

La connexion intertextuelle

Application au discours gouvernemental québécois*

Dominique Labbé
Denis Monière

CERAT - IEP F 38040 Grenoble Cedex 9 <cerat@iep.upmf-grenoble.fr>
Université de Montréal <moniered@pol.umontreal.ca>

Abstract

The intertextual connection measures the distance between the vocabularies of several texts. On each word, we calculate the difference between expected and observed frequencies. Inequality between texts' lengths does not influence the indicator. The calculation is applied to the 49 "discours du trône" by Québec' s Prime ministers since 1945. An automatic clustering brings to light the major events of Québec' s political life and underlines the singularity of two periods during which the nationalist party ("Parti Québécois") has been in charge (1977-84) and (1996-).

Résumé

La connexion intertextuelle mesure la distance entre les vocabulaires de plusieurs textes. Pour chacun des mots, on calcule la différence entre une fréquence théorique et la fréquence observée. L' indice est insensible aux différences de longueur entre les textes. Il est appliqué aux discours prononcés par les Premiers ministres québécois pour ouvrir les sessions parlementaires depuis 1945. Appliquée à ces données, la classification automatique met en valeur quelques grands épisodes dans la vie politique de la province et souligne la singularité des deux passages au pouvoir du parti québécois (1977-84 et 1996-).

Mots-clés : statistique textuelle - connexion intertextuelle - distance - classification - discours politique - québec

Dans un corpus, quels sont les textes qui sont les plus proches entre eux ou les plus éloignés ? Pour répondre à cette question, deux démarches doivent être distinguées. D' un e part, l' étude de la "connexion entre les vocabulaires" : on considère la présence ou l' absence de chaque vocable, dans chacun des textes, sans considération de sa fréquence (pour une solution à ce problème, voir Labbé et Hubert, 1998). D' autre part, la "connexion des textes" : la comparaison porte sur des surfaces et vise à répondre à la question : combien de mots sont communs ou encore : quelle proportion des deux textes leur est propre ? (d'où le terme de connexion "inter-textuelle"). C' est à cette dernière approche que nous allons nous intéresser. Le calcul sera illustré grâce au corpus des "discours du trône" prononcés depuis 1945 par les Premiers ministres du Québec.

I. La connexion intertextuelle

Ce problème a été abordé pour la première fois par Ch. Muller sous le nom de "connexion lexicale" (Muller, 1967 et 1977). Les indices utilisés n' étaient pas indépendants de la dimension des textes aussi Muller s' est-il contenté de quelques expériences sur des tranches égales extraites au hasard dans les textes à comparer. E. Brunet a repris le problème et en a

* Les auteurs remercient S. Bolasco, P. Hubert, L. Lebart et X. Luong pour leur lecture attentive de ce texte et leurs nombreuses remarques qui ont permis de l' améliorer sensiblement.

donné une application sur le vocabulaire de Giraudoux (Brunet, 1988). Il propose de calculer la distance entre deux textes A et B de la manière suivante :

$$\delta_{(A, B)} = \frac{(N_a - N_{ab}) + (N_b - N_{ab})}{N_a + N_b}$$

avec : N_{ab} : nombre de mots communs aux deux textes ; N_a : taille de A (nombre de mots) ; N_b : taille de B.

On obtient un indice qui peut varier entre zéro et un et rend les résultats de différentes expériences comparables entre eux. Cependant, 0 ne peut être obtenu qu'avec $N_a = N_b$, et 1 avec quelques cas particuliers, notamment : $N_a = N_b = 0,5N_{ab}$. Le minimum de la formule de Brunet s'accroît proportionnellement à la différence de taille entre les deux textes et il tend vers 1 lorsque cette différence est très grande et que le plus petit des deux vocabulaires est contenu dans le plus grand. Dans son manuel de 1977, Ch. Muller suggère incidemment que la solution consisterait, pour chaque vocable, à comparer des effectifs théoriques avec les fréquences observées. Nous proposons de développer cette piste.

Soit : x un vocable quelconque ; $F(x,a)$ sa fréquence dans le texte A ; $F(x,b)$ sa fréquence dans le texte B ; N_a et N_b les tailles respectives de ces deux textes.

Si x est également présent dans les deux textes, les fréquences attendues de x dans A et B, en fonction de sa fréquence dans l'autre texte sera :

$$E_{(x, b)} = F_{(x, a)} \times \frac{N_b}{N_a} \quad \text{et} \quad E_{(x, a)} = F_{(x, b)} \times \frac{N_a}{N_b}$$

On calcule la différence entre les observations et les valeurs attendues. La distance entre les fréquences d'emploi de x dans A et B sera la moyenne de ces deux différences :

$$D_{(x, a, b)} = \frac{|F_{(x, a)} - E_{(x, a)}| + |F_{(x, b)} - E_{(x, b)}|}{2}$$

Cette distance absolue varie donc entre zéro, si les valeurs observées sont égales aux valeurs attendues, et un maximum (δ_{\max}) en cas d'absence du vocable dans A ou B :

$$\text{avec } F_{(x, a)}=0 \quad \delta_{\max}(x_a, x_b) = E_{(x, a)} + F_{(x, b)}$$

$$\text{avec } F_{(x, b)}=0 \quad \delta_{\max}(x_a, x_b) = E_{(x, b)} + F_{(x, a)}$$

Calculé avec (1), $D_{(x, a, b)}$ présente un intérêt limité puisqu'il variera en fonction de la taille des textes. Pour neutraliser ces différences de taille, on calcule une distance relative :

$$D_{(x, a, b)} = \frac{|F_{(x, a)} - E_{(x, a)}| + |F_{(x, b)} - E_{(x, b)}|}{N_a + N_b}$$

Ainsi D variera entre "0" (identité) et "1" (écart maximum). Cet indice est directement inspiré des coefficients, dits "de Jaccard", habituellement utilisés pour l'étude des variables qualitatives de nature binaire. Le choix du dénominateur peut être discuté. Il s'agit de chiffrer la distance maximale que peuvent avoir théoriquement les deux fréquences.

Considérons maintenant l'ensemble des textes. La distance entre eux sera le total des distances, calculées sur les V vocables différents contenus dans l'ensemble A U B, rapporté à la distance maximale théorique, calculée en ramenant A à la taille de B puis B à la taille de A :



La signification des résultats de (3) est immédiatement accessible. Deux textes "identiques" — c'est-à-dire partageant la surface commune théorique maximale — sont séparés par une distance nulle ; le maximum théorique, 1, est obtenu quand les deux textes n'ont aucun mot en commun (leur surface est totalement différente) ; un indice de 0,5 signifie que les textes ont en commun la moitié de leur surface théorique commune maximale, etc.

On remarque que :

— l' "identité" ne signifie pas pure "répétition" (avec les différences d' échelle) : les mots peuvent être placés dans un ordre différent ; les temps des verbes modifiés ou le genre et le nombre des adjectifs... Autrement dit, deux textes dont la distance est faible, s' ils ne disent pas la même chose, partagent le même univers intellectuel et sont écrits dans un style semblable ;

— un indice égal à 1 est également improbable car, s' ils utilisent la même langue, les locuteurs sont condamnés à utiliser les outils de cette langue (déterminants, pronoms, verbes auxiliaires...) Mais plus on se rapprochera de 1, plus les textes appartiendront à des genres et à des univers intellectuels différents, plus ils développeront des thèmes éloignés ;

— les grandes distances peuvent aussi s' expliquer par des différences de genre. Chez un même auteur, l' écrit et l' oral, la prose et la poésie... utilisent des vocabulaires souvent éloignés. D'où la nécessité de comparer des textes appartenant au même genre (ce que nous faisons ci-dessous avec les déclarations gouvernementales).

La formule (3) remplit les principales conditions que l'on exige habituellement d'un indice de proximité et permet de comparer sans biais le vocabulaire de deux ou plusieurs textes de tailles différentes. Cependant, la technique ne doit pas être employée sur des textes trop brefs (empiriquement, nous avons fixé un minimum de 500 mots) et, quand les tailles sont fortement différentes, il faut résister à la tentation de mesurer les distances avec trop de décimales...

Lorsque le calcul aura été appliqué à un grand nombre de corpus divers, il sera même possible d' étalonner empiriquement une "échelle" des distances. Grâce à quelques expériences, il est possible d' esquisser cette échelle

— en dessous de 0,15 : les textes appartiennent au même registre et au même genre, ils ont un même auteur ou, s' ils ont des auteurs différents, ils développent un même thème avec des formulations extrêmement proches.

— de 0,15 à 0,25 : les textes appartiennent au même registre mais, s' ils sont d'un même auteur, ils diffèrent par le genre ou développent des thèmes assez éloignés. S' ils appartiennent aux mêmes registre et genre, mais sont d' auteurs différents, les thèmes sont relativement proches ;

— de 0,25 à 0,35 : pour un même auteur, les textes appartiennent probablement à des genres ou à des registres différents. En tous cas, les thèmes développés sont éloignés. Pour des textes de même registre et de même genre, mais d' auteurs différents, les thèmes sont assez éloignés

— au-dessus de 0,35 : les auteurs ou les registres sont différents ; les thèmes sont éloignés.

II. Les discours du trône au Québec (1945-1998)

Nous avons mis la méthode à l'épreuve sur les discours d'ouverture, prononcés devant le parlement québécois, de l'après-guerre jusqu'en 1996, soit 49 textes (voir bibliographie en annexe et Monière, 2000, pour une présentation du corpus).

Conformément à la tradition parlementaire britannique, le discours inaugural est prononcé au début de chaque session. Il est préparé par le Premier ministre et approuvé par le cabinet. "It is considered by cabinet to be the most important public statement of its political intentions" (Jackson-Atkinson, 1980). Il expose les objectifs généraux du gouvernement ainsi que les mesures législatives spécifiques qu'il entend faire adopter.

Ces discours ont été saisis selon les normes de Saint-Cloud puis lemmatisés (Labbé, 1990). Le corpus ainsi traité comprend 191 724 mots ou occurrences, 11 557 formes graphiques normalisées différentes et 6 654 vocables. La longueur de ces discours est très variable : le plus bref a 246 mots (1970) et le plus long 16 440 (1981). Les considérations présentées précédemment concernant l'influence de la longueur et l'inadaptation du calcul aux textes trop brefs, nous ont conduit à écarter les 6 discours de taille inférieure à 1.000 mots (1964, 1969, 1970(A et B), 1981B, 1983).

En premier lieu, nous avons mesuré la distance séparant chacun de ces textes à tous les autres. Cette opération permet de distinguer les discours les plus proches, par leur vocabulaire, et les plus éloignés (nous donnons dans le tableau 1 un aperçu des couples de discours les plus singuliers).

Les proximités les plus grandes se retrouvent généralement entre des gouvernements d'une même législature ou d'une même orientation politique. Dans ce dernier cas, ils peuvent être séparés par un laps de temps important. Ainsi le vingtième couple (Lesage 1966A et Bourassa 1989) enjambe plus de 20 ans d'histoire mouvementée et réunit deux gouvernements libéraux. Les plus fortes distances séparent essentiellement un texte de 1981 (René Lévesque, Parti Québécois) d'avec des gouvernements de l'Union nationale (1944-1958 et 1966B-70), du Parti libéral (1960-66 ; 1970-76 ; 1985-94) mais aussi avec un texte du même Lévesque prononcé la même année (1981C), ce qui signale l'ampleur de la rupture intervenue au cours de cette année cruciale.

De la même manière, on peut mesurer la plus ou moins grande singularité d'un texte dans un corpus en calculant la distance moyenne de ce texte par rapport à tous les autres (tableau 2).

Tableau 1. Distances les plus caractéristiques entre discours du trône

Les couples de textes les plus proches :			Les couples de textes les plus distants :		
Bourassa 1973B	Bourassa 1974	0,07	Duplessis 1945	Lévesque 1981A	0,53
Lévesque 1979	Lévesque 1980	0,08	Bourassa 1976	Lévesque 1981A	0,54
Bourassa 1973A	Bourassa 1973B	0,08	Bourassa 1975	Lévesque 1981A	0,54
Bourassa 1973A	Bourassa 1974	0,08	Bourassa 1973A	Lévesque 1981A	0,54
Lévesque 1977	Lévesque 1979	0,09	Bourassa 1973B	Lévesque 1981A	0,55
Lévesque 1977	Lévesque 1980	0,09	Lesage 1960	Lévesque 1981A	0,56
Bourassa 1974	Bourassa 1975	0,10	Duplessis 1957	Lévesque 1981A	0,56
Bourassa 1973B	Bourassa 1975	0,10	Bourassa 1974	Lévesque 1981A	0,57
Lévesque 1980	Lévesque 1984	0,10	Duplessis 1946	Lévesque 1981A	0,58
Lévesque 1979	Lévesque 1984	0,11	Duplessis 1954	Lévesque 1981A	0,59

Tableau 2. Les textes les plus centraux et les plus décalés du corpus

Les textes les plus centraux	1	Bouchard 1996	0,25
------------------------------	---	---------------	------

2	Bourassa 1994A	0,25	1	Lévesque 1981A	0,46
3	Duplessis 1958	0,26	2	Bourassa 1985	0,42
4	Johnson 1966B	0,26	3	Lévesque 1981C	0,33
5	Bourassa 1989	0,26	4	Bourassa 1976	0,32
6	Duplessis 1955	0,26	5	Lévesque 1978	0,32
7	Lesage 1966A	0,26	6	Lesage 1960	0,32
8	Bourassa 1992	0,26	7	Sauvé 1959	0,31
9	Lévesque 1984	0,27	8	Lévesque 1983	0,31
10	Duplessis 1948	0,27	9	Parizeau 1994B	0,30
Les textes les plus décalés			10	Duplessis 1947	0,30

Le dernier discours du corpus est aussi le plus "central". Il a été écrit par un cabinet du Parti québécois (Lucien Bouchard, 1996). Mais beaucoup de textes "décalés" proviennent de ce même parti, notamment de son premier passage au pouvoir (1976-84).

Enfin, on obtient une distance moyenne entre tous les textes qui est une mesure d'hétérogénéité du corpus. Dans les discours du trône québécois, elle est de 0,30 avec un coefficient de variation relative autour de cette moyenne de 29%. Cela signifie que, malgré la contrainte assez forte du moule institutionnel, les thèmes ont relativement fluctué au cours du dernier demi-siècle.

On peut donc se demander s'il n'existe pas des groupes ou des périodes plus homogènes dans ce corpus. La classification automatique répond à cette question. Les distances sont regroupées dans une matrice carrée où les textes sont rangés, en ligne et en colonne, par ordre chronologique ce qui donne une diagonale nulle. L'algorithme procède à la construction d'une classe en regroupant les deux textes séparés par la distance la plus faible, puis il recalcule les distances des autres textes par rapport à ce nouvel ensemble à l'aide de la formule (3), etc. Et ceci jusqu'à la constitution d'un groupe unique. Les regroupements successifs sont représentés dans un dendrogramme avec, en ordonnées, les distances correspondantes aux différents niveaux d'agrégation (tableau 4).

Nous avons indiqué plus haut que des textes ou groupes de textes d'auteurs différents, séparés par une distance inférieure à 0,25 peuvent être considérés comme ayant un vocabulaire proche et comme partageant une même thématique. En coupant le dendrogramme à cette hauteur, on obtient effectivement trois groupes assez homogènes.:

— le groupe (B), au centre du graphique, rassemble les discours antérieurs à 1960 et correspond à une période homogène idéologiquement : le règne ininterrompu de l'Union nationale, dirigée par M. Duplessis (sur le discours de cette période, on peut se reporter à Bourque-Duchastel, 1988) ;

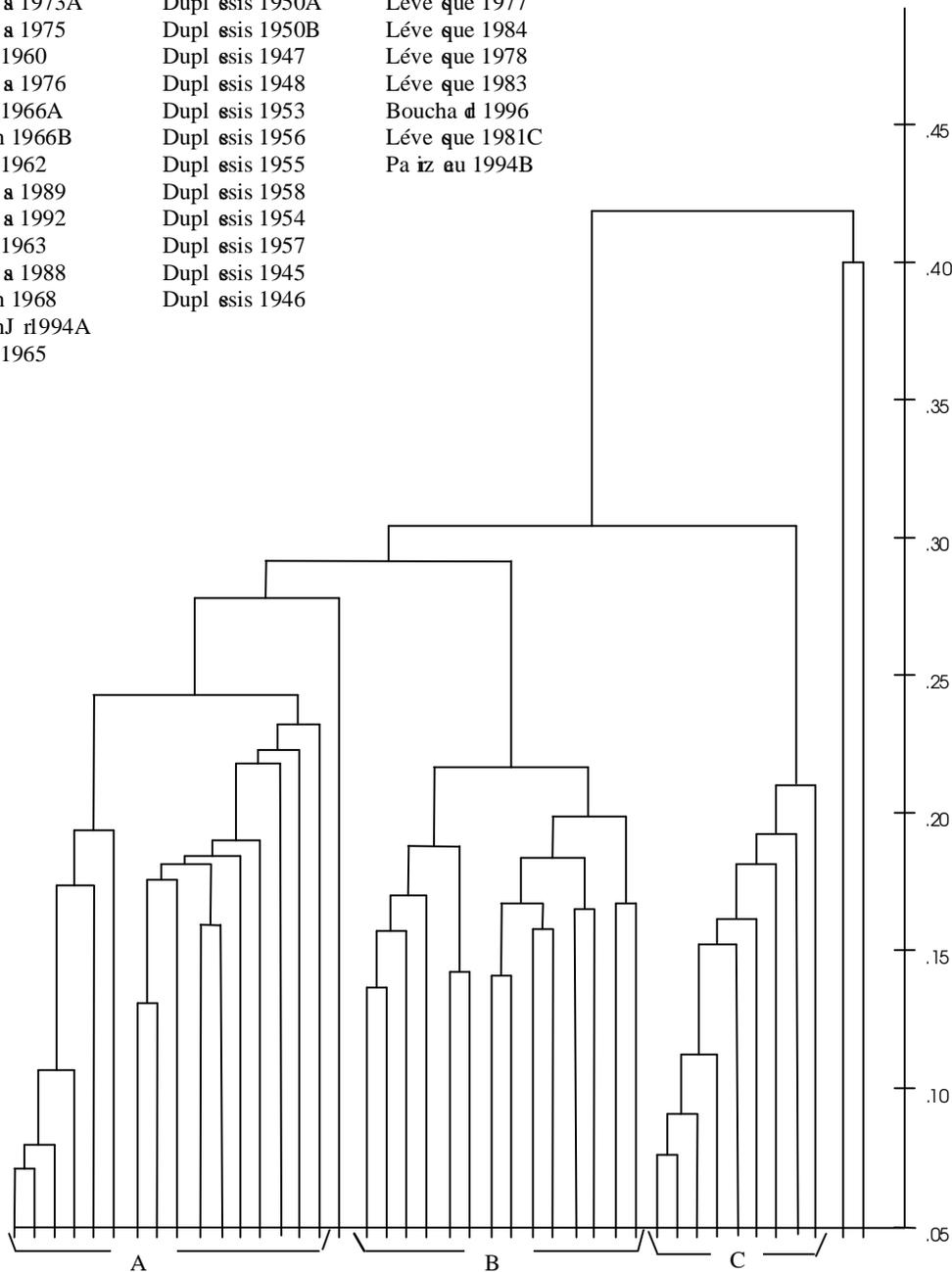
— le groupe (A), à gauche du graphique, paraît historiquement plus hétérogène puisqu'il regroupe des discours prononcés à des époques différentes, couvrant plus de trois décennies, par hommes appartenant à deux partis différents. Mais ces partis sont tous deux partisans du maintien de l'ordre constitutionnel canadien le Parti libéral (1960-1966, 1970-1976, 1985-1994) et l'Union nationale (1966-1970). Leur proximité de vocabulaire est significative d'une certaine parenté de pensée économique et sociale et d'une continuité dans la conception du rôle de l'Etat

Tableau 4. Classification automatique des discours inauguraux du Québec (1945-1996)

Groupe A :
 Bouras a 1973B
 Bouras a 1974
 Bouras a 1973A
 Bouras a 1975
 Lesage 1960
 Bouras a 1976
 Lesage 1966A
 Johnson 1966B
 Lesage 1962
 Bouras a 1989
 Bouras a 1992
 Lesage 1963
 Bouras a 1988
 Johnson 1968
 JohnsonJ r1994A
 Lesage 1965

Groupe B :
 Dupl ésis 1949
 Dupl ésis 1951
 Dupl ésis 1950A
 Dupl ésis 1950B
 Dupl ésis 1947
 Dupl ésis 1948
 Dupl ésis 1953
 Dupl ésis 1956
 Dupl ésis 1955
 Dupl ésis 1958
 Dupl ésis 1954
 Dupl ésis 1957
 Dupl ésis 1945
 Dupl ésis 1946

Groupe C :
 Léve que 1979
 Léve que 1980
 Léve que 1977
 Léve que 1984
 Léve que 1978
 Léve que 1983
 Boucha d 1996
 Léve que 1981C
 Pa iz au 1994B



— dans le troisième groupe (C), on trouve uniquement des discours du Parti québécois prononcés durant ses deux passages au pouvoir, même si ceux-ci sont séparés par une période de 9 ans (1985-1994), avec des gouvernements libéraux dont tous les discours se trouvent dans le groupe (A).

Enfin, la classification met en valeur trois textes dont le vocabulaire semble assez nettement

atypique : le discours de M. Sauvé, qui suit de peu la mort de M. Duplessis, se situe logiquement à mi-chemin des groupes A et B (à droite du graphique) : 1981A et 1985. Cette position à part ne surprend pas puisqu' il s' agit des deux discours les plus décalés par rapport à tous les autres (tableau 3). Cependant, ils ne sont pas tout à fait étrangers aux trois groupes précédents comme le montre le calcul de leur distance moyenne à chacun de ces groupes (tableau 3).

Tableau 3. Distance des vocabulaires de 1981A et de 1985 par rapport aux trois ensembles

distance de : à :	Groupe B (1945-1959)	Groupe A (1960-76 ; 1986-94)	Groupe C (1977-84 ; 1994-96)
1981A	0,491	0,516	0,323
1985	0,425	0,398	0,448

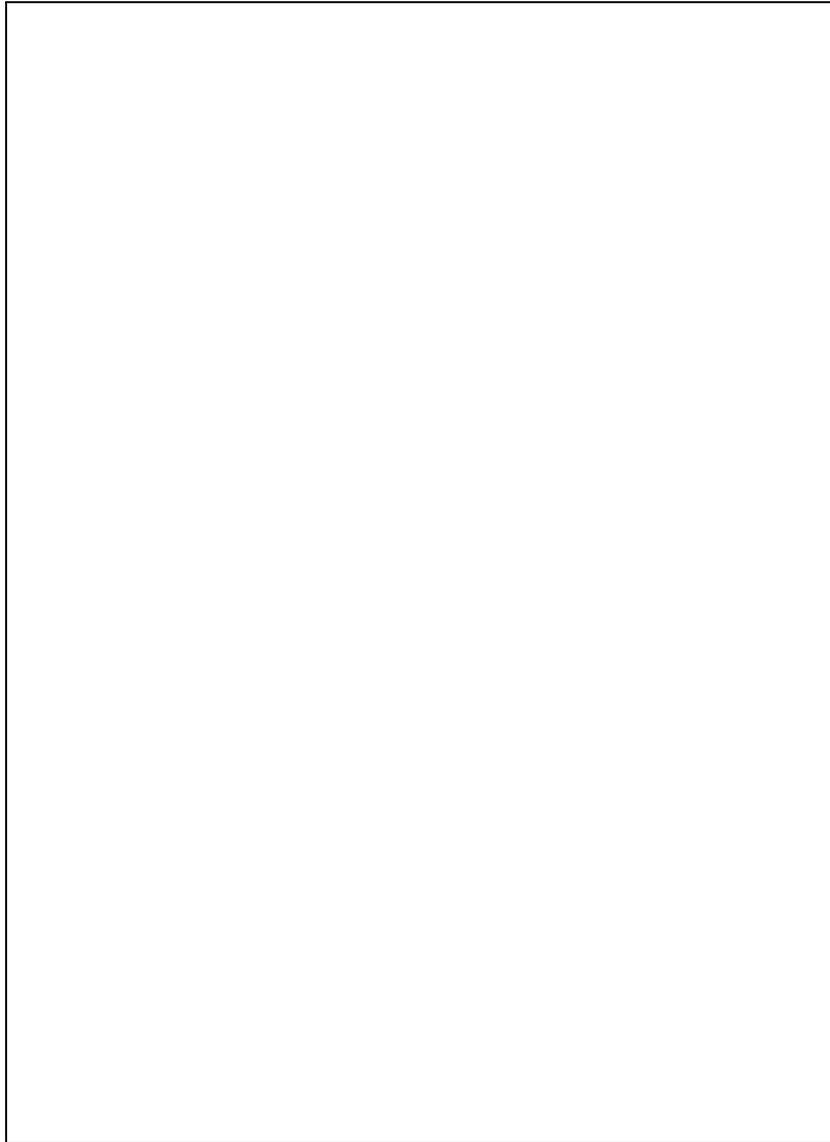
Il est donc possible de rattacher le premier texte au groupe C — ce qui est historiquement logique puisque le Premier ministre est R. Lévesque, leader du Parti québécois — et le second au groupe B, ce qui correspond, là encore, à la logique historique et politique... On obtient ainsi trois ensembles de textes relativement homogènes auxquels a été appliquée une recherche des spécificités du vocabulaire (Monière, 2000).

La technique du dendrogramme présente cependant quelques inconvénients connus. Certaines proximités entre textes ne sont pas discernables car les sommets qui les relie sont effacés par des agrégations effectuées à un niveau inférieur. D' autres représentations peuvent limiter cet inconvénient, notamment le graphe arboré (Barthélemy et Luong, 1998). La figure 4 a été réalisée par X. Luong sur le même tableau des distances (programmes dans Barthélemy-Guénoche, 1988). Cette représentation complète bien le dendrogramme classique — on retrouve les trois groupes mentionnés ci-dessous — en donnant une représentation des distances relatives séparant chaque texte de tous les autres considérés un par un.

Pour vérifier ces conclusions, nous avons effectué une analyse des correspondances (avec SPADT) sur le même corpus lemmatisé (tableau 6). Le profil général du nuage rappelle la courbe en fer à cheval caractéristique des corpus chronologiques. Cependant, les années ne s'y distribuent pas selon la chronologie. On distingue nettement les trois groupes évoqués précédemment. Dans le quart nord-est, on trouve les années 1945-1959 mais positionnées selon une sorte de boucle. Par exemple, le dernier discours (1959) est plus proche des deux premiers (1945 et 1946) que de ceux des années 1950... Dans la partie sud, figurent tous les discours du groupe B mais, là encore, la disposition des points dans le nuage n'ob éit pas complètement à la chronologie, même si tous les textes des années 1960-70 se trouvent à droite du deuxième axe et ceux des années 1985-1994 à gauche de celui-ci. Enfin, dans le quart nord-ouest, se trouvent les textes du Parti québécois.

Sur le fond, les deux analyses convergent : le corpus des discours inauguraux québécois n'ob éit que très partiellement au temps. Dans la première moitié du dernier demi-siècle, soit jusqu' en 1976, il y a une relative similitude dans le choix du vocabulaire gouvernemental, les différences entre le groupe A et le groupe B étant attribuables à des effets de conjoncture comme en témoigne d' ailleurs la présence de l' Union nationale dans les deux groupes (à des époques différentes). Mais à partir de 1976, il se produit une rupture avec l' arrivée au pouvoir du Parti québécois.

Figure 4. Représentation arborée des distances entre les discours du Trône québécois



Depuis les années 1970, le vocabulaire politique québécois semble moins influencé par les effets de conjoncture que par l' affrontement de deux positions idéologiques. En effet, si l' écoulement du temps avait eu une influence prédominante sur le choix des mots, alors on se serait attendu à ce que les discours les plus proches chronologiquement — même s' ils émanent de partis opposés — soient moins éloignés que ceux tenus par un même parti à plus de vingt ans de distance. Or c' est l' inverse qui se produit puisque les discours tenus par les libéraux et les souverainistes, au cours de la dernière décennie, sont assez fortement éloignés les uns des autres et nettement plus proches du vocabulaire utilisé par ces mêmes partis dans les années 1970. Pourtant, le monde, le Canada et le Québec ont beaucoup changé au cours de ces 25 dernières années ! Mais le passage du temps ne semble guère affecter le choix des mots ce qui traduit la stabilité des deux partis, tant du point de vue idéologique que du style oratoire.

Figure5. Analyse des correspondances simples sur les discours inauguraux du Québec

Conclusions.

La classification automatique et l'analyse factorielle mettent bien en valeur la singularité de ce corpus qui, bien que s'étendant sur un demi-siècle, n'est pas d'abord un corpus "chronologique".

Sur ce cas précis, la classification, combinée avec le calcul de la distance intertextuelle, semble donner des résultats plus immédiatement lisibles que l'analyse factorielle. De manière générale, ces graphes, comme les analyses factorielles, ne doivent pas être utilisés aveuglément. L'appartenance de chacun des textes à une classe donnée doit être contrôlée sur la matrice des distances. Il n'est pas mauvais non plus de recalculer la distance moyenne de chacun des textes à l'ensemble de ses "voisins" supposés, comme nous l'avons fait pour deux textes atypiques.

Nous ne prétendons pas résoudre d'un coup le problème classique et redoutable des mesures de proximité. Outre son caractère expérimental, ce calcul appelle quelques remarques :

— la formule (3) permet de comparer plusieurs textes de taille différents. Toutefois, ces différences quantitatives peuvent engendrer de véritables différences de nature quand les tailles sont trop éloignées. C'est pourquoi, de manière générale, il est recommandé d'exclure les "géants" et les "nains" ...;

— la distance intertextuelle n'a de signification que si les textes ont été dépouillés selon une norme stable et rigoureuse afin de ne pas interpréter les fluctuations dans les graphies comme

des distances réelles. Au fond, la comparaison exige que la toise soit la même pour tous les éléments comparés : il n' est pas sûr que les corpus "textuels" actuels répondent toujours à cette exigence.

— Enfin, si nos intuitions se trouvaient validées, il restera à explorer de nombreuses pistes, notamment le calcul de la "contribution à la distance" des différents mots et des catégories grammaticales.

Références

- Les discours du trône sont extraits des *Journaux de l'Assemblée législative de la province de Québec* (1944-1971) et des *Procès-verbaux de l'Assemblée nationale du Québec* (1971-1996).
- ARMONY Victor, DUCHASTEL Jules, BOURQUE Gilles (1998), "Quantitative Tools for Quantitative Analysis : an Indicator for Segment Typicality", in Sylvie MELLET (ed), *4e journées internationales d'analyse statistique des données textuelles*, Université de Nice, p. 11-18.
- BARTHELEMY Jean-Pierre, GUENOCHÉ Alain (1988), *Les arbres et les représentations des proximités*, Paris, Masson.
- BARTHELEMY Jean-Pierre, LUONG Xuan (1998), "Représenter les données textuelles par des arbres", in Sylvie MELLET (ed), *4e journées internationales d'analyse statistique des données textuelles*, Université de Nice, p. 49-71.
- BENZECRI Jean-Paul (1979), *L'analyse des données. 1. La taxinomie*, Paris, Dunod.
- BOURQUE Gilles, DUCHASTEL Jules, (1988), *Restons traditionnels et progressifs. Pour une nouvelle analyse du discours politique. Le cas du régime Duplessis au Québec*, Montréal, Boréal.
- BRUNET Etienne (1988), "Une mesure de la distance intertextuelle : la connexion lexicale", *Le nombre et le texte. Revue informatique et statistique dans les sciences humaines*, Université de Liège.
- JACKSON Robert, ATKINSON Michael (1980), *The Canadian Legislative System*, Toronto, MacMilan, p. 76.
- LABBE Dominique (1990), *Normes de saisie et de dépouillement des textes politiques*, Cahiers du CERAT n°7.
- LABBE Dominique, HUBERT Pierre (1998), "La connexion des vocabulaires", in Sylvie MELLET (ed), *4e journées internationales d'analyse statistique des données textuelles*, Université de Nice, p. 361-369.
- LEBART Ludovic, SALEM André (1994), *Statistique textuelle*, Paris, Dunod.
- MONIERE Denis, (2000), "Les mots du pouvoir : 50 ans de discours inauguraux au Québec : 1944-1996", à paraître dans *Mots*, automne 2000.
- MULLER Charles (1967), *Etude de statistique lexicale. Le vocabulaire de Pierre Corneille*, Paris, Larousse, p. 169-174.
- MULLER Charles (1977), *Principes et méthodes de statistique lexicale*, Paris, Larousse, p. 169-174, p. 145-154.