

# L'influence des plumes de l'ombre sur les discours des politiciens<sup>1</sup>

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

<sup>1</sup>Université de Montréal, denis.moniere@umontreal.ca,

<sup>2</sup>Institut d'études politiques de Grenoble, dominique.labbe@iep.upmf-grenoble.fr

## Abstract

By comparing the speeches of two Quebec prime ministers and by calculating the intertextual distances, the authors aim to identify the speeches written by a speech writer. The authors point out that there is a strong differentiation between the speeches in written or oral forms. They prove that more than one speech writer is the genius author of the speeches. The speeches of the prime ministers amalgamate various influences and combine oral and written styles. They control their speeches even if they do not write them personally.

## Résumé

Par la comparaison des discours de deux premiers ministres québécois et par le calcul des distances intertextuelles, les auteurs tentent d'identifier les discours qui sont attribuables à une plume de l'ombre. Les auteurs montrent qu'il y a une forte différenciation entre les discours écrits et les discours oraux et qu'il n'y a pas qu'une seule plume à l'œuvre. Les discours des chefs de gouvernement amalgament une diversité d'influences et combinent des styles oraux et écrits. Ils contrôlent leurs discours même s'ils ne les écrivent pas eux-mêmes.

**Mots-clés :** Plumes de l'ombre, discours politiques, distance intertextuelle, syntagmes verbaux.

## 1. Introduction

Les auteurs des discours politiques sont rarement les personnes qui les prononcent. La plupart des chefs politiques font appel à des plumes de l'ombre pour rédiger leurs allocutions et exigent de leurs rédacteurs qu'ils adoptent un style qui leur convient. Ils en contrôlent la forme et le contenu puisque si d'aventure les discours proposés ne correspondent pas à leur personnalité et à leur style, ils peuvent toujours les rejeter et chercher un autre rédacteur plus conforme à leurs attentes.

L'analyse empirique de la contribution des rédacteurs politiques n'a jamais pu être faite parce que l'existence des plumes de l'ombre relève du secret d'État (E. Faux et al, 1991 : 10) et que, pour mesurer statistiquement leur influence réelle, il faudrait neutraliser les variables idéologiques, temporelles et personnelles. Cela supposerait qu'on puisse tester un corpus de discours prononcés par deux chefs politiques différents ayant eu le même rédacteur, situation qui est plutôt exceptionnelle.

Les discours des premiers ministres québécois permettent de réaliser cette expérience puisque deux d'entre eux - Jacques Parizeau (entre 1994-1996) et Lucien Bouchard (de 1996 à 2001) qui ont été chefs du même parti et ont gouverné successivement - ont eu recours à la même personne pour rédiger leurs discours : le journaliste Jean-François Lisée. Si le rédacteur de

---

<sup>1</sup> Les auteurs remercient Cyril Labbé qui les a aidés dans l'élaboration des logiciels et dont l'assistance a été précieuse à toutes les étapes de cette expérience.

l'ombre imprime son style ou sa marque lexicale sur les discours, on doit s'attendre à une certaine proximité lexicale entre ces deux premiers ministres, puisque les variables temporelles et idéologiques sont relativement neutres.

## 2. La fabrique des discours

Dans une entrevue<sup>2</sup> avec J-F. Lisée, celui-ci nous a expliqué la procédure d'élaboration des discours des premiers ministres québécois. Comme le premier ministre prononce en moyenne quatre discours chaque semaine, même s'il le voulait, il n'aurait pas le temps d'écrire ses discours. Le rôle et l'influence du rédacteur varient selon les types de discours.

Pour les textes qui impliquent l'ensemble du gouvernement, comme le discours inaugural, le rédacteur dispose d'une liberté restreinte : il doit travailler de concert avec la machine gouvernementale et s'inspirer du « briefing book » préparé par les différents ministères. L'écriture se fait en collaboration avec le premier ministre, son directeur de cabinet, le secrétaire général du gouvernement et en consultation avec les sous-ministres.

Pour les discours qui définissent les positions politiques et stratégiques du premier ministre, le rédacteur est le seul responsable du discours qu'il soumet directement au premier ministre :

La plupart du temps, j'écrivais le discours la veille de la prestation. Dans le cas de M. Parizeau, j'écrivais le discours, je le faisais relire par un ou deux conseillers avant de le lui présenter. Il le lisait et c'était généralement approuvé à 98%. Dans le cas de Bouchard, qui avait été lui-même un rédacteur des discours du premier ministre canadien Brian Mulroney, j'écrivais le discours la veille de la prestation et quelques heures avant qu'il le prononce, il réunissait un comité de lecture pour l'améliorer ou apporter des nuances. Les changements étaient de l'ordre de 5 à 10%.

Enfin pour les discours annonçant des subventions ou des mesures gouvernementales, on faisait appel à d'autres employés du bureau du premier ministre qui s'inspiraient des discours d'orientation rédigés préalablement par le rédacteur.

Selon J-F. Lisée, les premiers ministres font deux types de discours : ceux du midi et ceux du soir. Les premiers sont rédigés et le premier ministre les lit. Il peut occasionnellement improviser pour introduire des anecdotes mais celles-ci sont souvent prévues d'avance. Le premier ministre peut présenter le même discours à plusieurs auditoires différents surtout dans une tournée en région, mais il doit se soucier d'offrir des éléments de nouveauté aux journalistes qui l'accompagnent. Pour les discours du soir, prononcés devant des assemblées partisans ou à l'occasion de dîners, le rédacteur prépare un canevas de discours, mais il y a rarement des textes écrits, sauf pour les dîners d'État avec des visiteurs étrangers.

Le bon rédacteur est celui qui réussit à s'imprégner du style du personnage qu'il sert :

Comme on travaille avec eux, on les entend constamment. On a leur rythme de voix et leurs expressions dans l'oreille et on les met sur la page. On adopte des tics langagiers qui sont les leurs et qu'on met dans le texte. Mais eux aussi en viennent à adopter nos expressions. Ça fonctionne par osmose.

J.-F. Lisée pense qu'il a assuré une continuité de style entre J. Parizeau et L. Bouchard. « *Je pense qu'il y a une similitude entre mes discours pour Parizeau et mes discours pour Bouchard. À tous deux, j'ai voulu donner un ton et un style de chef d'État et je ne me souviens pas que Bouchard ait voulu changer beaucoup de choses au plan stylistique* ». À l'appui de cette impression, il cite le chef de l'opposition, Daniel Johnson, qui, dans sa réponse au discours inaugural de L. Bouchard a reproché à celui-ci de parler comme J. Parizeau :

<sup>2</sup> Le 4 octobre 2005.

Et j'ai eu l'impression, dans le style... Je ne sais pas si c'est parce que c'est le même auteur ou collaborateur, là, qui est toujours, évidemment, au travail, mais j'ai eu l'impression d'un certain déjà vu dans, je dirais, la façon dont l'ensemble des points du premier ministre a été amené.<sup>3</sup>

### 3. Le corpus analysé

Pour cette expérience, nous avons utilisé le corpus de tous les *discours de circonstances* prononcés par les premiers ministres du Québec devant divers auditoires comme les chambres de commerce, les syndicats, les réunions partisans, ou à des occasions particulières, comme la fête nationale ou les remises de décorations. Ce corpus couvre 45 ans de vie politique, soit la période 1960-2005. Ces années ont été particulièrement intenses sur la scène politique québécoise, le Québec ayant connu une forte polarisation politique autour de la question nationale et de nombreuses crises sociales engendrées par la modernisation accélérée de la société québécoise et par les crises économiques mondiales.

Le corpus a été constitué à partir des archives numériques disponibles sur le site Internet du premier ministre du Québec où on trouve les principaux discours des chefs du gouvernement québécois depuis 1994. Pour la période antérieure à 1994, nous avons eu recours aux archives de la bibliothèque de l'Assemblée nationale du Québec qui possède une collection des discours des premiers ministres depuis 1960.

Tous ces textes sont soumis aux mêmes traitements : correction orthographique, standardisation des graphies puis lemmatisation (Labbé, 1990). Les calculs statistiques sont effectués sur les vocables (association d'un lemme et d'une catégorie grammaticale).

Pour la présente recherche, nous avons réduit le corpus Bouchard au nombre de mois du corpus Parizeau (17 mois), afin de contrôler la variable temporelle et limiter ainsi les effets de conjoncture sur le vocabulaire utilisé. Nous avons aussi exclu les premiers ministres qui ont succédé à L. Bouchard.

Premier ministre	N discours	N mots	N vocables
Lesage (1960-66)	141	307 328	8 326
Johnson (1966-68)	37	53 252	4 183
Bertrand (1968-70)	32	34 542	3 488
Bourassa 1 (1970-76)	70	205 370	5 884
Lévesque (1976-85)	82	408 014	9 719
Bourassa 2 (1985-94)	56	167 412	4 981
Parizeau (1994-96)	42	140 447	5 991
Bouchard (1996-01)	154	435 346	10 192

Tableau 1 : Caractéristiques du corpus des premiers ministres québécois (1960-2001)

### 4. Le calcul de la distance intertextuelle

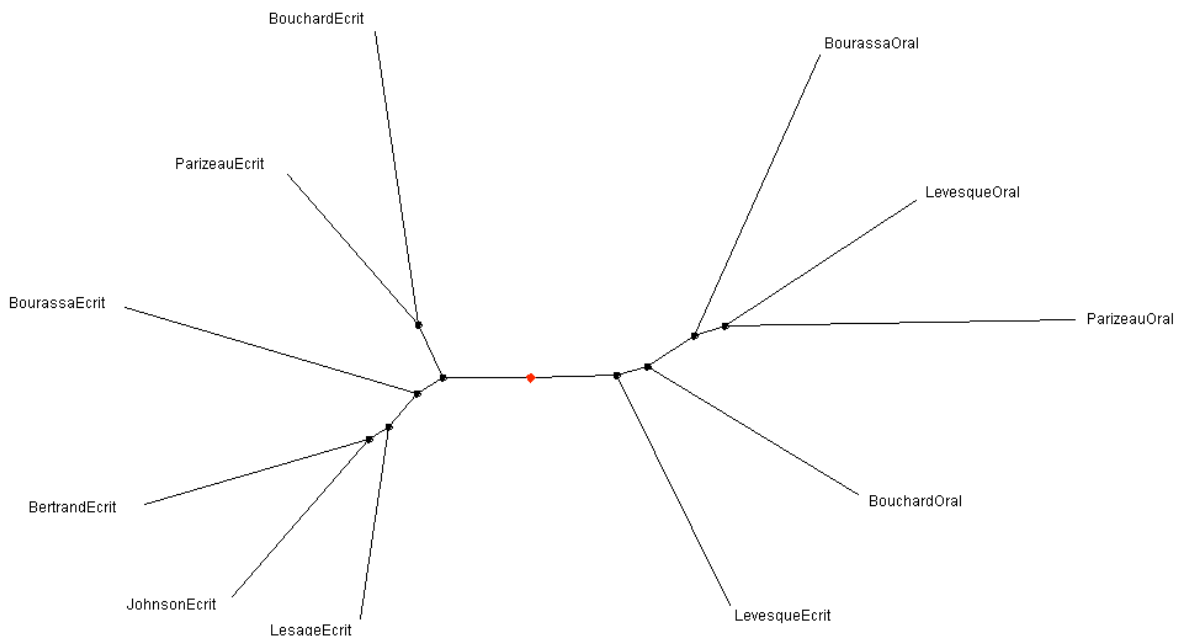
Pour savoir si l'on peut détecter une même plume de l'ombre derrière les discours des deux premiers ministres souverainistes, nous avons eu recours à la distance intertextuelle (Labbé & Monière, 2000 ; Labbé & Labbé, 2001 et 2003) et à la classification arborée (Luong, 1988). L'algorithme et les propriétés des arbres sont présentés dans (Rulman, 2003 et Labbé & Labbé, 2005)

<sup>3</sup> Journal des débats, 26 mars 1996.

Rappelons que quatre facteurs influent sur la distance entre deux textes. Dans un ordre décroissant d'importance : le genre, l'auteur, le thème et l'époque. Pour isoler la variable "auteur", il faut disposer, pour un même scripteur supposé, de textes de même genre, écrits à peu près à la même époque sur des thèmes voisins pour plusieurs "clients" différents.

Quelle est l'influence de la variable "genre" ? Dans le corpus figurent des allocutions lues par l'orateur - que l'on peut considérer comme de l'écrit - et des transcriptions de conférences de presse, émissions radio-télévisées, etc. qui sont de l'oral (même si certaines réponses sont préparées). Le graphique ci-dessous donne le résultat de la classification sur ces corpus "oraux" et "écrits" pour chacun des premiers ministres jusqu'à L. Bouchard.

Sur un arbre, la distance entre deux textes est figurée par la longueur du chemin à parcourir en suivant les « branches » et le « tronc » pour joindre les deux sommets correspondant. Par exemple, la plus longue distance sépare les écrits de J.-J. Bertrand (à l'ouest du graphe) et les interventions orales de J. Parizeau (à l'est du graphe).



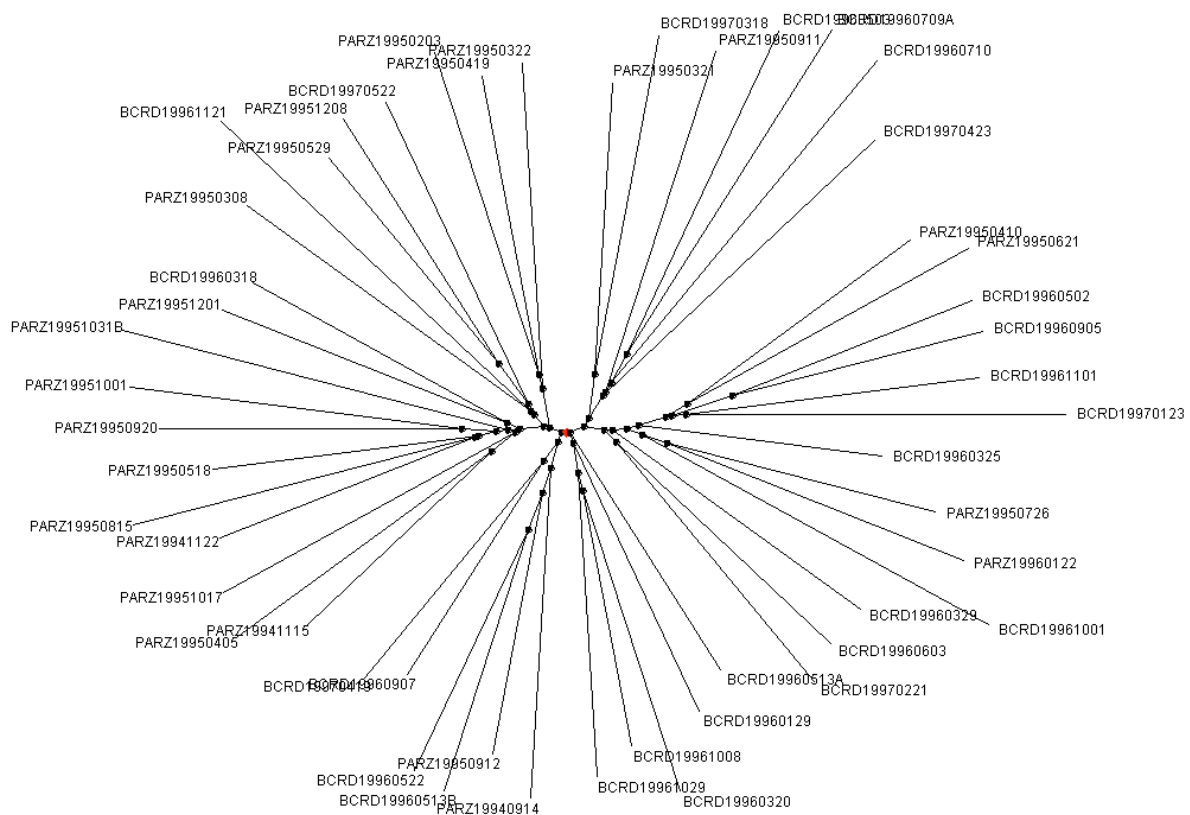
*Graphique 1. Classification arborée du corpus Premiers ministres du Québec  
(discours classés selon la nature : écrit lu ou dialogue spontané).*

Deux conclusions se dégagent de cette classification :

- la variable genre se révèle plus puissante que la variable auteur. Les corpus écrits sont classés à gauche, l'oral à droite ; les distances entre corpus oraux et écrits d'un même auteur sont systématiquement supérieures à celles séparant des auteurs différents dans un même genre. Cependant R. Lévesque semble ignorer la "barrière" du genre. Ses allocutions "écrites", sont classées avec l'oral (même si son "oral" est plus proche de celui de Bourassa ou de Parizeau que de ses propres écrits). Cette "mauvaise" classification confirme un fait de notoriété publique : R. Lévesque s'écartait souvent des notes qui lui avait été préparées et improvisait, même dans les circonstances les plus solennelles.

- Alors que "Bouchard Oral" est plus proche de "Lévesque Oral" que de "Parizeau Oral", la classification regroupe "Bouchard Écrit" avec "Parizeau Écrit" (c'est la distance la plus faible : .24). Il ne faut pas se hâter de conclure à un auteur unique. En effet, trois corpus écrits (Lesage, Johnson et Bertrand) sont pratiquement aussi proches. De plus, la partie gauche de l'arbre suggère l'existence d'une forte dimension chronologique (les premiers ministres sont regroupés selon leur succession dans le temps).

Pour aller plus loin, il est donc nécessaire de comparer non plus les corpus totaux mais chacun des textes contenus dans "Bouchard Écrit" et "Parizeau Écrit". La classification ci-dessous porte sur les 50 discours écrits de J. Parizeau et L. Bouchard dont la taille est supérieure à 1000 mots, seuil en dessous duquel la distance intertextuelle est trop instable (Labbé & Labbé 2003).



Graphique 2. Classification arborée sur les corpus Parizeau Écrit (PARZ) et Bouchard Écrit (BCRD)

La plupart des groupes isolés par l'algorithme mélangent des textes des deux premiers ministres. Il y a fort peu de clusters homogènes (L. Bouchard au sud-sud-est) ou relativement homogènes lorsque des textes "allogènes" sont rattachés tardivement à un groupe par un nœud proche de la racine (J. Parizeau à l'ouest du graphe). Cette classification amène deux conclusions : la forme étoilée du graphe suggère un corpus plutôt homogène (les textes sont relativement équidistants les uns des autres) ; la plupart des clusters mélangent des textes des deux "auteurs". Il y aurait donc bien une plume principale (sinon unique) ou un processus d'élaboration impliquant plus ou moins les mêmes personnes selon le même mode opératoire. L'examen de la matrice des distances (Labbé, 2006) permet d'isoler les couples de textes dont les distances sont inférieures au seuil en dessous duquel un même auteur est très probable

(groupe 1). À l'inverse, les distances plus grandes que le seuil supérieur isolent quelques textes dont les auteurs sont probablement différents de celui du groupe 1 (ou dont la confection n'a pas suivi le même mode opératoire) : groupe 2.

---

**Groupe 1** Textes ayant probablement un auteur unique dans les corpus Parizeau - Bouchard (classement par ordre chronologique)

**Parizeau (11 textes) :**

- Allocution devant la chambre de commerce de Montréal métropolitain (15 novembre 1994) ;
- Discours à Toronto (22 novembre 1994) ;
- Allocution devant la Chambre de commerce de Chaudière-Appalaches (5 avril 1995) ;
- Allocution devant la Chambre de commerce de Laval (10 avril 1995) ;
- Allocution devant le syndicat des communications, de l'énergie et du papier (18 mai 1995) ;
- Allocution devant la municipalité de Montréal (29 mai 1995) ;
- Allocution à l'occasion du dépôt du rapport du Comité consultatif sur l'autoroute de l'information (26 juillet 1995) ;
- Allocution à l'occasion du débat sur la question référendaire (12 septembre 1995) ;
- Allocution à la clôture du débat sur la question référendaire (20 septembre 1995) ;
- Allocution devant l'association des Partenaires pour la souveraineté (1 décembre 1995) ;
- Allocution devant la Chambre de commerce du grand Montréal (8 décembre 1995) ;

**L. Bouchard (14 textes) :**

- Discours d'assermentation (29 janvier 1996) ;
- Ouverture de la conférence sur le devenir économique et social du Québec (18 mars 1996) ;
- Discours de clôture de la conférence sur le devenir économique et social du Québec (20 mars 1996) ;
- Discours inaugural (25 mars 1996) ;
- Discours devant la chambre de commerce du Montréal métropolitain (29 mars 1996) ;
- Allocution devant les membres de la communauté italienne de Montréal (13 mai 1996) ;
- Point de presse à la suite du conseil des ministres (13 mai 1996) ;
- Dépôt d'une motion réaffirmant le droit du peuple québécois d'assumer son propre destin (22 mai 1996) ;
- Discours au Conseil national du parti québécois (7 septembre 1996) ;
- Allocution devant la Chambre de commerce de Laval (8 octobre 1996) ;
- Ouverture du sommet sur l'économie et l'emploi (29 octobre 1996) ;
- Clôture du sommet sur l'économie et l'emploi (1 novembre 1996) ;
- Allocution devant l'Alliance des manufacturiers et exportateurs du Québec (21 février 1997) ;
- Discours d'ouverture au Conseil national du Parti québécois (19 avril 1997).

**Groupe 2.** Textes ayant probablement un auteur différent de celui du groupe 1

**- J. Parizeau :**

- meeting commun des trois partis soutenant la souveraineté du Québec (sous la présidence conjointe de MM; L. Bouchard, M. Dumont et J. Parizeau, 15 août 1995) ;
- allocution radio-télévisée au lendemain de l'échec au référendum sur la souveraineté-association (31 octobre 1995) ;

**L. Bouchard :**

- Dîner offert en l'honneur du président de la Catalogne Jordi Pujol (9 juillet 1996) ;
  - Allocution à l'occasion de la rencontre Québec-Catalogne (10 juillet 1996) ;
  - Inauguration du nouvel édifice des HEC à Montréal (1 octobre 1996) ;
  - Allocution de présentation du rapport " Les enfants au coeur de nos choix " (23 janvier 1997).
- 

Au total, sur la cinquantaine de textes soumis à l'analyse (dimension supérieure à 1.000 mots), 25 peuvent être attribués à un même auteur principal. Ces deux listes appelleraient beaucoup de commentaires qui dépasseraient le cadre de cette note. Signalons en deux intéressants :

- les services du premier ministre ont classé l'un des textes de L. Bouchard comme étant un "point de presse", c'est-à-dire une discussion avec les journalistes. Or ce texte est rattaché au corpus écrit. À la lecture, il apparaît que c'est une déclaration solennelle lue devant les caméras de télévision et non suivie de questions. Si nous avons suivi la classification a priori des services officiels, ce texte aurait été rangé avec les oraux et n'aurait pas été soumis à l'expérience ;

- deux textes cruciaux de la période n'auraient pas été rédigés par la plume de l'ombre habituelle de J. Parizeau : le discours prononcé lors du grand meeting du 15 août 1995 qui lance la campagne pour le référendum du 30 octobre - mais il s'agit d'un "discours du soir" donc largement improvisé - et la déclaration faite le lendemain de ce référendum, par laquelle le premier ministre reconnaît la défaite, annonce son départ prochain et tente de justifier sa petite phrase, prononcée la veille au soir, sur "les votes ethniques".

Au total, 32 textes sur 50 sont donc "attribués" soit à une même "source" principale soit à une ou plusieurs autres différentes de cette source principale.

Que faire avec les 18 restants ? La classification automatique (non reproduite ici) et la classification arborée ci-dessus permettent d'en "récupérer" quelques-uns. Par exemple, l'arbre comporte au nord, un couple de textes de L. Bouchard (18 mars 1997) et J. Parizeau (21 mars 1995). De même, au sud, un cluster regroupe 2 couples : Parizeau (14 septembre 1994 et 12 septembre 1995), Bouchard (22 mai 1996 et 13 mai 1996). Comme ce dernier texte appartient au groupe 1, on peut supposer que les trois autres ont été confectionnés selon le même mode opératoire. De même, à l'ouest-sud-ouest, une feuille de l'arbre regroupe un ensemble de textes de J. Parizeau comprenant notamment un des textes figurant dans le groupe 1 (15 novembre 1995). Les autres peuvent donc être supposés avoir été composés de la même manière (au moins pour l'essentiel).

Toutefois, il faut se garder d'aller trop loin dans ces déductions car les algorithmes de classification - toujours imparfaits - ne donnent que le "meilleur" classement possible au sein d'une grande collection d'objets. Pour valider ces regroupements, il est donc nécessaire de se reporter à la matrice originale et, lorsque la distance séparant les textes est élevée (ce qui est le cas pour le dernier ensemble de textes de J. Parizeau), il est plus prudent de ne pas conclure...

Au passage, ce "déchet" ne doit pas étonner puisque la méthode ne vise pas à attribuer tous les textes en acceptant un taux d'erreur inévitable mais, au contraire, à éviter les erreurs, quitte à laisser une proportion notable de cas irrésolus.

Est-il possible de compléter cette attribution par d'autres indices ? L'essai sur Corneille et Molière (Labbé, 2003) signale que les combinaisons "verbe+verbe" du type "vouloir faire" sont un moyen possible de repérer un auteur unique. Le tableau ci-dessous récapitule les résultats de l'expérience pour les deux corpus.

<b>Bouchard</b>		<b>F absolue</b>	<b>%o</b>	<b>Parizeau</b>		<b>F absolue</b>	<b>%o</b>
Devoir	être	24	4,76	pouvoir	Faire	17	3,99
Devoir	faire	22	4,36	aller	Faire	14	3,29
Pouvoir	faire	20	3,96	devoir	Être	13	3,05
Aller	faire	12	2,38	pouvoir	Compter	12	2,82
Devoir	prendre	12	2,38	vouloir	Dire	12	2,82
Faire	avancer	11	2,18	pouvoir	Dire	10	2,35
Faire	connaître	10	1,98	falloir	Faire	9	2,11
Pouvoir	être	10	1,98	vouloir	Être	8	1,88
Devoir	dire	8	1,59	faire	Avancer	7	1,64
Faire	reculer	8	1,59	pouvoir	Parler	7	1,64
Falloir	faire	8	1,59	aller	Dire	6	1,41
Faire	progresser	7	1,39	cesser	Emprunter	6	1,41
Vouloir	dire	7	1,39	devoir	Pouvoir	6	1,41
Vouloir	faire	7	1,39	aller	Cesser	5	1,17
Aller	prendre	6	1,19	devoir	Faire	5	1,17
Avoir	faire	6	1,19	emprunter	Payer	5	1,17
Faire	valoir	6	1,19	faire	Baisser	5	1,17
Aller	demander	5	0,99	faire	Entendre	5	1,17
Aller	réduire	5	0,99	falloir	Dire	5	1,17
Aller	tenter	5	0,99	pouvoir	Constater	5	1,17
Falloir	agir	5	0,99	pouvoir	Mettre	5	1,17
Falloir	savoir	5	0,99	réussir	Faire	5	1,17
Pouvoir	compter	5	0,99	vouloir	Faire	5	1,17
Savoir	faire	5	0,99	aller	Prendre	4	0,94
Vouloir	saluer	5	0,99	avoir	Faire	4	0,94
Être	dire	5	0,99	devoir	Avoir	4	0,94
Être	faire	5	0,99	devoir	Dire	4	0,94
Aller	savoir	4	0,79	devoir	Prendre	4	0,94
Devoir	avoir	4	0,79	décider	Faire	4	0,94
Devoir	payer	4	0,79	faire	Bouger	4	0,94
Devoir	savoir	4	0,79	faire	Reculer	4	0,94
Entendre	parler	4	0,79	falloir	Donner	4	0,94
Faire	profiter	4	0,79	falloir	Prendre	4	0,94
Falloir	donner	4	0,79	pouvoir	Trouver	4	0,94
Falloir	mettre	4	0,79	pouvoir	Être	4	0,94
Permettre	créer	4	0,79	venir	Discuter	4	0,94
Pouvoir	constater	4	0,79	vouloir	Parler	4	0,94
Vouloir	être	4	0,79	être	Dire	4	0,94

Tableau 2 : Comparaison des syntagmes verbes/verbes chez Parizeau écrit et Bouchard Écrit (classement par ordre décroissant de fréquence)

La moitié des syntagmes verbaux les plus fréquents sont communs (21 sur 40). Ces syntagmes communs occupent 62% de la surface couverte par les combinaisons "verbe+verbe" dans les deux textes<sup>4</sup>. Quelques-unes de ces combinaisons sont "banales" (on les trouve aussi dans les autres corpus), notamment : devoir faire, falloir faire, vouloir faire... Les autres sont caractéristiques d'un style personnel : aller prendre, devoir prendre ; c'est dire, c'est faire ; faire avancer, reculer ; pouvoir compter, constater... Ces combinaisons n'ont rien d'extraordinaire mais elles sont propres à ces deux corpus et à eux seuls. Comme le résultat

<sup>4</sup> Aller (faire, prendre) ; avoir faire; devoir (avoir, dire, être, faire, prendre) ; être (dire, faire) ; faire (avancer, reculer) ; falloir (donner, faire) ; pouvoir (compter, constater, être, faire) ; vouloir (dire, être, faire).



est moins homogène que ceux obtenus par D. Labbé sur des corpus à auteur unique, cela permet d'écarter l'hypothèse d'un seul auteur pour la totalité des deux corpus, ce que suggérait déjà la classification.

Trois explications complémentaires peuvent être fournies :

- il y aurait une proportion faible mais non négligeable d'oral dans certains textes (les "discours du soir" dont parle J.-F. Lizée), donc J. Parizeau et L. Bouchard "polluent" l'expérience ;
- les Premiers ministres utilisaient certainement plusieurs scripteurs et ont parfois "collé" ensemble des textes d'origine diverses (par exemple, il serait logique que L. Bouchard ait demandé aux auteurs du rapport sur l'enfance - dernier texte du groupe 2 - de lui soumettre un projet d'allocution) ;
- enfin, il est probable que la relation d'influence entre le Premier ministre et la plume de l'ombre est mutuelle...

## Conclusions

Il n'est pas possible de conclure que J.-F. Lisée est bien l'auteur des 25 textes attribués à un seul auteur. Pour cela il faudrait disposer d'un corpus "étalon", comportant des textes attribuables en toute certitude à ce seul écrivain, ayant été écrits à peu près à la même époque et portant sur des thèmes proches.

Enfin, quelle confiance attribuer à ces classifications ? Selon le procédé présenté par Labbé (2006), les groupes 1 et 2 sont constitués avec moins de 5 chances sur 100 de se tromper. En ce qui concerne les arbres, Labbé et Labbé (2005) proposent de calculer la proportion de l'information initiale contenue dans la matrice des distances qui se trouve conservée dans l'arbre (ou ce qui revient au même : la proportion de l'information perdue lors du tracé de cet arbre). L'indice pouvant être calculé pour la totalité de l'arbre, pour un nœud, un chemin ou un cluster. Ainsi, le premier arbre conserve 98,2% de l'information totale contenue dans la matrice initiale et le second : 96,8%. Pour les deux arbres, tous les nœuds ont une qualité supérieure à 95%. En revanche, certains chemins ont une qualité plus médiocre. Par exemple, le second graphe comporte 1225 chemins différents pour relier chacun des sommets à tous les autres. 42 de ces chemins ont une qualité comprise entre 95 et 90 % et 12 une qualité comprise entre 90 et 85. Aucun de ces tracés "douteux" ne relie les textes attribués au même auteur.

Au total, nous espérons avoir montré que le calcul de la distance intertextuelle combiné avec la classification arborée offre un outil efficace.

## Références

- Faux E. et al (1991). *Les plumes de l'ombre*. Paris, Ramsey.
- Labbé C., Labbé D. (2001). Intertextual Distance and Authorship Attribution : Corneille et Molière, *Journal of Quantitative Linguistics*, 8-3 : 213-231.
- Labbé C. et Labbé D. (2003). La distance intertextuelle. *Corpus*. 2-2003 : 95-118.
- Labbé C. et Labbé D. (2005). A Tool for Literary Studies : Intertextual Distance and Tree Classification. *Literary and Linguistic Computing*. 20-4.
- Labbé D. (1990). *Normes de saisie et de dépouillement des textes politiques*. Grenoble, Cahier du CERAT.

- Labbé D. (2003). *Corneille dans l'ombre de Molière*. Bruxelles, Impressions nouvelles.
- Labbé D. (2006, à paraître). Experiments on Authorship Attribution by Intertextual Distance in English. *Journal of Quantitative Linguistics*. 13/1.
- Labbé D. et Monière D. (2000). "La connexion intertextuelle : application au discours gouvernemental québécois", Martin Rajman (eds), *Actes des 5<sup>e</sup> journées d'analyse des données textuelles*, Lausanne, École polytechnique fédérale : 85-94.
- Labbé D., Monière D. (2003). *Le vocabulaire gouvernemental. Canada, Québec, France (1945-2000)*. Paris, Champion.
- Luong X. (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Université de Paris V.
- Ruhlman M. (2003). *Analyse arborée. Représentation par la méthode des groupements*. Grenoble, Polytech<sup>1</sup>-CERAT, <http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/Labbe>.