

## REFERENCE

### Citation

Duchastel, Jules, François Daoust et Dimitri della Faille (2008) "Le problème de l'interprétation des données à partir d'un corpus bilingue. L'exemple du discours des trois chefs de parti sur la motion de reconnaissance du « Québec comme société distincte au sein du Canada »", dans Heiden, Serge et Bénédicte Pincemin (sous la dir.), Actes des 9es Journées d'analyse de données textuelles, Lyon, Presses Universitaires de Lyon, pp. 421-431.

### RIS (EndNote)

TY - CONF

AU - Duchastel, Jules

AU - Daoust, François

AU - Della Faille, Dimitri

PY - 2008

TI - Le problème de l'interprétation des données à partir d'un corpus bilingue. L'exemple du discours des trois chefs de parti sur la motion de reconnaissance du « Québec comme société distincte au sein du Canada »

T2 - Actes des 9es Journées d'analyse de données textuelles

SP - 421

EP - 431

CY - Lyon

PB - Presses Universitaires de Lyon

A2 - Heiden, Serge

A2 - Pincemin, Bénédicte

ER -

# Le problème de l'interprétation des données à partir d'un corpus bilingue. L'exemple du discours des trois chefs de parti sur la motion de reconnaissance du « Québec comme société distincte au sein du Canada »

Jules Duchastel<sup>1</sup>, François Daoust<sup>2</sup>, Dimitri della Faille<sup>3</sup>

<sup>1 et 3</sup>UQAM – Chaire MCD – Québec – Canada

<sup>2</sup>UQAM – Centre ATO – Québec – Canada

## Abstract

We are examining the problem of interpreting data with different statistical techniques (SATO, Lexico and Alceste) applied to bilingual corpora. We are interested in the variation of interpretations based on the same discursive data along two main lines: language and software. We are comparing the results obtained on both linguistic versions of the speeches pronounced by the three party leaders on the motion recognizing “Quebec as a distinct society within Canada”. Based on the observation that results generated by similar statistical techniques vary from one linguistic version to the other, we are questioning the validity of some interpretations induced by different sets of results. Our conclusion is that particular results may only be considered as indications leading to local interpretations that must be reformulated in a more global context. Our findings also show that taking into account the other linguistic version can produce new hermeneutical cues.

## Résumé

Nous proposons d'examiner le problème de l'interprétation des données produites par des analyseurs statistiques (SATO, Lexico et Alceste) à partir d'un corpus bilingue. Nous nous intéressons à la variation des interprétations des mêmes données discursives selon deux axes : linguistique et logiciel. Nous comparons les résultats d'analyse, dans les deux langues officielles du Canada, du discours des trois chefs de parti, prononcés à la chambre des communes d'Ottawa à l'occasion de la présentation par le gouvernement de la « motion visant à reconnaître le Québec comme société distincte au sein du Canada ». Sur la base du constat de la variation des résultats générés par les mêmes techniques statistiques à partir des deux versions linguistiques, nous nous interrogeons sur la validité de l'interprétation des données ainsi produites. Nous formulons la conclusion que les résultats obtenus à partir de ces analyses ne représentent que des indices partiels menant à des interprétations locales devant nécessairement être rapportées à un ensemble englobant d'interprétation globale. Nous établissons également que le recours aux analyses dans l'autre langue est susceptible de faire surgir de nouveaux indices herméneutiques.

**Mots-clés :** interprétation, lexicométrie, corpus bilingue et multilingue, sociologie politique, Québec, Canada.

## 1. Introduction

L'importance de corpus bilingues ou multilingues est de plus en plus actuelle dans le contexte des systèmes fédéraux et des organisations supra ou trans-nationales (Deroubaix, 2004 ; Jalam et al., 2004). Nous fûmes confrontés à ce problème au moment de la publication de *L'identité fragmentée* (Bourque et Duchastel, 1996). Nous avons dû à cette époque décliner l'invitation d'un éditeur de publier une édition anglaise de notre livre sur l'analyse du discours constitutionnel canadien effectuée à partir de la version française des comptes-rendus des conférences constitutionnelles au Canada. Des pré-tests indiquaient que l'intervention des

traducteurs officiels pouvait changer la nature des conclusions auxquelles nous étions arrivés. Par exemple, la notion de *peuple*, qui revêt une importance particulière dans le discours des différents Premiers ministres du Québec, est traduite en anglais de trois manières différentes : *people*, *population*, *nation*. Il n'aurait donc pas suffi de traduire notre livre mais il aurait fallu reprendre entièrement le traitement des données dans leur version officielle anglaise. De la même manière, au moment où nous avons voulu répliquer, sur la version anglaise, l'exemple d'une analyse de discours politique en version française sur le site de SATO (<http://www.chaire-mcd.ca/sato/>), nous avons constaté que certains résultats donnant lieu à des interprétations fortes à partir du corpus en français ne se retrouvaient pas à l'identique à l'examen de la version anglaise. Fallait-il conclure que l'interprétation était prématurée ou incorrecte ou ne fallait-il pas y voir plutôt une variation du sens des mots, induite par l'usage de l'autre langue ? Cette expérience nous a conduit à poursuivre la réflexion sur le traitement de corpus bilingues et sur la variation des résultats pouvant affecter l'interprétation des données.

Notre réflexion porte avant tout sur le problème de l'interprétation<sup>1</sup>. Dans un chapitre récent (Daoust et Duchastel, 2007), nous distinguons pluralisme des interprétations comme « multiplicité des points de vue possibles dans l'observation et l'analyse d'un même objet » et pluralité des interprétations en tant que « pluralité des choix herméneutiques qui s'imposent à toutes les étapes de la recherche ». C'est dans ce second sens que nous allons nous intéresser à la variation possible des interprétations des mêmes données discursives selon deux axes : linguistique et logiciel. Après avoir présenté le contexte historique et le contenu des trois discours, nous examinerons l'effet de l'application de diverses techniques lexicométriques proposées par les logiciels SATO, Lexico et Alceste dans les deux langues. Notre objectif est de montrer non pas la fragilité des interprétations, mais leur nécessaire complémentarité dans un processus d'analyse incrémentiel.

## 2. Contexte et corpus

Nous nous sommes intéressés aux discours de trois chefs de Parti prononcés à la chambre des communes d'Ottawa à l'occasion de la présentation par le gouvernement de la « motion visant à reconnaître le Québec comme société distincte au sein du Canada ». Cette motion fut présentée en décembre 1995, suite aux engagements pris par le gouvernement fédéral à la toute fin de la campagne référendaire d'octobre 1995. Celle-ci fut couronnée par une très courte victoire du Non (50,6%) à la question : « Acceptez-vous que le Québec devienne souverain, après avoir offert formellement au Canada un nouveau partenariat économique et politique, dans le cadre du projet de loi sur l'avenir du Québec et de l'entente signée le 12 juin 1995 ? ». La notion de « société distincte » constitue le point d'achoppement de l'Accord du Lac Meech (1987-1990) qui visait à réintégrer le Québec dans la Constitution canadienne rapatriée en 1982 sans son accord. Le premier des trois locuteurs qui introduit la motion en Chambre est le Premier ministre du Canada, Jean Chrétien dont l'action avait contribué à l'échec de l'accord du Lac Meech, au terme du processus de ratification par les gouvernements provinciaux. Lors de la présentation de la motion, le second locuteur est le chef de l'opposition aux communes, Lucien Bouchard qui a conduit les forces souverainistes à une quasi-victoire du Oui au référendum de 1995. Enfin, le dernier locuteur est le chef du

---

<sup>1</sup> La question de l'interprétation, dans son acception large, se comprend comme toute action d'accorder une signification à une unité de discours dans le cadre d'une discipline et d'un système de référence théorique. Dans l'exemple abordé, nous nous situons dans une perspective d'interprétation sociologique.

Parti réformiste, Preston Manning, qui représente une tendance nettement conservatrice et régionaliste, opposée a priori à toute concession en faveur du Québec.

Dans ce contexte, il n'est pas surprenant de voir les trois locuteurs utiliser le vocabulaire entourant le thème de la société distincte de manière fort différente. Notre questionnement méthodologique consiste à voir si des outils lexicométriques nous permettent d'effectuer une lecture nuancée de ces discours sur les identités canadienne et québécoise et si la traduction du discours dans l'autre langue officielle introduit des nuances dans cette compréhension. Notons que seul le discours de Manning est prononcé en langue anglaise. Les deux autres le sont en français. Dans l'ordre, le discours de Bouchard représente plus de 60% du total des lexèmes du corpus (63,35% en français et 61,74% en anglais). Le discours de Chrétien compte pour un peu plus de 20% (respectivement 21,20% et 21,80%) et celui de Manning pour un peu plus de 15% (15,46%, 16,46%). Avant de procéder à l'application des diverses techniques statistiques aux deux versions du corpus, il est utile pour les fins de l'expérience de procéder à une première lecture flottante du matériel. Ainsi, nous acquérons une connaissance élémentaire des points de vue des locuteurs. L'application des outils servira ultimement à valider notre capacité d'interprétation des résultats statistiques.

Jean Chrétien explique que la motion sur la reconnaissance du Québec comme « société distincte au sein du Canada » constitue une réponse aux engagements qu'il a personnellement pris au cours de la dernière campagne référendaire. Rappelant l'échec du projet de « séparation mis de l'avant par le parti québécois », il souhaite favoriser « l'évolution de la fédération canadienne ». La motion vise la reconnaissance de la différence du Québec et appelle à la solidarité des concitoyens des autres provinces dans l'esprit canadien de promotion de la diversité. Le Premier ministre refuse de revenir à des pourparlers constitutionnels, mais affirme que la motion aura des effets réels sur les pouvoirs législatif et exécutif. Il prend la peine cependant de rappeler que la reconnaissance d'un statut distinct pour le Québec ne saurait être interprété comme une négation des droits des peuples autochtones à l'endroit desquels il réaffirme la protection « de leurs droits, ancestraux ou issus de traités, [...] y compris le droit inhérent à l'autodétermination ». Jean Chrétien assure également que la constitution ne pourra pas être modifiée sans le consentement du Québec, mais étend ce quasi-droit de veto à chacune des cinq grandes régions du Canada. Il oppose, en terminant, l'attitude de compromis, de respect mutuel, de générosité et de respect qui vise à assurer le « progrès sans rupture » et à « préserver un Canada uni », au projet des séparatistes qui a pour effet de monter « les canadiens les uns contre les autres » et de « détruire le Canada ».

Preston Manning profite du dépôt de la motion pour s'en prendre à l'inaction du Gouvernement canadien qui n'est intervenu qu'à la dernière heure pour sauver le Canada de la « menace séparatiste » lors du dernier référendum. Pour contrer le « rêve séparatiste », le locuteur propose de développer « une vision fédéraliste d'un Canada nouveau et meilleur ». Il est donc en faveur de modifications constitutionnelles incluant l'élaboration de « conditions de la séparation [...] pour toute tentative de sécession d'une province ». Il se dit prêt à appuyer la motion du Premier ministre à la condition que trois amendements soient acceptés. Le premier porte sur la sauvegarde de l'égalité des provinces ; le second sur la protection des droits des minorités au Québec ; la troisième sur la protection de l'intégrité du Canada. Il insiste enfin sur le fait que la motion ne saurait être considérée comme ayant une portée constitutionnelle.

Le discours de Lucien Bouchard se distingue autant par sa longueur, son caractère répétitif que par sa thématique. Il attaque Jean Chrétien pour son attitude « dure et rigide », pour sa

façon improvisée et précipitée de réagir devant la menace d'une victoire du Oui au référendum. Le chef de l'opposition s'étendra longuement sur l'histoire du rapatriement de la Constitution en 1982 et l'enchâssement de la Charte des droits et libertés dans le texte constitutionnel, sur celle de l'accord du Lac Meech dont il était à l'époque l'un des artisans, enfin sur celle de l'entente de Charlottetown. Son propos vise à démontrer le long processus d'édulcoration de la notion de « société distincte ». Pour lui, cette notion a perdu toute substance. Il jouera en permanence dans son discours sur l'opposition entre « société distincte » et « caractère distinctif du Québec ». Il critique enfin la notion de « société distincte au sein du Canada » en montrant que la perspective fédérale pose l'existence d'une seule nation au détriment d'une représentation alternative du peuple québécois. Enfin, il souligne le caractère « non liant » de la motion. Autant dire qu'elle n'aura aucune portée réelle.

### **3. Application de diverses techniques lexicométriques**

Nous nous sommes d'abord intéressés à ce corpus dans le contexte d'élaboration d'une démonstration en ligne des fonctionnalités de SATO. La première démonstration a été effectuée sur la version en langue française des discours des chefs de Parti. Le commentaire principal sur le calcul de distance entre le discours de Lucien Bouchard et celui des autres chefs concernait l'utilisation du terme « distinctif » chez le premier et « distincte » chez les seconds. L'explication sur laquelle nous reviendrons réside dans le travail sémantique opéré par le chef de l'opposition dans son analyse de la compréhension différentielle du caractère particulier du Québec dans l'ensemble canadien. Lorsque nous avons voulu reproduire cette démonstration sur la version en langue anglaise du même corpus, cet élément central de l'analyse n'apparaissait plus dans le tableau de distance. Nous avons donc décidé d'appliquer un ensemble de techniques lexicométriques aux deux versions du corpus afin d'évaluer la variation des résultats et leurs conséquences sur l'interprétation sociologique que nous pouvons en faire. Nous procéderons donc systématiquement à la présentation de résultats d'analyses statistiques produites par divers outils logiciels et tenterons d'évaluer les variations selon la langue pouvant induire des interprétations différentes ou divergentes.

#### ***3.1. Analyse factorielle des correspondances (Lexico)***

Le module d'AFC de Lexico nous permet de comparer la situation des trois locuteurs dans un plan factoriel sans pour autant nous fournir l'information sur le vocabulaire qui nous permettrait d'interpréter les axes. Les trois locuteurs apparaissent très éloignés. Sur un premier axe, on constate une nette distinction entre Bouchard et les deux autres locuteurs. Sur le deuxième axe, les discours de Chrétien et de Manning se distinguent tout aussi nettement, ce qui correspond à l'intuition de la lecture initiale. Leur éloignement relatif est le même peu importe que l'on choisisse le corpus en français ou en anglais. Il existe un effet de miroir entre les deux langues qui laisse apparaître une image inversée de la position de chaque locuteur, mais ce renversement n'a aucune incidence sur l'interprétation.

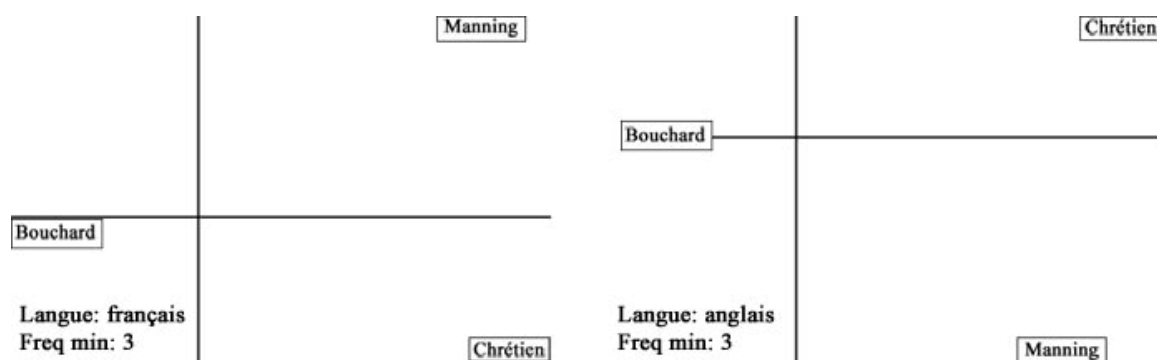


Figure 1 : AFC de Lexico pour le français et l'anglais

La représentation par AFC du lexique en français ou en anglais est également stable selon que l'on choisisse un seuil de fréquence minimale bas ou élevé. On peut donc dire qu'à la lumière d'une analyse globale des lexèmes et de leurs relations, tel que nous l'offre l'AFC, l'interprétation du corpus en français ou en anglais ne dépend pas de la langue choisie.

### 3.2. Analyseur de distance (SATO)

L'indice de distance en SATO permet d'établir la distance sur la base du chi carré entre deux corpus pris deux à deux. On peut ainsi contraster le discours de Lucien Boucharde avec celui des deux autres chefs, établir la distance entre deux locuteurs uniques ou encore comparer le discours d'un locuteur avec l'ensemble du corpus. Le calcul de distance établit le poids relatif et cumulatif des mots expliquant la distance entre les divers locuteurs. Le calcul peut s'effectuer sur l'ensemble des formes (fonctionnelles ou pleines) et tient compte des mots faiblement occurrents. Lorsque l'on considère la distance entre le discours de Lucien Boucharde et celui des deux autres locuteurs, on constate d'abord la présence forte de formes fonctionnelles (en français : « , », « on », « a », « - », « ça », « qui », « qu' », « là », « était », « il », « tout », « avait », « ...mais », « beaucoup », « donc », ...; en anglais, « , », « was », « - », « ? », « », « he », « but », « so », « his », « very », « did », « had », ...). Évidemment, on n'obtient pas exactement les mêmes formes fonctionnelles dans les deux langues, mais leur présence massive indique, d'une part, un style narratif, d'autre part, la présence de phrases complexes emboîtant les propositions. En français, on remarque des marqueurs d'arguments (« mais », « donc ») et en anglais, les marqueurs (« but », « so ») qui semblent correspondre à la nature rhétorique du discours de Boucharde.

Lorsque l'on compare les formes pleines dans les deux versions, on obtient des résultats largement comparables. Une des deux versions linguistiques étant la traduction officielle du discours prononcé dans l'autre langue, il n'est pas surprenant de noter une très grande concordance entre les deux lexiques. Cependant, deux mots expliquant le plus la distance ne sont présents que dans une langue : en français, le mot « distinctif », en anglais, le mot « political ». Dans ce dernier cas, en comparant les concordances générées à partir des occurrences de « political » et de « politique », on obtient une première explication de l'usage prépondérant du mot « political » par le locuteur Boucharde. Dans l'ensemble du corpus en anglais, 18 occurrences sur 19 appartiennent à Boucharde, alors que dans la version française, l'usage du mot « politique » chez Boucharde compte pour 18 occurrences sur 23, ce qui expliquerait la moindre prépondérance de « politique » par rapport à

Français	Distance	Spécificité	Anglais	Distance	Spécificité
peuple	0,37	6	Accord	0,37	5
<b>distinctif</b>	0,30	5	English	0,29	4
accord	0,30	5	People	0,29	4
Meech	0,28	5	<b>Political</b>	0,27	4
anglais	0,25	5	1982	0,23	4
caractère	0,23	4	Nature	0,21	3
1982	0,18	4	Agreement	0,16	3
Lac	0,14	Nil	Sovereignty	0,16	3
souveraineté	0,11	Nil	Constitution	0,15	3
canadien	0,10	Nil	Charter	0,14	2
charte	0,10	Nil	Negotiate	0,12	Nil

Tableau 1. : Mots pleins expliquant le plus la distance du texte de Bouchard

« political ». Mais, l'examen des contextes nous indique un usage très particulier du terme « politique » dans le discours original prononcé par Bouchard. Alors que dans la version française les deux autres chefs utilisent l'expression pour désigner la dimension institutionnelle du politique (« formation » et « parti » chez Chrétien et « chef », « actuelle » et « linguistique » chez Manning), Bouchard se réfère presque exclusivement aux pratiques et aux attitudes politiques (paysage, itinéraire, carrière, efficacité, capacité, courtoisie, rectitude, instinct). Le traducteur a donc rendu l'usage que fait Bouchard de l'épithète « politique » en français par le qualificatif « political ». Il est intéressant de noter que cet usage particulier du mot politique chez Bouchard ne nous est pas apparu à l'examen du lexique généré par l'analyseur de distance en français, mais à partir de la présence significative du mot « political » en anglais. Cela s'explique du fait de la polysémie du mot politique en français, alors qu'en anglais, il existe plusieurs mots pour désigner les divers aspects du politique (« politics », « policy », « policies », « political »). Il resterait à comprendre comment le mot « politique » utilisé cinq fois en français par les autres locuteurs est traduit en anglais. Un outil d'alignement des deux versions du corpus serait ici d'une grande utilité (Zimina, 2004 ; Martinez et al, 2002).

L'importance que revêt le mot « distinctif » en français ne se retrouve pas dans la traduction anglaise. Alors qu'en français, nous retrouvons respectivement 22 et une occurrences des adjectifs « distinctif » et « distinctive », en anglais, nous avons une seule occurrence de l'adjectif « distinctive » et neuf occurrences du nominal « distinctiveness ». Cela montre de nouveau que les résultats sont sensibles à la langue et que leur interprétation est toujours sujette à caution. Ici, l'analyse sur le corpus français a bien saisi le travail rhétorique effectué dans sa langue natale par Bouchard sur le caractère distinct du Québec. La traduction a dû recourir en anglais à une phraséologie plus complexe pour rendre la même idée, ce qui fait disparaître un indice important d'un fonctionnement discursif dans le tableau de distance en anglais.

Fréqtot	français	Fréqtot anglais
10	distinct	57 distinct
32	distincte	1 distinctive
22	distinctif	9 distinctiveness
1	distinction	
1	distinctive	
2 distincts		

Tableau 2. : *Lexiques comparés des formes dérivées de distinct*

Lucien Bouchard se fait un devoir de critiquer le concept de société distincte qui, selon lui, a été progressivement édulcoré jusqu'à ne plus rien signifier. Les 32 occurrences de l'adjectif « distincte » renvoie à l'expression « société distincte » qui est reprise dans la motion du Premier ministre, mais qui appartient aussi à l'histoire des négociation entourant l'accord du Lac Meech, autre notion qui caractérise le discours de Bouchard. Pour le reste, les termes « distinct », « distinctif », « distinction », « distinctive » et « distincts » renvoient à une affirmation du caractère particulier du Québec. En anglais, le traducteur a trouvé quatre expressions différentes pour traduire ce propos : « distinctiveness », « distinct nature », « distinct character », « distinct identity », ce qui a pour effet d'indifférencier les divers emplois du terme « distinct ».

### 3.3. *Analyseurs de spécificité et de segments répétés (Lexico)*

Le tableau 1 présente un extrait du lexique des distances caractéristiques du discours de Bouchard, augmenté d'une propriété accordant à chaque entrée lexicale l'indice de spécificité de Lexico. Le choix d'un seuil de cinq pour l'application de l'analyse de distance nous fournit un lexique beaucoup plus élaboré que celui produit par l'analyse des spécificités. Ainsi, en français, les quatre derniers mots caractéristiques n'apparaissent pas dans le lexique des spécificités. Cependant l'essentiel des résultats en SATO sur la présence significative des mots « political » et « distinctif » est confirmé par Lexico.

Segments répétés en français		Segment répétés en anglais	
<i>Les segments de 4 mots et plus :</i>		<i>Les segments de 4 mots et plus:</i>	
Bouchard	90	Bouchard	32
Chrétien	Moins de 30	Chrétien	33
Manning	Moins de 30	Manning	Plus de 50
<i>Les segments de 5 mots et plus:</i>		<i>Les segments de 5 mots et plus:</i>	
Bouchard	26	Bouchard	9
Chrétien	17	Chrétien	11
Manning	7	Manning	17
<i>Les segments de 6 mots et plus:</i>		<i>Les segments de 6 mots et plus:</i>	
Bouchard	10	Bouchard	4
Chrétien	8	Chrétien	9
Manning	1	Manning	12

Tableau 3. : *Nombre de segments répétés selon les locuteurs*



Lorsqu'on applique l'analyseur des segments répétés selon les locuteurs, on constate une variation en fonction de la langue pour deux d'entre eux. Par contre, le discours de Jean Chrétien conserve un nombre relativement semblable de segments répétés dans les deux langues. En ce qui concerne Bouchard et Manning, l'importance des segments répétés s'inverse dans le processus de la traduction. En français, le sous-texte de Bouchard contient le plus de segments, quel que soit le nombre de mots retenus. En anglais, c'est le discours de Manning qui semble en comporter davantage. Ces résultats s'inversent tout simplement dans le passage à l'autre langue. Faut-il penser que le style de ces deux locuteurs est moins neutre que celui du Premier ministre dont le discours semble comporter une dimension plus institutionnelle ? Son discours serait ainsi soigneusement produit dans les deux langues officielles, en contrôlant le libellé de chaque énoncé.

Retenons l'exemple suivant : le Premier ministre Chrétien utilise quatorze fois le segment répété « au sein du Canada » soit pour parler de la société distincte ou de la place que le Québec doit y occuper. Cette expression est systématiquement traduite en anglais par le syntagme « within Canada ». On n'insistera jamais assez sur l'importance capitale dans son discours de figer l'expression « société distincte au sein du Canada » et que cette expression reçoive une formulation tout aussi figée en anglais. Il s'agit d'encadrer cette reconnaissance dans le cadre de l'unité canadienne et non d'insister sur la différence du Québec.

Le nombre très élevé de segments répétés dans le discours prononcé en français par Bouchard indique le caractère réitératif de son discours. Ce discours qui est trois fois plus long que chacun des deux autres comporte un grand nombre de redites. Parmi celles-ci, le locuteur utilise 22 fois le segment « caractère distinctif du Québec » qu'il oppose au segment « société distincte » qui est compris dans le sens de sa reconnaissance dans « l'accord du Lac Meech ». Il s'agit, pour lui, de disqualifier toute tentative de réintroduire cette reconnaissance atténuée dans un contexte d'affirmation de l'unité canadienne et d'insister, au contraire, sur la réalité de la différence du Québec. On trouve enfin un grand nombre formules ou tics de langage qui sont propre à un discours improvisé (« c'est-à-dire que », « qu'il n'y a », « on a pas », « ce qu'il y a »). La première explication du nombre réduit de segments figés en anglais tient probablement au fait que le mot « distinct » (57 occurrences) est combiné dans plusieurs expressions différentes, réduisant d'autant la redondance des expressions figées dans cette langue. La seconde tient au travail stylistique du traducteur qui fera varier les expressions courantes trop souvent répétées.

Le seul discours originalement prononcé en anglais est celui de Manning. Il comporte dans cette langue le plus grand nombre de segments répétés. L'explication tient au caractère très répétitif de son discours dont l'intention est de proposer des amendements à la motion du Premier Ministre. On y trouve ainsi les expressions : « the Prime Minister's motion » et « of the federal government to ». Par ailleurs, Manning reprend des formulations à l'identique comme « the inclusion of a clear statement that nothing shall » (trois fois) visant à prévenir le gouvernement fédéral du danger de reconnaître un statut au Québec au détriment des minorités qui y résident. Il reprend de même la formulation « if it puts its commitment to distinct society ahead of » (trois fois) pour signifier de nouveau son opposition à tout traitement privilégié du Québec face aux minorités. On peut aisément comprendre que le traducteur a fait varier ces expressions en français, faisant ainsi disparaître la trace de cet indice du discours de Manning.

### 3.4. Analyse de classification descendante hiérarchique (Alceste)

Nous avons vu que l'AFC distingue, sur le premier axe, le discours de Bouchard des deux autres et qu'elle sépare, sur le second facteur, les discours de Chrétien et de Manning. Elle ne nous apprend rien, du moins dans la version Lexico, sur les mots qui expliquent ces différences. Alceste propose une classification descendante hiérarchique, tout en nous donnant un aperçu du contenu des classes et de leur proximité avec les locuteurs. Toujours selon notre protocole, nous avons appliqué l'analyseur aux deux versions linguistiques du corpus. La lecture des résultats nous rassure immédiatement sur le sens général qui ressort des données. Cependant, un examen plus attentif de la classification nous fait constater une différence selon la langue des corpus. De manière générale, on observe que l'analyse appliquée à la version française distingue bien le discours de Bouchard de celui de Manning et Chrétien qui forment un seul groupe. En anglais, la classification distingue plus clairement les trois locuteurs tout en montrant une certaine intersection entre Manning et Chrétien. En s'arrêtant au discours de Bouchard, on constate que l'analyse sur le Français distingue quatre classes portant sur les thématiques du « peuple », de la « finance », de « Meech » et de « distinctif ». En anglais, n'apparaissent que deux classes dont la plus importante est dominée par le thème « people » et l'autre par le « Meech Lake Accord ». Deux classes relativement importantes disparaissent, celle de la « finance » et celle du « caractère distinctif ». Dans le cas de *finance*, la connaissance du texte nous incite à croire que ce thème est relativement marginal dans le discours de Bouchard. Dans le second, nous retrouvons de nouveau le problème de la traduction du terme distinctif en anglais. L'analyseur est sensible à cette variation dans les expressions introduites par la traduction. Par contre, une nouvelle application de l'analyseur au seul texte de Bouchard en anglais nous permet de retrouver quatre classes qui sont assez proches des résultats en français. Nous obtenons ainsi trois classes à peu près d'égale importance dont les thèmes sont pour la première « Meech », pour la seconde, « distinct », « identity » et pour la troisième, « charter », « recognition », « distinctiveness », « nature ». La quatrième classe de moindre importance correspond au thème de « Charlottetown ». On ne retrouve pas finance, mais l'essentiel des classes d'abord identifiées dans l'analyse de la version française resurgit dans celle de la version anglaise du discours amputé des sous-textes des deux autres locuteurs. Ce qui est intéressant ici, c'est que la comparaison des analyses sur les deux versions du corpus intégral ne donnent pas les mêmes résultats, mais qu'une analyse sur le corpus restreint en anglais restitue les classes d'abord identifiées en français.

Sachant que Chrétien et Manning se distinguent de Bouchard sur le premier axe de l'AFC, il n'est pas surprenant de constater une certaine proximité entre leur discours. Ce qui est intéressant cependant, c'est de noter que leur discours n'est pas distingué par l'analyse Alceste appliquée au corpus en français. Les deux locuteurs sont présents dans la même classe bien qu'avec un poids inégal (Chrétien, Chi2 : 67,38 ; Manning, Chi2 : 38,10). La classe se caractérise par les thèmes : « gouvernement », « société », « motion », « distinct », « sein », « droit », « majorité », « minorité ». L'analyse de la version anglaise montre à la fois une intersection entre les deux discours et une différence. Le discours autour des mots « society », « motion », « recognize » et « veto » est associé d'abord à Chrétien (Chi2 : 38,48) puis à Manning (Chi2 : 5,26). Le discours caractérisé par « amend », « minority », « language », « respect », et « clear » appartient avant tout à Manning (Chi2 : 38,48) et à Chrétien (Chi2 : 3,14). Comment expliquer cela ? La traduction en français fait probablement perdre la spécificité du discours de Manning qui, rappelons-le est le seul à prononcer son discours en anglais. Lorsque l'on produit l'analyse en français, le discours de Manning se rapproche de

celui de Chrétien. Lorsque l'on refait l'analyse en version anglaise, alors la distinction se fait jour. Cette fois, c'est le discours de Chrétien qui est traduit. Mais, comme nous l'avons suggéré plus haut, le discours du Premier ministre semble avoir été préparé dans les deux langues ce qui expliquerait une moindre variation entre les versions linguistiques. Encore une fois, nous n'avons pas complètement perdu les indices conduisant à une interprétation possible du texte, mais nous voyons que la variation selon la langue et les outils que nous utilisons introduisent des modulations dans les interprétations possibles.

#### **4. Conclusion : l'interprétation**

La présence de corpus bilingues ou multilingues dans les régimes fédéraux ou les organisations transnationales pose de nouveaux défis aux chercheurs. Les traitements dans plusieurs langues peuvent générer des variations dans les résultats obtenus. Il est donc intéressant de voir comment, dans le contexte d'un corpus canadien bilingue, de telles variations apparaissent à la suite de l'application de divers analyseurs lexicométriques. Notre questionnement porte avant tout sur la validité des interprétations produites sur la base des résultats ainsi obtenus. L'idée qu'il existe une pluralité de choix herméneutiques qui s'imposent à toutes les étapes de la recherche implique que le processus herméneutique ne se déploie pas seulement au terme d'une recherche, mais dans l'ensemble des opérations de gestion, de description, d'exploration et d'analyse des données (Duchastel et Laberge, 1999). L'interprétation globale dépend donc d'une constellation d'interprétations locales effectuées tout au long du processus. Dans ce sens, les résultats obtenus suite à l'application d'analyseurs statistiques ne peuvent être considérés que comme des indices herméneutiques contribuant progressivement à produire une interprétation globale du phénomène étudié.

C'est ainsi que nous avons pu observer tout au long de la démarche d'analyse que certains indices apparaissent ou disparaissent selon que l'on applique l'un ou l'autre type d'analyse à l'une ou l'autre version linguistique du même discours. Nous avons d'abord constaté que la macroanalyse du corpus (AFC) appliquée dans les deux langues produit les mêmes résultats, distinguant dans un premier temps le discours de Bouchard et ceux des deux autres locuteurs et, dans un deuxième temps, les discours de Manning et de Chrétien. L'application de l'analyseur de distance (SATO) produit en gros des résultats comparables dans les deux langues, mais introduit des différences significatives concernant la présence ou l'absence de mots comportant un fort potentiel herméneutique. Nous avons vu que dans les cas de « distinctif » et de « political », l'explication tenait à la différence terminologique attachée à chacun de ces mots dans les deux langues. Cependant, si dans le cas de « distinctif », la production du lexique en langue anglaise faisait perdre un précieux indice explicatif du discours de Bouchard, dans le cas de « political », la traduction en anglais faisait apparaître un indice absent du vocabulaire en français. Toutefois, dans les deux cas, la comparaison entre les deux analyses poussait la réflexion plus loin. Le travail sémantique opéré par Bouchard était ainsi mis à nu de manière beaucoup plus évidente. L'analyse bilingue des segments répétés a ajouté à notre connaissance des pratiques de traduction au Parlement canadien. En effet, la similitude du nombre de segments répétés dans les deux langues nous a suggéré que le discours du Premier ministre était fort probablement écrit dans les deux langues officielles du Canada, alors que l'inversion de la prépondérance des segments chez les deux autres locuteurs dans le passage à l'autre langue, nous a fourni des indices sur la dimension stylistique du discours prononcé dans la langue d'origine et leur transformation par les traducteurs. Enfin, l'analyse Alceste, tout en confirmant l'essentiel des observations faites jusqu'alors, a montré que nous ne pouvions interpréter de manière littérale les classes, sans

procéder à une comparaison des versions linguistiques, voire sans reproduire les analyses sur des sous-ensembles du corpus.

Trois conclusions découlent de ce travail. D'abord, les observations particulières que nous autorisent les résultats d'analyse doivent être considérées comme des indices menant à des interprétations locales devant être corroborées par de nouvelles observations. Ensuite, les variations en fonction de la langue des corpus ou des outils d'analyse doivent servir d'incitation à approfondir la recherche. Enfin, la comparaison des analyses sur des versions linguistiques différentes a l'avantage de nous faire découvrir de nouvelles pistes d'interprétation.

## Références

- Bourque G. et Duchastel J. (1996). *L'identité fragmentée. Nation et citoyenneté dans les débats constitutionnels canadiens, 1941-1992*. Montréal, Fides.
- Daoust F. (1996, 2005). *SATO 4.3, Manuel de référence*. Centre ATO, UQAM, Montréal.
- Daoust F. et Duchastel J. (2007). Pluralisme et pluralité des interprétations. In *Humanités numériques I, nouvelles technologies cognitives et épistémologie*, 257-268.
- Deroubaix (2004). Que faire des corpus multilingues parallèles ? Une expérience. In *Le poids des mots, Actes des 7<sup>es</sup> journées internationales d'analyse statistique des données textuelles*, UCL Presses Universitaires de Louvain : 295-303.
- Jalam R. et al. (2004). Cadre pour la catégorisation de textes multilingues. In *Le poids des mots, Actes des 7<sup>es</sup> journées internationales d'analyse statistique des données textuelles*, UCL Presses Universitaires de Louvain : 650-660.
- Martinez W. et Zimina M. (2002). Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues, In *Actes des 6<sup>es</sup> journées internationales d'analyse statistique des données textuelles JADT. 2002*.
- Reinert, M (2002). *Alceste, Manuel de référence*. Université de Saint-Quentin-en-Yvelines, CNRS.
- Salem A. et al. (2003). *Manuel Lexico 3. Version 3.41*.
- Zimina M. (2004). Alignement textométrique des unités lexicales à correspondances multiples dans les corpus parallèles, In *Le poids des mots, Actes des 7<sup>es</sup> journées internationales d'analyse statistique des données textuelles*, UCL Presses Universitaires de Louvain : 1195-1202.