HAL
open science

Un siècle et demi de discours gouvernemental au Canada Contribution de la lexicométrie à l'Histoire politique

Denis Monière, Dominique Labbé

To cite this version:

Denis Monière, Dominique Labbé. Un siècle et demi de discours gouvernemental au Canada Contribution de la lexicométrie à l'Histoire politique. 12th International Conference on Textual Data Statistical Analysis., Jun 2014, Paris, France. pp.485-494. hal-01154716

HAL Id: hal-01154716

https://hal.science/hal-01154716

Submitted on 22 May 2015

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.

Un siècle et demi de discours gouvernemental au Canada Contribution de la lexicométrie à l'Histoire politique

Denis Monière[1], Dominique Labbé[2]

[1] Université de Montréal (denis.moniere@umontreal.ca)

[2] Institut d'Etudes Politiques de Grenoble (dominique.labbe@iep-grenoble.fr)

Abstract

The Canadian Prime Minister opens each session of the Ottawa federal parliament with a speech of the "throne". From the origin of these institutions (1867), there are 128 speeches and a total of 260,836 words. The automatic segmentation of the corpus highlights two major turning points defining three periods that are characterized by their own topics and by their own vocabularies. The first turning point occurs at the beginning of the Second World War when the central government extended considerably, and the second one occurred in 1968 with the arrival in power of Trudeau who wants to build a nation and a strong federal government. Within these three main periods, there are several important events which also delimit some sequences. Thus, lexicometry provides some useful tools for the periodization of political history.

Résumé

Le Premier ministre canadien ouvre chaque session du parlement d'Ottawa par un discours du "trône" soit, depuis l'origine de ces institutions (1867), 128 discours comportant 260 836 mots. La segmentation automatique de ce corpus met en lumière deux tournants majeurs délimitant trois périodes dont les thèmes propres sont déterminés grâce à leurs vocabulaires caractéristiques. Le premier tournant correspond à la seconde guerre mondiale et à l'après-guerre où les activités du gouvernement central s'étendent considérablement ; le deuxième survient en 1968 avec l'arrivée au pouvoir de Trudeau qui veut bâtir une nation canadienne et un pouvoir fédéral fort. Au sein de ces trois périodes principales, on distingue plusieurs épisodes secondaires importants dont on établit également les vocabulaires caractéristiques. Ainsi, la lexicométrie fournit des outils intéressants pour la périodisation de l'histoire politique.

Keywords: lexicometry ; vocabulary growth ; segmentation of large corpora ; Canada ; political discourses

In Née Emilie, Daube Jean-Michel, Valette Mathieu, Fleury Serge (dir.). *Proceedings of the 12th International Conference on Textual Data Statistical Analysis*. Paris : June 3-6 2014, p 485-494.

1. Introduction

L'analyse lexicométrique de vastes corpus diachroniques est un outil précieux pour l'étude des langues (Bowker & Pearson 2002). Elle peut aussi venir à la rescousse de l'histoire politique et lui offrir le moyen de déterminer les moments de rupture et de stabilité dans l'évolution de la communication politique. Par exemple, les discours des huit premières années du Premier ministre anglais T. Blair (la méthode met en valeur le tournant décisif induit par la guerre en Irak : Arnold 2005) ou l'évolution de la presse écrite envers le principe de précaution (Delanoë 2010).

Nous avons présenté en 2010, une méthode de segmentation entièrement automatique et une première expérience sur les discours gouvernementaux québécois de 1867 à 2010 (Monière & Labbe 2010b). Une fois le corpus découpé en périodes homogènes, leur vocabulaire caractéristique peut-être établi, ce qui permet de mettre en valeur les singularités de chacune (Monière & Labbé 2012).

La présente communication appliquera cette démarche à un nouveau corpus qui vient s'ajouter à notre bibliothèque des discours politiques (annexe 1) : soit l'ensemble des 128 discours du trône du gouvernement canadien depuis la fondation de la confédération canadienne, par l'acte de l'Amérique du nord britannique de 1867 (annexe 2).

La production de ces textes s'étale donc sur près d'un siècle et demi, dans des conditions formelles d'énonciation inchangées : ouverture de la session parlementaire par un discours prononcé par le représentant du souverain britannique mais écrit par le gouvernement pour présenter son programme législatif. Cette stabilité institutionnelle rend ce corpus particulièrement intéressant pour mesurer l'impact de l'histoire sur le discours politique.

2. Corpus et méthodes

Les discours de 1945 à nos jours avaient déjà été recueillis dans le cadre de notre analyse sur les discours gouvernementaux (Monière & Labbé 2003). Nous avons depuis complété ce corpus en ajoutant les discours prononcés de 1867 à 1944 en dépouillant le *Journal des débats*. Nous avons retenu la version française de ces discours qui a même valeur officielle que la version anglaise.

Chacun de ces textes a été traité selon les mêmes procédures : balisage, correction et standardisation orthographiques, lemmatisation selon les normes présentées dans Labbé (1990). Les 128 discours ainsi traités comportent au total 260 836 mots et un vocabulaire de 6 901 vocables différents.

Pour segmenter ce corpus en périodes homogènes, l'accroissement du vocabulaire est ajusté par sa tendance (Labbé, Labbé & Hubert 2004). Ensuite, un algorithme de segmentation associé à des tests de validité donne le découpage optimal du corpus (Monière & Labbé 2010). Une série d'indicateurs mesurent l'ampleur des mouvements de vocabulaire caractérisant chacune des périodes.

3. Localisation des principales ruptures

3.1. Vue d'ensemble

Le corpus est découpé en T tranches de 200 mots (N_1 = 200 ; N_2 = 400 … N_{1035} = 206 836). Pour chacune de ces tranches, le nombre de vocables apparus depuis le début du corpus (V_t) est centré sur la tendance calculée à l'aide du modèle de partition du vocabulaire (Labbé &

Hubert 1988) puis rapporté à l'écart-type (variable centrée et réduite : V'_t). Dans la figure 1 ci-dessous, l'accroissement moyen du vocabulaire est figuré par l'axe horizontal au milieu du graphique ; le trait gras représente les valeurs observées centrées et réduites.

Lorsque le trait gras est orienté vers le haut, cela signifie que se produit à cet endroit un afflux de mots nouveaux (on suppose qu'il s'agit d'un tournant marqué par l'apparition de nouveaux thèmes) ; à l'inverse, une pente descendante indique que les orateurs répètent des choses déjà dites auparavant.

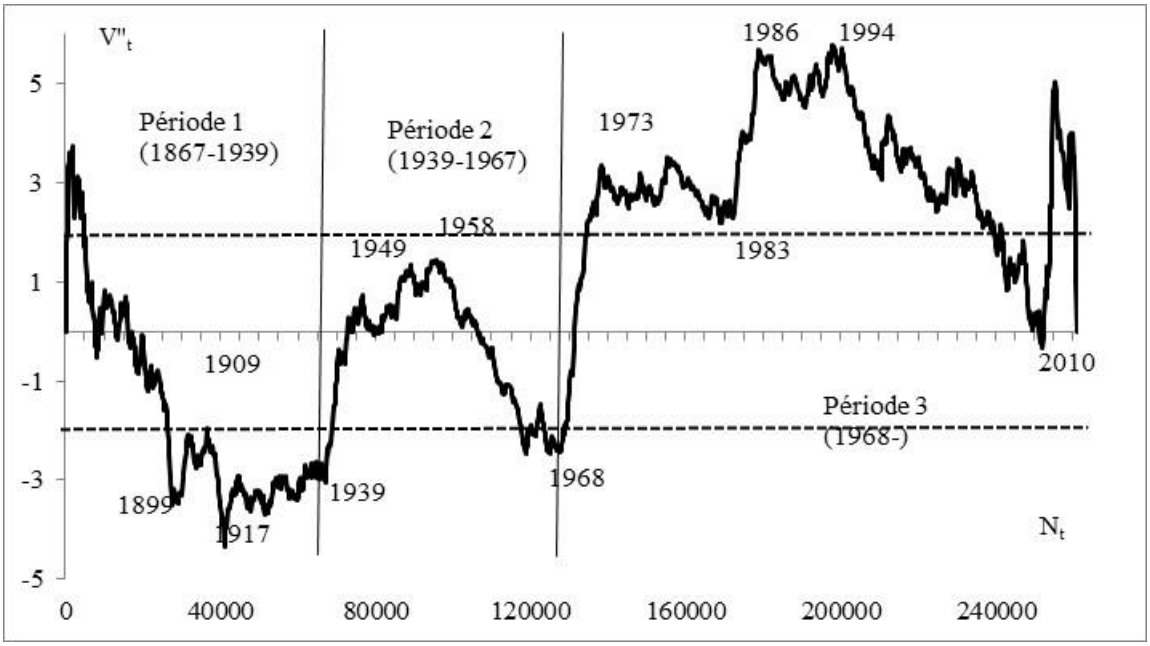

Figure 1.Accroissement du vocabulaire dans les discours du trône (valeurs centrées et réduites)

Deux ruptures majeures – septembre 1939 et 1968 – délimitent trois périodes principales. La troisième période semble plus longue qu'elle n'est en réalité car, à partir de 1968, la longueur des discours s'allonge nettement (on passe d'une moyenne de 1 200 mots par discours à plus de 3 000). Chacune de ces périodes s'ouvre par une séquence d'innovation lexicale – courbe ascendante – suivie de séquences de faible innovation (courbe proche de l'horizontale), voire de simple répétition (courbe descendante).

Il y a également des ruptures secondaires, délimitant des sous-périodes ou "séquences", notamment :
– 1899-1909 : correspondant aux gouvernements libéraux de W. Laurier ;
– 1984-1991 correspondant à l'arrivée au pouvoir du Parti conservateur dirigé par B Mulroney, après 20 ans de règne libéral (aucun discours du trône en 1992 et 1993).

Dans le cadre limité de cette communication, nous nous limiterons aux deux dernières périodes (1939-1967 et 1968-2013).

3.2. Segmentation automatique des périodes

La segmentation automatique permet de distinguer 11 séquences, trois durant la seconde période et huit séquences durant la troisième. Dans la Figure 2, le trait maigre représente les données centrées sur la tendance moyenne et réduites, les segments de droite en gras l'ajustement de ces données centrées et réduites.

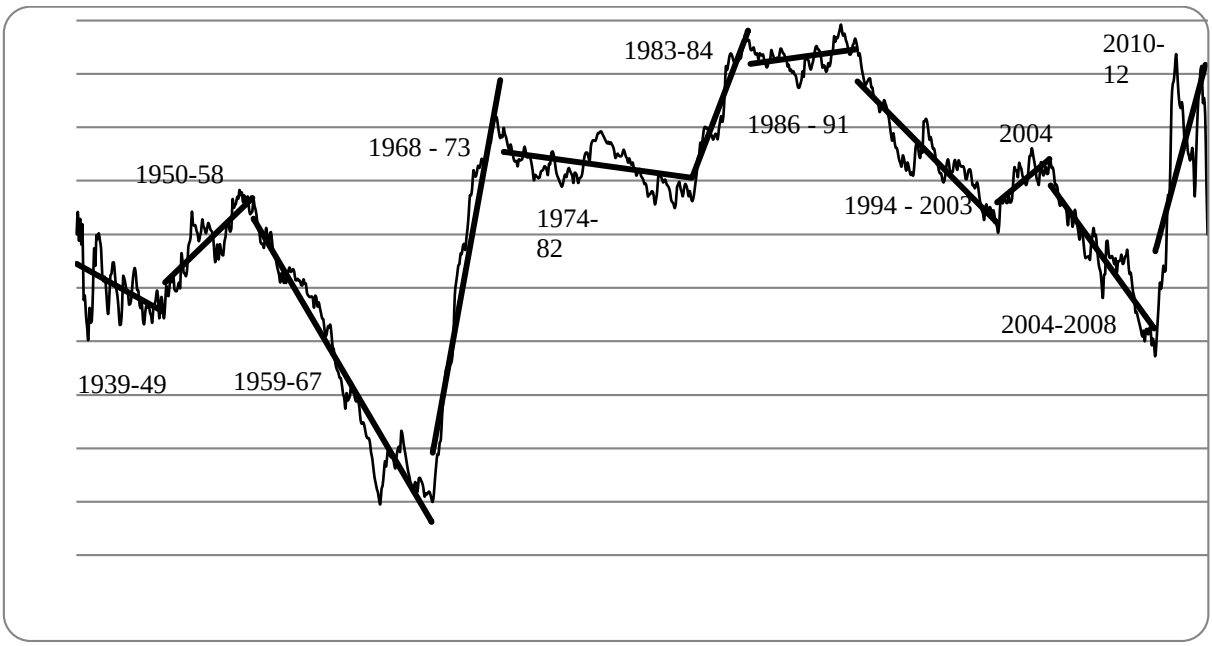

Figure 2. Segmentation automatique de l'accroissement du vocabulaire dans les discours du trône depuis 1939 (valeurs centrées et réduites et ajustements linéaires de ces valeurs)

On a imposé au logiciel deux contraintes : respecter les frontières entre les discours (± 200 mots) et donner à chaque séquence une longueur minimale de 10 000 mots. Ces deux contraintes expliquent pourquoi le discours de février 2004 prononcé par P. Martin juste avant la dissolution de la chambre constitue à lui seul une séquence (la n° 9), alors que d'autres ne franchissent pas ces contraintes (comme le premier discours de L. Pearson ou celui de J. Chrétien en 1999 – eux aussi marqués par des innovations mais trop brèves pour introduire une rupture dans la tendance générale à la baisse).

La tendance de l'accroissement lexical dans chacune des séquences est donnée par le coefficient directeur de la droite d'ajustement (tableau 1 et 2).

Séquences	Coefficient directeur de la droite d'ajustement	Coefficient de détermination (R^2)	Nombre de points (ddl)
1	-0,012	0,254	75
2	+0,021	0,717	77
3	-0,037	0,944	154
4	0,116	0,917	61
5	0,003	0,266	162
6	0,057	0,906	49
7	0,003	0,115	92
8	-0,019	0,811	121
9	0 ,018	0,536	46
10	-0,030	0,870	90
11	0,098	0,721	44

Tableau 1. Coefficient directeur et qualité de l'ajustement pour chacune des séquences.

N°	Début	Fin		Tendance
1.	sept 1939	1949	Mackenzie King	↘
2	1950	1957	Saint-Laurent	↗
3	1958	1967	Diefenbaker / Pearson	↘
4	1968	1973	Trudeau 1	↗
5	1974	1983	Trudeau 2 / Clark	→
6	1984	1986	Mulroney 1	↗
7	1988	1991	Mulroney 2	→
8	1994	2003	Chrétien	↘
9	fév.2004	fév.2004	Martin 1	↗
10	oct. 2004	2008	Martin 2 / Harper 1	↘
11	2010	2012	Harper 2	↗

Tableau 2. Délimitation chronologique des séquences homogènes

Trois situations sont possibles :
– coefficient négatif (l'apport en vocabulaire nouveau est moins important que la moyenne : période de répétition) ; les séquences 3 puis 10 sont les plus marquées ;
– coefficient positif (apport supérieur à la moyenne : période d'innovation), dans l'ordre : séquences 4 puis 11 puis 6 ;
– un coefficient quasi nul (variable stationnaire : apport moyen) : séquences 5, 7.

Qualité de l'ajustement

La qualité de l'ajustement est donnée par le coefficient de détermination de V'_t par N_t – c'est-à-dire la détermination du vocabulaire par le temps - et par le nombre de points compris dans la séquence (tableau 1). Ce coefficient donne une indication concernant l'homogénéité de la séquence.

L'ajustement est bon ou excellent pour 7 des dix séquences (2, 3, 4, 6, 8, 10 et 11). On peut donc postuler que les discours tenus pendant chacune de ces séquences sont homogènes et partagent les mêmes thèmes.

A l'inverse, l'ajustement est médiocre pour 3 séquences (1, 5, 7). Dans ces trois cas, le coefficient directeur est quasiment nul : ce sont manifestement les cas les plus difficiles à traiter. Les apports de vocabulaire, comme les passages répétitifs sont trop brefs pour que l'algorithme puisse les segmenter, de telle sorte que, à l'échelle choisie, la stabilité du phénomène est l'hypothèse la plus acceptable. Mais un grain plus fin pourrait peut-être permettre de repérer plusieurs sous-séquences.

La séquence 9 illustre les difficultés de l'exercice (figure 3). Il s'agit du premier discours de P. Martin qui vient de remplacer J. Chrétien (juste avant les élections de 2004). Etant donné le nombre de points inclus dans le calcul, le coefficient de détermination est à la limite du seuil de signification, de telle sorte que l'hypothèse d'homogénéité peut être retenue.

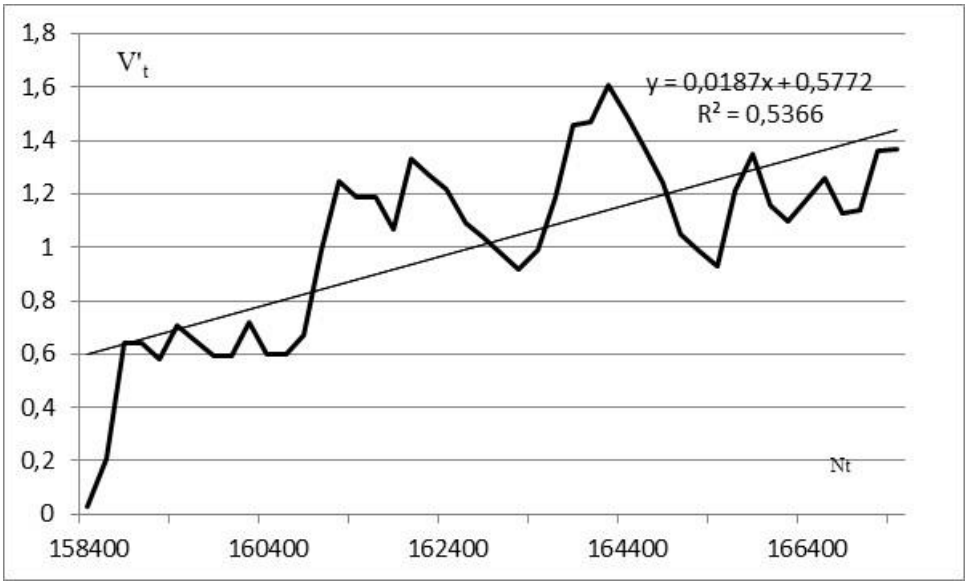

Figure 3. Discours de février 2004 (Martin) : accroissement du vocabulaire (variable centrée et réduite) et ajustement linéraire.

C'est un discours de plus de 10 000 mots – ce qui permet à l'algorithme de le distinguer des précédents et des suivants. Ce discours est manifestement un collage de plusieurs textes dont certains apportent des éléments nouveaux – portions de courbe ascendantes – alors que d'autres passages reprennent des thèmes déjà présents dans d'autres discours antérieurs (portions descendantes). Il présente les caractéristiques du discours électoral (Labbé & Monière 2010a), notamment la faible personnalisation, la valorisation du "nous", la prédominance du verbe et des adverbes (valorisation de l'action sur les idées), une forte densité des chiffres et des noms propres, enfin une forte dimension polémique.

En conclusion sur ce découpage : la plupart des séquences sont délimitées par un changement de majorité parlementaire, mais tous les changements de majorité n'introduisent pas de rupture. Le cas le plus frappant concerne la période 3. Pour cette période il y a 5 ans de gouvernement conservateur majoritaire (J. Diefenbaker de 1958 à 1963) et 5 ans de gouvernement libéral minoritaire (L. Pearson de 1963 à 1968). Pour cette période, le coefficient de détermination est pratiquement égal à 1. L. Pearson semble donc en continuité avec J. Diefenbaker qui lui-même reprend la thématique mise en place durant la séquence 1950-57 (L. Saint-Laurent, libéral) qui a manifestement été une période d'innovation.

L'étude du vocabulaire et des thématiques permet de caractériser le contenu de chacune des séquences ainsi délimitées.

4. Vocabulaires et thématiques caractéristiques des principales séquences chronologiques

L'algorithme utilisé pour calculer le vocabulaire et les phrases caractéristiques d'un locuteur a été présenté dans Monière et Labbé 2012. Cet algorithme utilise le calcul des "spécificités du vocabulaire" (Lafon, 1984) qui a été adapté pour tenir compte du poids des catégories grammaticales et des classes de fréquences. Une fois connu le vocabulaire caractéristique de la séquence (par rapport au reste du corpus), l'algorithme attribue à chaque phrase (de cette séquence) un score déterminé en fonction du nombre de vocables caractéristiques qu'elle contient. Les phrases les plus caractéristiques donnent une synthèse des thèmes qui singularisent cette séquence.

1939-1949

Les discours tenus après septembre 1939 sont caractérisés par de nouvelles thématiques reliées à l'effort de guerre qui entraîne de nombreuses mesures législatives concernant la mobilisation des ressources matérielles et humaines dont la tenue d'un plébiscite en 1942. On observe un suremploi de noms propres et de mots étrangers pour désigner les belligérants ainsi qu'un afflux de substantifs reliés à l'effort de guerre : *guerre, nation, force, paix, combattant, production, liberté, victoire, armée, conflit, logement, reconstruction* etc.

Comme en Angleterre, cette époque innove aussi en matière de sécurité sociale puisqu'on adopte un programme d'assurance-chômage en 1940, on étend le régime des pensions (1943), on instaure un régime d'allocations familiales (1944) et on légifère en matière de logement social, certains de ces programmes comme l'assurance-chômage, nécessitant un amendement constitutionnel. Voici l'une des phrases les plus caractéristiques de cette période :

> "Au Canada, on a déjà considérablement pourvu à la sécurité sociale par des mesures fédérales établissant des rentes viagères, l'assurance-chômage, les pensions aux vieillards, aux aveugles et aux anciens soldats invalides, puis par des mesures provinciales relatives aux accidents, à la maladie, à l'hospitalisation, aux pensions des mères et des veuves et aux secours de maternité." (1943).

La construction de l'État providence canadien avive alors les conflits avec les provinces qui contestent les ingérences du gouvernement fédéral dans leurs champs de compétence. Ces innovations en matière sociale se poursuivront jusqu'à la fin des années cinquante avec l'adoption d'une loi sur la sécurité de la vieillesse (1951), une loi sur les aveugles (1951), une autre sur les invalides en 1954 et une autre sur l'assurance-hospitalisation en 1957.

1968-1983

Même si ses adversaires reprochaient à P. Trudeau son manque d'intérêt pour l'économie, les discours sous sa gouverne sont dominés par la première crise économique et l'interventionnisme accru de l'Etat central dans les problèmes économiques et sociaux comme en témoignent les substantifs les plus sur-employés durant cette période tels : *emploi, prix, droit, ressource, revenu, inflation, croissance, personne, pétrole…* ainsi que les adjectifs *économique, fédéral, social, industriel…* Voici la phrase la plus caractéristique qui marque bien "l'idéologie Trudeau".

> "Prévoyance, cran et esprit d'entreprise à l'appui, nous pouvons, si nous le voulons, faire de notre société, une société au sein de laquelle la diversité humaine devient un atout et non une entrave ; une société au sein de laquelle la liberté individuelle et l'égalité des chances demeurent les plus précieux de tous les biens ; une société au sein de laquelle le bonheur de vivre se mesure en termes qualitatifs et non quantitatifs ; une société qui encourage la créativité, l'audace, l'ingéniosité et l'initiative, non pas froidement et de façon impersonnelle à des fins de stricte utilité, mais avec chaleur et cordialité, comme entre amis. " (1970).

Dès le début, ces discours sont marqués par l'insistance sur "fédéral" et "fédération" mais ce n'est qu'après 1977 qu'apparaît *ouvertement* la volonté de modifier les institutions canadiennes comme le révèlent deux des phrases caractéristiques de cette ambition :

> "Ces propositions porteront notamment sur l'essence de la fédération canadienne et sur ses objectifs, sur certaines libertés et certains droits fondamentaux dont tous les Canadiens devraient jouir, puisqu'ils sont indispensables à la survie du Canada comme société libre et démocratique, et sur certains éléments de la formule fédérative canadienne qui importent à son bon fonctionnement" (1977). "Nous nous engageons par ailleurs dans une nouvelle ère de planification et de consultation fédérales provinciales" (1983).

Notons enfin l'allongement des discours et leur caractère "idéologique" (peu de verbes, beaucoup de substantifs) et surtout la volonté de fonder une nouvelle "société canadienne".

"Un renouveau de la fédération canadienne est tout aussi indispensable" (1978). Même s'il a largement contribué au "nation building" canadien par ses politiques publiques, P. Trudeau n'a pas misé sur une rhétorique nationaliste puisque les vocables identitaires *Canada* et *Canadien* sont en spécificités négatives.

1984-1991

Les principaux leitmotiv de B. Mulroney sont la prospérité économique, le libre-échange, l'environnement et la constitution canadienne. « Réconciliation et unité nationale : la législature que nous inaugurons aujourd'hui doit marquer le début d'une ère de réconciliation nationale, de renouveau économique et de justice sociale. » (1984) Les conservateurs introduisirent une nouvelle politique constitutionnelle plus favorable aux provinces et tentèrent de solutionner la question québécoise en proposant l'accord du lac Meech.

> « Il importe donc au plus haut point de rendre la Constitution plus conforme à ce que nous sommes, à ce que nous représentons en tant que pays et à ce que nous aspirons à devenir en tant que peuple ». (1991)

Ils se démarquèrent aussi en faisant adopter une politique de libre-échange avec les Etats-Unis et par la suite avec le Mexique. Ces nouvelles orientations de la politique canadienne sont attestées par le suremploi des substantifs et adjectifs suivants : *prospérité, initiative, environnement, concurrence, déficit, libre-échange, constitution, marché, identité ; national, économique, mondial, compétitif, écologique, multilatéral, constitutionnel*.

Conclusions

La méthode de segmentation présentée ci-dessus est aisée à mettre en œuvre et elle se déroule de manière automatique, ce qui donne une grande solidité aux découpages obtenus. Combiné avec la recherche des ruptures stylistiques dans le même corpus (Hubert & Labbé 2002), cet outil est bien adapté aux corpus de grandes dimensions. Une fois chaque période délimitée, l'étude de son vocabulaire caractéristique permet de mettre en valeur les principaux thèmes et les singularités. Ces méthodes permettent de supprimer l'un des points "aveugles" de la statistique textuelle : le découpage des corpus en autant de sous-parties qu'ils en contiennent effectivement et non plus selon les *a priori* de l'observateur ou la doxa des historiens.

Naturellement, ces techniques sont "exploratoires". Elles assistent le lecteur sans se substituer à lui ; elles n'épargnent pas le "retour au texte" mais le facilitent. L'observateur saura où il doit chercher.

En ce qui concerne la politique canadienne, notre analyse confirme que le renouvellement du vocabulaire correspond soit à des événements majeurs comme le déclenchement d'une guerre (1939), l'arrivée au pouvoir d'un nouveau chef (1968) ou encore d'un nouveau parti qui a fait un long séjour dans l'opposition comme ce fut le cas pour le parti conservateur en 1984 et en 2011 où – après 5 ans de gouvernement minoritaire – les conservateurs ont eu les coudées franches pour imposer leur ordre du jour avec l'obtention d'une majorité de sièges. Cette exploration lexicométrique met aussi en relief la difficulté d'analyser les situations de gouvernement minoritaire comme ce fut le cas entre 1962 et 1968 et entre 2004 et 2010. Même si deux partis différents ont gouverné, on n'observe pas de renouvellement significatif du vocabulaire, comme si l'absence de majorité imposait la continuité au moins dans le discours. On peut aussi observer que lorsqu'un gouvernement reste au pouvoir pour plus d'un mandat, il a tendance à se répéter comme l'indique les séquences 4 et 5 pour P. Trudeau et 6 et 7 pour B. Mulroney. Pour aller plus loin, il faudrait reconstituer l'ensemble des discours

des Premiers ministres canadiens comme nous l'avons fait pour leurs homologues québécois (Labbé et Monière, 2008).

Remerciements

Pierre Hubert a mis au point le modèle de partition du vocabulaire avec D. Labbé. Cyril Labbé a écrit les programmes informatiques avec D. Labbé et il a contribué à l'élaboration de la grande bibliothèque (avec D. Monière pour les corpus politiques). E. Arnold, C. Labbé et J. Savoy ont relu une première version de ce texte et ils ont proposé plusieurs corrections et améliorations importantes. Enfin, Les auteurs remercient les deux relecteurs anonymes pour leurs remarques qui ont permis d'améliorer sensiblement ce texte.

References

Les discours des Premiers ministres québécois et canadiens sont en ligne sur le site de l'Assemblée nationale du Québec : http://archivespolitiquesduquebec.com/

Arnold Edward (2005). Le discours de Tony Blair (1997-2004). *Corpus*, 4, p. 55-77.

Bowker L. & Pearson J. (2002). *Working With Specialized Language: A practical guide to using corpora*. London: Routledge.

Delanoë A. (2010). Statistique textuelle et séries chronologiques sur un corpus de presse écrite. Le cas de la mise en application du principe de précaution. In Bolasco S., Chiari I., Giuliano L. (Eds). *Proceedings of 10th International Conference Statistical Analysis of Textual Data*. Rome: Edizioni Universitarie di Lettere Economia Diritto, 2010, Vol 1, p. 561-572.

Hubert P. et Labbé D. (1988b). Un modèle de partition du vocabulaire. In Labbé D., Serant D. et Thoiron P. *Etudes sur la richesse et la structure lexicales*. Paris-Genève : Slatkine-Champion, 1988, p 93-114.

Hubert P., Labbé C. & Labbé D. (2002). Segmentation automatique des corpus. *Voyages de l'autre côté* de J.-M. Le Clezio. In Morin A. et Sébillot P. (Eds). *VIe Journées Internationales d'Analyse des Données Textuelles (Saint-Malo 13-15 mars 2002)*. Rennes : IRISA-INRIA, 2002, Vol. 2, p. 359-369.

Labbé C., Labbé D. & Hubert P. (2004). Automatic Segmentation of Texts and Corpora. *Journal of Quantitative Linguistics*, december 2004, 11-3, p. 193-213.

Labbé D. (1990). *Normes de saisie et de dépouillement des textes politiques*. Cahier du CERAT n° 7. Grenoble : CERAT-IEP.

Labbé & Monière (2008).*Les mots qui nous gouvernent*, Montréal, Éditions Monière-Wollank.

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.

Labbé D. & Monière D. (2003). *Le discours gouvernemental*. Paris-Genève : Slatkine-Champion.

Labbé D. & Monière D. (2010a). Quelle est la spécificité des discours électoraux? Le cas de Stephen Harper. *Canadian Journal of Political Science / Revue canadienne de science politique*, 43:1, (March/ mars 2010), p. 69–86.

Lafon P. (1984). *Dépouillements et statistiques en lexicométrie.* Genève-Paris : Slatkine:Champion.

Monière D. & Labbé D. (2010b). Segmentation des corpus chronologiques : 143 ans de discours gouvernemental au Québec. In Bolasco S., Chiari I., Giuliano L. (Eds). *Proceedings of 10th International Conference Statistical Analysis of Textual Data*. Rome: Edizioni Universitarie di Lettere Economia Diritto, 2010, Vol 2, p. 805-816.

Monière D. & Labbé D. (2012). Le vocabulaire caractéristique du Premier ministre du Québec J. Charest comparé à ses prédécesseurs. In Dister Anne, Longrée Dominique, Purnelle Gérald (éds). *Proceedings of the 11th International Conference on Textual Data Statistical Analysis*. Liège : LASLA - SESLA, 2012, p.737-751.

Annexe 1.
La bibliothèque des discours politiques

Corpus	Dates	Nombre de textes	Longueur (mots)
Premiers ministres canadiens	1867-2012	701	1 098 172
Premiers ministres québécois	1867-2012	1 051	2 994 863
Présidents français	1958-2012	1 247	3 807 945
Premiers ministres français	1945-2013	53	299 425
Campagnes présidentielles	2007 & 2012	2 373	2 583 192
Divers et autres pays	1958-2012	108	740 058
Total		5 533	11 523 655

Annexe 2.
Le corpus des discours du trône canadien

Premier ministre	Parti	Dates	Nombre de discours	Longueur (mots)
John Alexander Macdonald	conservateur	1867-1873	7	5497
Alexander Mackenzie	libéral	1874-1878	5	5000
John Alexander Macdonald	conservateur	1879-1891	14	12573
John Joseph Caldwell Abbott	conservateur	1891-1892	1	807
John Sparrow David Thompson	conservateur	1892-1894	2	1 593
Mackenzie Bowell	conservateur	1894-1896	2	1 683
Wilfrid Laurier	libéral	1896-1911	14	11 370
Robert Laird Borden	conservateur	1911-1920	6	5 330
Arthur Meighen	conservateur	1920-1921	1	992
William Lyon Mackenzie King	libéral	1921-1930	8	9 669
Richard Bedford Bennett	conservateur	1930-1935	5	4 364
William Lyon Mackenzie King	libéral	1935-1948	10	18 560
Louis Stephen Saint-Laurent	libéral	1948-1957	9	13 594
John George Diefenbaker	conservateur	1957-1963	7	16 369
Lester Bowles Pearson	libéral	1963-1968	5	14 926
Pierre Elliott Trudeau	libéral	1968-1978	10	20 455
Charles Joseph Clark	conservateur	1979	1	2 595
Pierre Elliott Trudeau	libéral	1980-1983	2	12 243
Martin Brian Mulroney	conservateur	1984-1993	5	20 589
Jean Chrétien	libéral	1993-2004	5	26 777
Paul Martin	libéral	2004-2006	2	14 886
Stephen Harper	conservateur	2006-	4	21 136
Total			128	260836