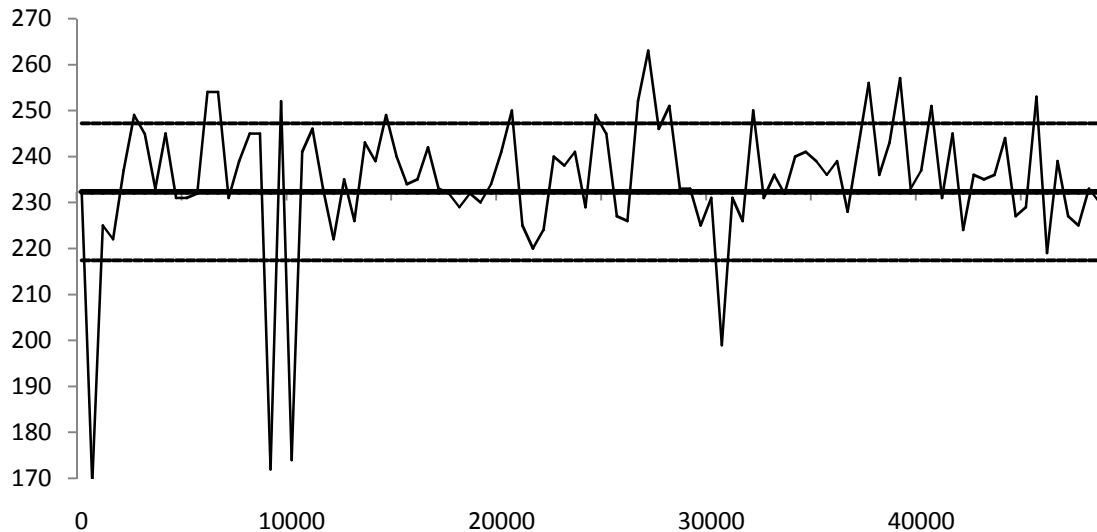
Il est donc impossible d'appliquer une recherche chronologique sur ce corpus. Restent les communiqués (graphique 10).

Graphique 10. Croissance du vocabulaire dans les communiqués de M. Le Pen et du FN depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Plus de 95% des points se trouvent dans la plage de fluctuation normale. Il n'y a donc aucune fluctuation thématique dans la communication de M. Le Pen et de son parti, tout au long de la campagne électorale. Quelques encoches brusques dans la courbe – notamment au début et vers 18 000 mots – indiquent simplement que quelques phrases sont reprises quasiment à l'identique dans deux communiqués. En dehors du début, l'accroissement le plus notable se situe durant la troisième semaine de mars et correspond aux événements de Toulouse. La composante stylistique confirme cette stabilité (graphique 11).

Graphique 11. Évolution de la diversité du vocabulaire dans les communiqués de M. Le Pen et du FN depuis le 1^{er} janvier. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques (trait plein gras)



Les trois points anormalement bas correspondent à des passages répétés à l'identique – ou quasiment. En dehors de ces points, plus de 95% des points sont situés dans l'intervalle de fluctuation normale. L'algorithme de segmentation n'a pu repérer aucun segment s'écartant significativement des autres. C'est pourquoi la droite d'ajustement (trait gras) se confond avec la moyenne.

Les communiqués du FN et de M. Le Pen sortent donc tous du même moule et probablement de la même plume, même quand ils sont signés par des responsables différents, ou bien ils ont été soigneusement relus et "formatés" par une seule personne.

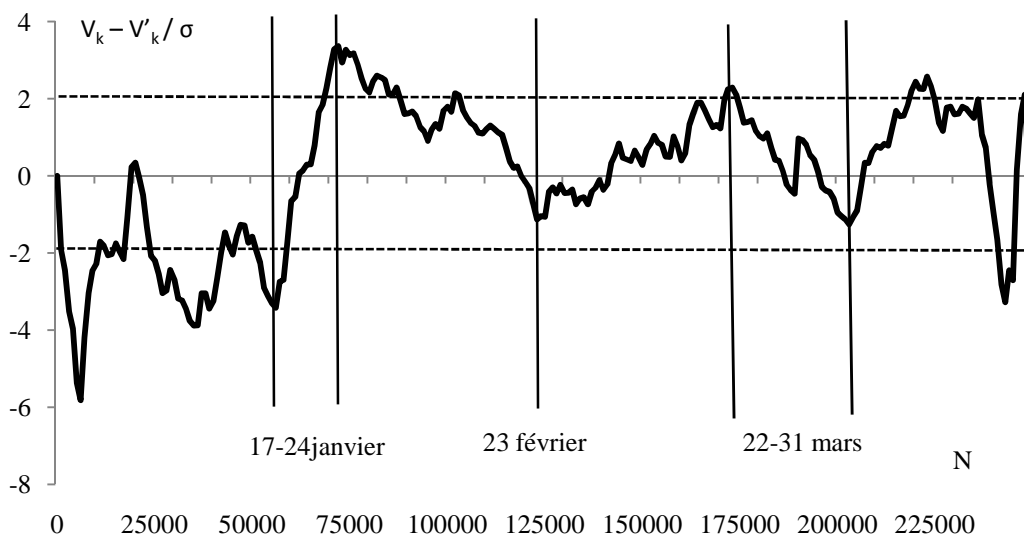
N. Sarkozy et l'UMP

Le corpus de N. Sarkozy présente une caractéristique singulière par rapport aux trois autres. Président en titre, il a pu retarder l'annonce de sa candidature bien après celle de ses rivaux. Il est donc intéressant de se demander si cette annonce a introduit une rupture dans sa communication et dans celle de son parti (UMP).

Les discours et entretiens de N. Sarkozy avant et après sa déclaration de candidature

Pour répondre à cette question, on considère toutes ses prestations publiques depuis le 1^{er} janvier 2012.

Graphique 11. Croissance du vocabulaire dans les discours et entretiens de N. Sarkozy depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Les inflexions mentionnées sur le graphique correspondent aux événements suivants :

17 janvier : vœu au monde rural

20 janvier : vœux au corps diplomatique

21 janvier : inauguration de l'aéroport de Cayenne

22 janvier : vœux à la France d'outremer

24 janvier : vœux au monde de la culture

Le 15 février, N. Sarkozy annonce officiellement sa candidature mais sa campagne ne prend vraiment son élan qu'avec les événements suivants :

23 février : meeting à Lille

24 février : inauguration du salon de l'agriculture (entretien avec Agra-presse)

28 février : meeting de Montpellier

12 mars : meeting de Villepinte

Après les événements de Toulouse (voir Labbé & Monière g), la campagne de N. Sarkozy a encore été marquée notamment par les discours suivants :

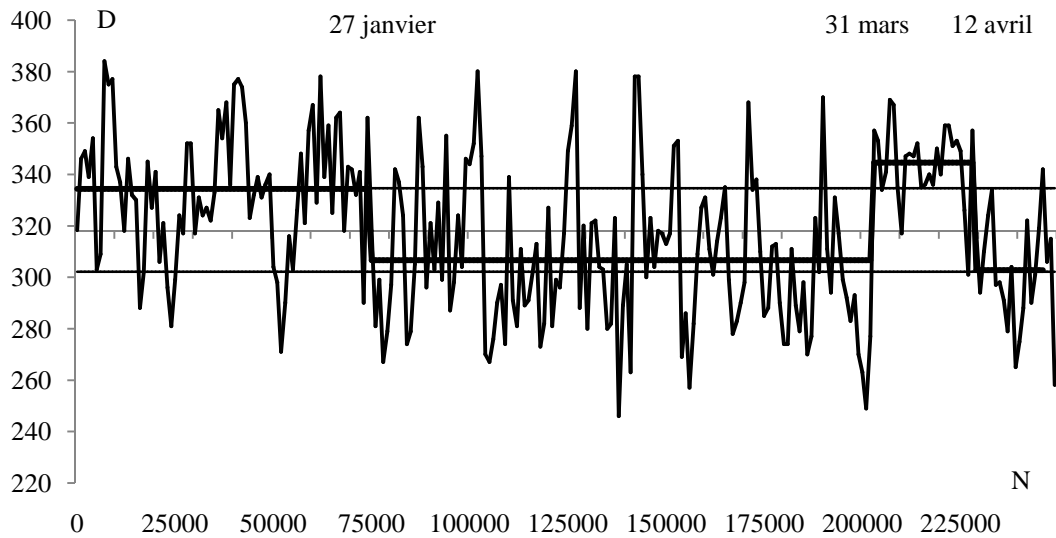
22 mars : meeting à Strasbourg (mise en cause de l'Europe de Schengen à propos de l'immigration)

24 mars : meeting à Rueil-Malmaison (respect de l'autorité)

31 mars : meeting à la Porte de Versailles (Paris) : discours à la jeunesse

Mais, au total, à partir du 17 janvier, plus de 9 points sur 10 de la courbe d'accroissement sont inclus dans l'intervalle de fluctuation normale. Il est évidemment impossible de savoir à quel moment le président a décidé d'être candidat, mais il est clair que depuis la mi-janvier, il a mis en place son discours de campagne et qu'il s'y est tenu depuis lors. L'examen de la diversité du vocabulaire confirme cette conclusion (graphique 12)

Graphique 12. Évolution de la diversité du vocabulaire dans les discours et les entretiens de N. Sarkozy depuis le 1^{er} janvier. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques.



Avant le 27 janvier, la communication de N. Sarkozy est essentiellement écrite et lue. De la fin janvier à la fin mars, il donne la priorité à la communication orale, avec une part d'improvisation significative (les pointes vers le haut correspondent toutes à des textes rédigés).

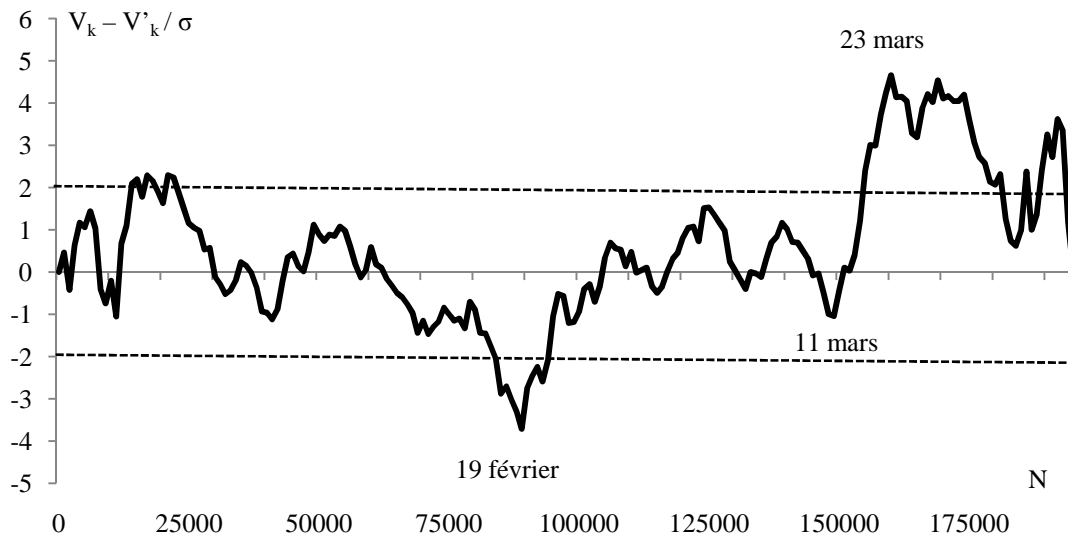
De la fin mars à la mi-avril, ses services ne communiquent plus que des textes écrits. Cette ligne de conduite culmine avec la "lettre aux Français" (5 avril). Dans la dernière période, on redonne plus de part à la spontanéité et à la simplicité.

On note enfin que le niveau moyen se situe plus bas que celui de F. Hollande, ce qui confirme le choix fait en faveur de la simplicité dans l'expression.

Les communiqués de N. Sarkozy et l'UMP

Cette stabilité dans les choix stratégiques se retrouve dans les communiqués émis par l'UMP puis par le président après son entrée en campagne (graphique 13).

Graphique 13. Croissance du vocabulaire dans les communiqués de N. Sarkozy et de l'UMP depuis le 1^{er} janvier 2012 (nombre de vocables nouveaux par tranches de 1000 mots, variable centrée et réduite)



Avant le 19 février, l'UMP est la seule à émettre des communiqués. Après l'annonce de la candidature du président, un certain nombre de communiqués sont également placés sur le site de campagne de N. Sarkozy ("lafranceforte") et ces textes sont intégrés dans la série. Il faut aussi rappeler qu'il y a, théoriquement, une bonne dizaine de rédacteurs différents à l'UMP.

Il est donc tout à fait remarquable que près de 90% des points de la courbe se trouvent dans l'intervalle de fluctuation normale. Depuis le 1^{er} janvier jusqu'à la fin de la campagne, c'est la même thématique qui a été adoptée par tous les rédacteurs de communiqués (officiels ou officiels) du président. Avec deux points d'inflexion (19 février et 11 mars).

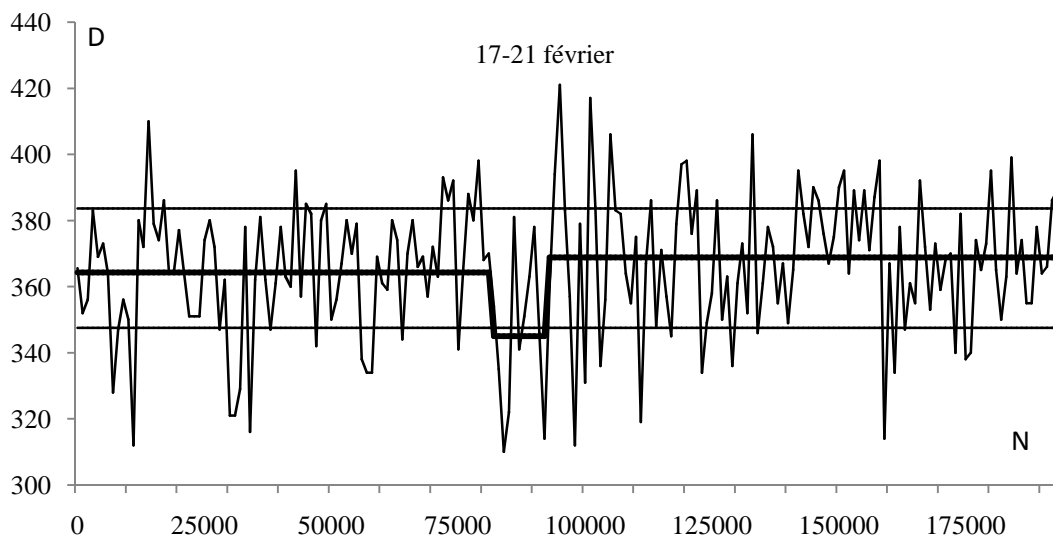
Avant l'entrée en campagne de N. Sarkozy, la courbe suit un lent déclin : l'essentiel de la communication est consacrée à la critique de F. Hollande, de manière assez répétitive (voir notamment les radioscopies 1 et 3) ;

De l'entrée en campagne de N. Sarkozy (15 février) au meeting de Villepinte (12 mars), on assiste à un renouvellement très limité.

Le meeting de Villepinte est suivi d'une brève période de renouvellement thématique qui s'interrompt au lendemain des événements dramatiques de Toulouse (Labbé et Monière h).

La conclusion essentielle est la suivante : en dehors d'une dizaine de jours – 11-23 mars -, la campagne de N. Sarkozy n'a connu aucun renouvellement thématique significatif. L'examen de la diversité du vocabulaire confirme ce constat (graphique 14).

Graphique 14. Évolution de la diversité du vocabulaire dans les discours et les entretiens de N. Sarkozy depuis le 1^{er} janvier. Variable centrée autour de la moyenne et de la plage de variation normale (traits pointillés). Segmentation automatique et localisation des principales ruptures stylistiques.



Au-delà des petites fluctuations erratiques, on retrouve la même stabilité remarquable. En dehors des 5 jours correspondant à l'entrée en campagne de N. Sarkozy et à la mise en route de son site de campagne – pendant lesquels, on constate un appauvrissement significatif d'un style déjà pauvre - la diversité du vocabulaire est remarquablement stable et basse.

Il s'agit d'un choix stylistique délibéré. Comme pour ceux émis par F. Hollande et son parti, le grand nombre apparent des signataires ne doit pas tromper. Ces textes sont certainement rédigés ou supervisés et révisés par une seule personne et celle-ci agit en étroite relation avec le candidat et les responsables de sa campagne.

Conclusions

Lorsqu'on détecte une sous-période homogène, on lui applique la technique du vocabulaire caractéristique présentée dans la note n°7 à propos des événements de Toulouse. Dans le cadre limité de cette note, il n'était pas possible de présenter ces résultats qui ne sont pas indispensables pour tirer les conclusions principales.

Du point de vue du vocabulaire, les principaux candidats et leurs équipes ont privilégié la simplicité tant dans leurs prestations orales que dans leurs communiqués. A l'exception de M. Le Pen, ils ont utilisé, dans leur discours et leurs entretiens, un vocabulaire peu diversifié et assez général, réservant les expressions plus techniques ou spécifiques aux communiqués.

Du point de vue thématique, tous les candidats se sont tenus aux thèmes qu'ils s'étaient fixés au départ, sans y apporter de modifications importantes en cours de campagne. F. Bayrou a présenté son programme fin janvier et l'a développé sans y apporter de modifications ou de nouveautés majeures. F. Hollande a développé pendant six mois, les thèmes présentés lors de la convention d'investiture d'octobre 2011, accentuant simplement les critiques contre N. Sarkozy en fin de campagne. N. Sarkozy a profité des vœux que le président adresse en début d'année aux différents corps constitués – puis de la crise de l'Euro

- pour esquisser son programme pour un second quinquennat. En dehors de son discours de Villepinte, il n'a rien apporté de neuf à cette thématique.

F. Hollande est le seul à présenter des fluctuations stylistiques importantes, comme s'il avait eu du mal à imprimer une marque personnelle à sa communication ou comme si ses rédacteurs avaient changé à plusieurs reprises. Les autres candidats font preuve d'une stabilité stylistique remarquable qui suggère que leur communication a été confiée à un très petit nombre de collaborateurs sous une direction unique.

De ces choix stratégiques, il est résulté une campagne monotone et sans grand relief. Dans la conjoncture économique et financière présente, ces choix étaient sans doute justifiés. Mais les électeurs ont-ils été intéressés par cette campagne sans surprise et dominée par les invectives échangées par les principaux candidats ?

Remerciements

Denis Monière a pu effectuer cette étude grâce à l'invitation de l'Institut d'Etudes Politiques de Grenoble et du laboratoire PACTE.

Les programmes informatiques utilisés pour ces analyses ont été réalisés avec l'aide de Cyril Labbé (Laboratoire d'Informatique de Grenoble – Université Joseph Fourier). Bernard Denni (PACTE) a relu ces notes et nous a fait de très utiles commentaires de forme et de fond. La mise en ligne a été assurée par Annie Salomon (PACTE).

Bibliographie

Arnold Edward (2005). Le discours de Tony Blair. *Corpus*, 4, 2005, p 55-78.

Hubert Pierre & Dominique Labbé (1988). Un modèle de partition du vocabulaire. In Labbé Dominique, Seran Daniel & Thoiron Philippe (eds.) : *Etudes sur la richesse et la structure lexicales*. Paris-Genève : Champion-Slatkine, 1988, p. 93-114.

Labbé Cyril, Labbé Dominique et Hubert Pierre (2004). "Automatic Segmentation of Texts and Corpora". *Journal of Quantitative Linguistics*. December 2004, 11-3. p 193-213 (texte consultable sur HAL-SHS).

Labbé Cyril & Labbé Dominique (2010). La diachronie dans le discours politique. Le général de Gaulle. In Banks David (Ed). *Aspects diachroniques du texte de spécialité*, Paris, l'Harmattan, 2010, p. 129-148 (consultable sur HAL-SHS).

Labbé Dominique & Monière Denis (2012a). *Radioscopies de la campagne présidentielle. 1. La pré-campagne (1er janvier - 4 février)*. Note de recherche publiée sur le site www.trielec2012.fr.

Labbé Dominique & Monière Denis (2012b). *Radioscopies de la campagne présidentielle. 2. Les mots et les thèmes de la pré-campagne (1e janvier -11 février 2012)*. Note de recherche publiée sur le site www.trielec2012.fr.

- Labbé Dominique & Monière Denis (2012c). *Radioscopies de la campagne présidentielle. 3. L'entrée en scène du Président sortant (5-18 février). (1er janvier -11 février 2012)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Labbé Dominique & Monière Denis (2012d). *Radioscopies de la campagne présidentielle. 4. Sale mec ou gentil garçon ? Portraits croisés des principaux candidats (1er janvier-25 février 2012)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Labbé Dominique & Monière Denis (2012e). *Radioscopies de la campagne présidentielle. 5. Faire campagne contre les autres ? (19 février-3 mars)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Labbé Dominique & Monière Denis (2012f). *Radioscopies de la campagne présidentielle. 6. L'effet Villepinte (4-17 mars)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Labbé Dominique & Monière Denis (2012g). *Radioscopies de la campagne présidentielle. 7. Les attentats de Montauban et de Toulouse : un tournant dans la campagne électorale ? (18-24 mars 2012)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Labbé Dominique & Monière Denis (2012h). *Radioscopies de la campagne présidentielle. 8. La spirale de la négativité (25 mars au 7 avril)*. Note de recherche publiée sur le site www.trielec2012.fr.
- Monière Denis & Labbé Dominique (2003). *Le vocabulaire gouvernemental. Canada, Québec, France (1945-2000)*. Paris : Champion.
- Monière Denis & Labbé Dominique (2008). *Les mots qui nous gouvernent. Le discours des premiers ministres québécois : 1960-2005*. Montréal : Monière-Wollank Editeurs. (ouvrage consultable sur HAL-SHS).
- Monière & Labbé (2010). "Segmentation des corpus chronologiques : 143 ans de discours gouvernemental au Québec". In Bolasco Sergio, Chiari Isabella, Giuliano Luca (Eds). *Proceedings of 10th International Conference Statistical Analysis of Textual Data*. Rome : Edizioni Universitarie di Lettere Economia Diritto, Vol 2, p. 805-816 (document consultable sur HAL-SHS).

Annexe 1
Le corpus et les principaux sous-corpus

(au 7 mars 2012)

		Nombres textes	Nombres mots	Vocabulaire
Bayrou	Communiqués Bayrou	119	68 070	5 035
	Communiqués Modem	103	44 301	3 846
	Discours et entretiens	50	205 149	6 809
Total Bayrou		272	317 520	9 069
Hollande	Communiqués Hollande	101	34 727	3 641
	Communiqués PS	369	123 915	7 304
	Discours et entretiens	76	307 721	7 377
Total Hollande		546	466 363	10 551
Le Pen	Communiqués Le Pen	59	11 322	2 306
	Communiqués FN	159	34 396	4 590
	Discours et entretiens	21	82 035	6 091
Total Le Pen		239	127 753	8 144
Mélenchon	Communiqués Mélenchon	33	3 457	1 054
	Communiqués Front Gauche	94	25 461	3 662
	Discours et entretiens	2	14 244	1 896
Total Mélenchon		129	43 162	4 630
Sarkozy	Communiqués Sarkozy	66	33 101	3 461
	Communiqués UMP	623	161 895	7 164
	Discours et entretiens	66	248 907	7 713
Total Sarkozy		755	443 903	10 665
Total		1 941	1 398 701	18 162

Annexe 2

Méthode

Les textes sont rangés par ordre chronologique pour former un corpus "ordonné" de taille N mots qui est découpé en K tranches - ici la taille des tranches est de 1000 mots, sauf pour M. Le Pen où elles sont de 500 mots - et à observer, à la fin de chaque tranche, le nombre des V_k vocables différents apparus depuis le début du texte.

Grâce à ces observations, l'algorithme calcule le poids relatif de deux vocabulaires :

- *vocabulaire général* de l'auteur qu'il utilise quelles que soient les circonstances et les thèmes traités. Cette proportion de texte est considérée comme le résultat d'un tirage exhaustif (sans remise) parmi ce vocabulaire général. Cette composante de l'accroissement du vocabulaire est décrite par la loi hypergéométrique ;

- un ou de plusieurs *vocabulaires spécialisés* - de l'économie, de la finance, des relations internationales... - que l'auteur mobilise en fonction des circonstances et des thèmes qu'il aborde. Cette composante de l'accroissement du vocabulaire est décrite par un ajustement linéaire.

On obtient ainsi l'équation de la courbe d'ajustement du phénomène : celle qui décrit son profil théorique en l'absence de toute variation aléatoire ou conjoncturelle. A chaque observation (V_k) est associé une valeur théorique (V'_k).

Deux questions se posent alors :

- l'écart entre les deux valeurs est-il significatif ? Ou bien peut-il être considéré comme "normal", c'est-à-dire sans autre origine probable que le hasard ?

- en cas d'écart significatif, s'agit-il d'un événement unique où bien se répète-t-il pour constituer une variation significative par rapport à la tendance ?

Pour répondre à ces deux questions,

- les observations sont centrées sur la tendance ($V_k - V'_k$). La courbe idéale devient une droite horizontale et les valeurs observées se répartissent autour d'elle. Si l'observation est supérieure à ce qui est attendu, c'est qu'en ce point, il est apparu un excédent de mots nouveaux par rapport au rythme moyen. L'orateur aborde probablement un sujet nouveau. A l'inverse, un point inférieur au rythme moyen signale un épuisement relatif du vocabulaire : l'orateur redit des choses qui ont déjà été énoncées auparavant.

- on calcule la variabilité "normale" de la variable observée, donnée ici par l'écart-type de la loi hypergéométrique (σ). Tout écart supérieur à $1,96\sigma$ signale une déviation significative (avec un risque d'erreur inférieur à 5%).

Pour faciliter la lecture des données, la valeur centrée est divisée par l'écart type ($(V_k - V'_k) / \sigma$). On obtient la **variable centrée et réduite** ou "normée". C'est celle qui est reproduite dans les graphiques de cette note.

Si le résultat est supérieur à $+ 2\sigma$ ou inférieur à $- 2\sigma$, l'opérateur sait immédiatement qu'il peut conclure avec très peu de chances d'erreur à une déviation significative.

En calculant les pentes locales de cette courbe centrée et réduite, on peut localiser la présence de ruptures thématiques et découper le corpus en autant de séquences homogènes.

Références

- Le modèle proprement dit a été présenté pour la première fois dans : Hubert & Labbé (1988).
- Les calculs – de l'accroissement, de la diversité du vocabulaire et de la segmentation automatique - et leurs propriétés sont présentés et discutés dans : Labbé, Labbé et Hubert (2004).
- Pour des exemples d'applications au discours politique : Arnold Edward 2005 ; Monière & Labbé (2003) ; Labbé & Labbé (2010) ; Monière & Labbé (2008).
- Le repérage des ruptures thématiques et le découpage de corpus en séquences temporelles homogènes sont décrits dans : Monière & Labbé (2010).