

Le discours gouvernemental. Canada, Québec, France (1945-2000)

Dominique Labbé et Denis Monière
2003, Paris, Champion, 181 p.

**Paroles de président. Jacques Chirac (1995-2003)
et le discours présidentiel sous la V^e République**

Damon Mayaffre
2004, Paris, Champion, 291 p.

La collection « Lettres numériques », dirigée par Étienne Brunet et François Rastier chez Champion, publie depuis 2000 des ouvrages de statistique linguistique et de linguistique de corpus proposés par des chercheurs chevronnés. Ouvrages couteux pour des bourses d'étudiants, mais passionnants et fort utiles pour tous les enseignants et chercheurs concernés par le développement des approches numérisées des discours. Dans le domaine littéraire, exploré et quantifié depuis de longues années par E. Brunet, Valérie Beaudouin présente un outil original de phonétisation et de versification, le Métromètre, qu'elle applique à l'analyse de la tragédie classique¹. Marc Hug présente des protocoles d'analyse permettant de décrire et de quantifier des usages linguistiques à partir de la base Frantext², tandis que Jean-Marie Viprey fournit un moteur de recherche pour l'analyse textuelle et hypertextuelle des poèmes baudelairiens³. Dans le domaine de la lexicologie politique quantitative, deux ouvrages retiennent particulièrement l'attention.

Le livre de Dominique Labbé et Denis Monière constitue une illustration claire et précise de la démarche lexicométrique devenue maintenant quasiment standard. Les auteurs se proposent d'analyser et de comparer, à l'aide des outils statistiques probabilistes, trois types de discours gouvernementaux fortement institutionnalisés sur une durée de 55 ans : 47 discours du Trône au Canada (soit 147 000 mots-occurrences et 5 400 formes-vocables); 50 discours inauguraux des ministères québécois (soit 204 000 occurrences et 6 800 formes); 48 déclarations gouvernementales des ministères français de la Quatrième et de la Cinquième République (soit 258 000 occurrences et 7 400 formes). Les trois corpus⁴, du même ordre de grandeur, saisis selon la

1. V. Beaudouin, 2002, *Mètre et rythme du vers classique. Corneille et Racine*, Paris, Champion. On comparera avec intérêt l'analyse rigoureuse des données pratiquée ici et les extrapolations un peu fantaisistes de D. Labbé, récemment diffusées comme un scoop par les grands médias, et qui visait à attribuer à Corneille l'ensemble de l'œuvre de Molière sur la seule foi des résultats d'un programme quantitatif. En ces temps de confusion entre application scrupuleuse des méthodes quantitatives et interprétations hasardeuses, livrées aux appétits médiatiques, on apprendra avec soulagement que les analyses métrométriques ne conduisent pas à de telles conclusions. On regrettera seulement que l'outil lui-même ne soit pas encore disponible pour d'autres tâches que l'alexandrin classique. On pourrait penser à l'analyse des slogans politiques.
2. M. Hug, 2002, *L'enquête d'usage linguistique*, Paris, Champion.
3. J.-M. Viprey, 2002, *Analyses textuelles et hypertextuelles des Fleurs du mal*, Paris, Champion.
4. Les discours du Trône canadien sont étudiés dans leur version française officielle ; les problématiques multilingues, d'alignements ou de corpus reliés ne sont pas abordées dans le présent ouvrage.

norme graphique de Saint-Cloud, puis catégorisés morphologiquement et lemmatisés selon la norme du CERAT de Grenoble, sont d'abord étudiés séparément, avec des outils lexicométriques identiques, élaborés ou développés par D. Labbé. Les programmes d'accroissement lexical, de distance intertextuelle, de classification automatique hiérarchique, de voisinage sémantique des vocables les plus fréquents, de spécificités lexicales des parties et de phrases représentatives sont ainsi appliqués systématiquement. Les panoramas lexicaux des trois corpus et de leurs parties sont dressés successivement. Ceux-ci sont ensuite réunis en un corpus étiqueté important de 600 000 formes, et comparés à l'aide de programmes mesurant notamment la diversité, la spécialisation et la richesse du vocabulaire.

Sans entrer dans le détail des résultats⁵ et de leur interprétation politique, qui reste limitée, on notera d'abord que les discours du Trône canadiens manifestent une continuité et une « chronologicité » lexicale (c'est-à-dire une évolution régulière faible dans le temps) remarquables, à peine marquées par les personnalités politiques et les choix partisans. Les discours inauguraux des ministères québécois soumis aux mêmes instruments révèlent des contrastes idéologiques plus accusés et qui l'emportent sur la conjoncture événementielle. Entre le Parti québécois et le Parti libéral, « le passage du temps ne semble pas affecter le choix des mots utilisés pour exprimer les politiques de ces deux partis, ce qui traduit la stabilité de leur idéologie respective qui se distingue par des différences de styles, de valeurs et d'orientations » (p. 55). Les mêmes programmes permettent de préciser sur cette base les oppositions proprement stylistiques entre dirigeants représentatifs des divers courants – Maurice Duplessis, Robert Bourassa, René Lévesque –, ce dernier présentant un fort taux de syntagmes verbaux, une forte adverbialité, un taux élevé de pronoms, de la première personne du pluriel notamment, caractéristique d'une parole empathique.

L'application des procédures lexicométriques standards aux discours gouvernementaux français de 1945 à 1997 apporte une masse plus importante de résultats dus aux changements de régimes et aux alternances politiques multiples et complexes. Il met en jeu aussi plusieurs expérimentations qui relativisent la fiabilité des outils utilisés, ou précisent les résultats.

L'analyse des 48 déclarations de politique générale, soumises ou non au vote du Parlement, permet de contraster la Quatrième et la Cinquième République sous l'angle évènementiel (de la décolonisation à l'accession de la gauche au pouvoir) et institutionnel (du régime parlementaire au régime présidentiel), mais il établit surtout la continuité lexicale du discours gouvernemental. Les phases d'accroissement lexical les plus marquées ne sont pas toujours les plus attendues : gouvernements Mollet, Pflimlin, Messmer, Rocard, Cresson, Bérégovoy, Jospin. Les proximités lexicales mettent bien sûr en évidence les paramètres

5. *Mots. Les langages du politique*, n° 62, mars 2000, *Le Programme de gouvernement, un genre discursif*, avait présenté plusieurs études sur ce domaine.

de l'identité personnelle, de l'appartenance partisane, de la proximité historique, mais ne permettent pas de savoir lequel de ces paramètres l'emporte sur les autres. La classification hiérarchique et les programmes de spécificités permettent néanmoins de dégager quatre grands types de déclarations gouvernementales: les demandes de confiance, constituées essentiellement par des déclarations de gauche dans les deux Républiques ainsi que par Jacques Chirac durant la cohabitation; les discours programmatiques ou fondateurs, notamment à l'orée des nouveaux régimes ou des changement de majorité, incarnés aussi par l'ensemble très homogène des discours pompidoliens; les déclarations à forte teneur personnelle tant de la Quatrième – René Pleven, Edgard Faure, René Mayer – que de la Cinquième République – Édith Cresson, Edouard Balladur, Alain Juppé, Lionel Jospin; enfin, les déclarations nettement parlementaristes, caractéristiques de la Quatrième République. On est ainsi conduit à des constats qui peuvent surprendre: il y a finalement peu de contrastes idéologiques et peu de différences temporelles dans la déclaration gouvernementale française, ce que les auteurs expliquent un peu rapidement à partir de l'homogénéité sociale et culturelle des groupes dirigeants, considérés même comme « une sorte de "locuteur collectif" » (p. 129). Interprétation purement sociologique qu'il faudrait sans doute affiner à l'aide d'une observation plus fine de la phraséologie, de la terminologie institutionnelle et des modes d'argumentation.

Le dernier chapitre regroupe les trois corpus et contraste globalement les vocabulaires utilisés dans les trois États analysés. Les longues listes de formes spécifiques à chacun, classées par catégories grammaticales, nécessiteraient évidemment plus qu'un bref commentaire impressionniste du type: « Les Français privilégient l'économie, la chronologie et l'histoire [...] le groupe verbal et spécialement les pronoms personnels [...] des premières personnes du singulier et du pluriel [...] les pseudo auxiliaires modaux ». Le commentaire, trop expéditif, sans retour au contexte, montre ici ses limites sur le plan linguistique, comme sur les hypothèses communicationnelles et sociopolitiques, la définition du vocabulaire spécialisé et la confusion entre discours et vocabulaire. Il n'empêche que le livre possède dans son ensemble une clarté didactique digne d'un manuel, qui, même si quelques coquilles brouillent parfois les résultats (p. 29 et ailleurs), illustre une démarche lexicométrique remarquablement maîtrisée et approfondie.

L'ouvrage de Damon Mayaffre, lui aussi clair et lisible, repose sur des principes analogues, mis en œuvre en l'occurrence à travers l'association de l'éti-queteur Cordial⁶ et du logiciel lexicométrique Hyperbase développé à l'université de Nice par Étienne Brunet. Ses vertus didactiques et épistémologiques sont toutefois moins évidentes (absence d'index ; caractéristiques de l'éti-queteur

6. Produit par la société Synapse (Toulouse), ce logiciel commercialisé, relativement robuste mais non dépourvu de faiblesses, assigne à tous les mots du corpus leur catégorie grammaticale, ce qui permet de les trier, de les dénombrer selon des critères morphosyntaxiques.

morphologique ni présentées, ni mises en question malgré ses limitations et ses imprécisions). Le programme Hyperbase, quant à lui, est utilisé comme fournisseur systématique de résultats, plus que comme outil d'expériences et de comparaison avec d'autres méthodes. Ses ambitions sont moins descriptives et analytiques qu'historiques et politiques. Ses hypothèses interprétatives sont fortes et intéressantes sur le plan discursif, et par conséquent plus discutables. Le travail se présente lui-même d'emblée comme une étape nouvelle et originale dans les analyses statistiques. L'auteur propose même, pour la désigner, le terme néologique de *logométrie*, littéralement: « analyse quantitative du discours, de la parole », qui nous paraît ambigu, voire contestable à plus d'un titre, comme nous le verrons en conclusion.

Pour analyser le discours de Jacques Chirac à l'aune de ses prédécesseurs, en suivant cette démarche logométrique, D. Mayaffre a réuni un vaste corpus⁷, remarquable en soi, de « tous les discours grand public » (p. 22) des présidents de la Cinquième République, un peu moins de deux millions d'occurrences pour près de 600 discours de différentes natures et provenant de supports très divers: messages au Parlement et autres, allocutions de circonstances, déclarations officielles, entretiens écrits ou oraux, télévisés ou donnés à des magazines, conférences de presse. Notons en passant que les notions de « grands discours » et de « genre » mériteraient des définitions. En réalité, il ne s'agit pas, pour D. Mayaffre, d'étudier le discours à travers un ou plusieurs genres, mais la « parole des présidents » dans sa globalité et son évolution diachronique, le volume et la diversité considérables des données étant censés unifier leur hétérogénéité constitutive. Dépourvue de réflexion sur les paramètres de la communication politique, des stratégies argumentatives, mais aussi des conditions de transcription, voire d'établissement des textes, cette globalité de l'objet ne correspond ni au concept philosophique classique du *logos*, inscrit dans la dialectique rationaliste, ni à sa déclinaison contemporaine dans les nouvelles rhétoriques, dans ses rapports avec la *doxa*, ni à son sens restreint de parole orale.

Divisé en six parties selon les mandats présidentiels successifs de la Cinquième République, le corpus lemmatisé et catégorisé est soumis à une démarche apparentée à celle de D. Labbé et D. Monière, moins développée toutefois sur le plan statistique. C'est essentiellement la distance lexicale, puis les spécificités et les classifications qui sont employées pour contraster les discours des cinq présidents dans la première partie du livre, et, dans les derniers chapitres, les différentes périodes chiraquiennes.

L'analyse s'articule en trois phases. La première s'attache à montrer l'évolution lexico-grammaticale des discours des cinq présidents de la République. D. Mayaffre examine les index hiérarchiques lemmatisés des cinq locuteurs.

7. Les données de D. Mayaffre, en accès libre sur le site POLITEXT, témoignent d'une remarquable volonté, encore insuffisamment partagée dans la communauté, de mettre non seulement les résultats mais aussi les données à disposition.

En l'espèce, il ne retient que les dix premiers lemmes et commente ces « mots pleins » en concluant, un peu rapidement selon nous, à un lexique peu original. Même sur une liste aussi brève, on pourrait noter quelques contrastes remarquables, par exemple l'absence paradoxale d'un lexique européen chez Valéry Giscard d'Estaing. L'index complet portant sur un vaste corpus, dispersé sur le plan thématique et typologique, n'est pas analysé dans sa diversité. L'évolution du discours présidentiel est ensuite étudiée sous l'angle grammatical. Une « grammaire du discours », terme peu heureux au regard de ce qu'on appelle traditionnellement « grammaire de discours »⁸ en linguistique de corpus et en analyse de discours, propose une quantification des catégories grammaticales. Elle permet d'établir, par exemple, que le verbe s'est substitué dans les derniers septennats au substantif qui domine le discours des premiers présidents – Charles de Gaulle, Georges Pompidou, V. Giscard d'Estaing. Le discours présidentiel aurait ainsi perdu de sa « tenue littéraire » (p. 43) au profit d'improvisations orales et le dire importerait désormais plus que le dit. Cette évolution marquerait un déclin de la parole politique, la conceptualisation gaullienne ou marxiste ayant laissé progressivement la place au dire communicationnel. Il faudrait ici s'interroger sur la dégénérescence du discours politique⁹, mais aussi sur la transformation des supports et des pratiques politico-médiatiques.

Le vocabulaire spécifique des locuteurs est ensuite analysé dans des termes essentiellement psychologiques. Ch. de Gaulle y apparaît comme le guide, le chef, au travers d'emplois tels que *peuple, État, Algérie*. G. Pompidou « l'humaniste » est caractérisé par des suremplois tels que *civilisation, bonheur, individu*. V. Giscard d'Estaing, « pédagogue face à la crise », accuse un excédent pour les formes *emploi, énergie, pétrole, situation, problème* mais aussi à *l'heure actuelle* et *indiquer*. François Mitterrand est caractérisé par une forte implication dans le discours et un suremploi des marques de la première personne du singulier, mais aussi de la forme *Europe*.

La parole chiraquienne forme l'objet de la deuxième section, « Structure grammaticale du discours chiraquien », puis de la troisième, « Chirac et les aléas du temps », où son évolution est approfondie. La centralité de la parole chiraquienne, c'est-à-dire le fait que son vocabulaire et les catégories grammaticales spécifiques de son discours apparaissent comme peu originaux, banals, intermédiaires entre les autres, y est établie à l'aide des outils standards de classification, dendrogrammes, analyses factorielles. Dans les derniers chapitres, une nouvelle application

8. Développée en Allemagne depuis les années soixante, la *Textgrammatik* a été approfondie en France par J.-M. Adam notamment, mais aussi par les sémioticiens du texte. Il s'agit, dans cette perspective, de dégager des modèles structuraux des textes, lesquels nécessitent d'autres étiquetages que les catégories grammaticales, qui n'en sont qu'une approche préalable selon nous. Voir M.-A. Paveau & G.-E. Sarfati, 2003, *Les grandes théories de la linguistique*, Paris, Armand Colin, chapitre 10.
9. Sur la notion de déclin, voir S. Bonnafous, 2003, « La dégénérescence du discours politique: un lieu commun de l'Antiquité à nos jours? », *Argumentation et discours politique*, Rennes, Presses universitaires de Rennes, p. 249-257.

des programmes établit, en corollaire, que le discours de Lionel Jospin, dans les années 2000-2002, aurait subi lui aussi une centralisation qui pourrait expliquer son échec à l'élection présidentielle de 2002.

Somme toute, pour reprendre les termes de D. Mayaffre, l'étude systématique des discours présidentiels permettrait de montrer, en trois temps, que Ch. de Gaulle, G. Pompidou et V. Giscard d'Estaing appuient leur éloquence sur le *logos*, F. Mitterrand sur le *pathos*, J. Chirac sur l'*ethos*. En d'autres termes, la *pensée* politique s'efface au profit de la *communication* politique dans laquelle le prédicat importe plus que le thème, le dire que le dit, l'illocutoire que le locutoire, le symbole que le sens, la personnalité du locuteur que les arguments avancés, la forme phatique que le fond idéologique. Cette thèse générale, séduisante, est développée dès l'introduction de l'ouvrage: J. Chirac, « président de la parole » (p. 5), faute de pouvoir agir, multiplie les interventions; il place dans le verbe ce qu'il ne peut faire dans la réalité. Il est « condamné à gouverner par les mots » (p. 13). J. Chirac serait l'héritier d'une lignée de chefs d'État dont il ne peut ni ne veut se démarquer dans le discours. « Déférente filiation à Charles de Gaulle » (p. 20), déférence vis-à-vis des quatre « grands » présidents à qui il succède, complexe d'infériorité envers F. Mitterrand, telles sont les thèses sociopolitiques. D. Mayaffre explique ainsi la relative centralité – statistique, linguistique et discursive – du discours chiraquien, le caractère peu discriminé de son lexique, l'importance de la verbalité et de l'adverbialité chez lui (*naturellement* étant son mot fétiche et passe-partout), ce qu'il appuie au moyen de quantifications et d'analyses multidimensionnelles.

Pourtant, ces traits distinctifs une fois établis sur la base de la morphologie et de la syntaxe et après retour aux contextes, il apparaît aussi que la « centralité » est principalement le fait d'une série d'indices par lesquels le président pourrait chercher à insuffler à son discours une dimension mobilisatrice, un volontarisme, paradoxalement inédits jusqu'ici dans le discours présidentiel. Cela fait-il de J. Chirac le *président de la parole*? Tout est question d'interprétation. Il serait évidemment nécessaire d'analyser son évolution avant et après la cohabitation, mais également au lendemain de la dernière élection présidentielle, dont la configuration fut inédite. Il existe des précédents dans ce schéma politique: Mitterrand/Chirac en 1986 et Mitterrand/Balladur en 1993. Observe-t-on chez F. Mitterrand une variation dans les emplois des formes verbales? Par ailleurs, peut-on expliquer les caractéristiques chiraquiennes par une dégénérescence du politique? Là encore, tout est question d'interprétation et de définition du politique. Plutôt que d'évoquer une dégénérescence du politique, on pourrait parler d'évolution des pratiques discursives en politique. Encore faudrait-il pour cela analyser un corpus politique qui ne soit pas uniquement présidentiel pour l'affirmer avec certitude. Comment, en outre, porter un jugement comparatif sur des positionnements énonciatifs si ceux-ci ne s'inscrivent pas dans les mêmes conditions d'énonciation? Comment juger de la personnalisation du discours

et d'autres évolutions chronologiques si, pour chaque locuteur, on ne dispose pas du même échantillonnage? Une certaine conception de l'analyse des données textuelles tend à montrer que le grand nombre compense – dans une certaine mesure – ces déséquilibres. Encore faut-il alors revenir au texte, prendre en compte les différentes situations de communication, monter des contre-expériences avant toute phase interprétative. On peut également choisir l'exhaustivité – souvent illusoire – et considérer qu'on analyse alors l'ensemble du discours présidentiel, l'ensemble du discours gouvernemental. Mais même dans ce cas, les analyses contrastives devraient tenir compte de chaque situation d'énonciation et comparer des sous-corpus avant d'entreprendre des interprétations sur l'ensemble. Nous ne critiquons nullement l'usage des corpus de grande ampleur et la loi du grand nombre qui constituent une formidable opportunité pour le chercheur, mais défendons une cohérence méthodologique et un retour au texte plus stricts. Cela permettrait sans doute de montrer l'importance grandissante du cadrage de la parole, de la parole mobilisatrice, voire incantatoire, des formules, au détriment du discours conceptuel, des bilans, des synthèses, des clarifications didactiques.

Revenons pour finir sur la notion même de logométrie en soulevant deux questions. L'ouvrage de D. Mayaffre, pour intéressantes et suggestives que soient ses analyses et ses interprétations, ne constitue pas selon nous une étape radicalement distincte de la lexicométrie ou de son dérivé immédiat, la textométrie¹⁰. D'abord historien, D. Mayaffre est sans doute plus attentif aux tendances statistiques lourdes, aux évolutions solidement établies, interprétables immédiatement dans des catégories historiques, politiques ou psychosociologiques qu'aux phénomènes discursifs, aux définitions grammaticales, aux variations contextuelles, aux fonctionnements textuels, aux stratégies communicationnelles et argumentatives. Sa démarche exploite un étiqueteur qui nécessiterait une intervention manuelle importante pour réduire sa part d'indétermination, sous peine d'approximations nombreuses qui demeurent dans ses analyses, et un dispositif lexicométrique certes excellent et très robuste, mais qui ne fournit pas d'outils descriptifs radicalement nouveaux. Elle ne constitue pas une nouvelle étape inédite, logométrique, dans ce que nous appelons pour notre part le domaine du TAM (traitement automatisé des discours).

En second lieu, la logométrie telle que D. Mayaffre la décrit programmiquement ne nous paraît pas être l'étape la plus cruciale d'une nouvelle lexicométrie ou textométrie. Fruit, selon lui, d'un « glissement de la lexicométrie originelle vers une logométrie pleine et entière, susceptible de renouveler la discipline » grâce aux progrès quantitatifs des lemmatiseurs et des catégoriseurs, la démarche logométrique serait « un ensemble de traitements documentaires et statistiques du texte qui ne s'interdit rien pour tout s'autoriser; qui dépasse

10. Terme qui souligne d'abord le cadre interprétatif textuel, linéaire, plus étendu, dans lequel les propriétés quantitatives prennent leur signification.

le traitement des formes graphiques sans les exclure ou les oublier; qui analyse les lemmes ou les structures grammaticales sans délaisser le texte natif auquel nous sommes toujours renvoyés. C'est finalement un traitement automatique global du texte dans toutes ses dimensions: graphiques, lemmatisées, grammaticalisées. L'analyse ainsi portera sur toutes les unités linguistiques, de la lettre aux isotopies, en passant par les n-grams¹¹, les mots, les lemmes, les codes grammaticaux, les bi-codes ou les enchaînements syntaxiques ».

Nous sommes loin encore d'un traitement transversal global des textes, articulé, fondé sur des expérimentations solides qui donneraient lieu à des interprétations sémantiques intégrées. Les catégorisations morphosyntaxiques, sans parler même des sémantiques, ne peuvent encore être mises en œuvre sans une minutieuse vérification manuelle des données, et les résultats quantitatifs ne doivent pas être considérés comme fiables sans de solides contre-expériences ni un retour permanent aux outils contextuels que sont d'abord les concordanceurs, mais aussi les liens hypertextuels. La dénomination de logométrie nous semble par conséquent trop globalisante, laissant entendre que l'on appréhende le discours sous tous les angles simultanément, que l'on mène une « mesure du discours », alors que ces dimensions, bien que complémentaires, ne peuvent être mises sur le même plan. La conception logométrique, qui ne contraindrait plus le chercheur à un type d'analyse particulier mais lui autoriserait catégorisations, lemmatisations et travail en surface, sans protocole établi, est une approche possible, mais qui risque d'être approximative. La tâche lexicométrique prioritaire reste pour nous, d'une part, le développement de questions susceptibles d'être quantifiées (toutes ne le sont pas) et d'hypothèses idoines, l'accumulation de données et de résultats comparables, d'autre part la comparaison d'outils divers, d'expériences variées et de résultats partiels, la mise en visibilité topographique de mesures successives procédant par rapprochements progressifs jusqu'à saisir les phénomènes les plus fins, la transition de l'analyse statique des segments répétés vers les analyses dynamiques des constructions terminologiques d'un côté, des circulations phraséologiques de l'autre. C'est à ces conditions, entre autres, qu'il sera possible peut-être de passer de la mesure des mots et des textes à l'appréhension d'un objet, le discours, qui, tout autant que la langue, est le résultat d'une construction de points de vue multiples.

Pierre Fiala

Université Paris 12, Céditec
fiala@univ-paris12.fr

Jean-Marc Leblanc

Université Paris 12, Céditec
Jean-Marc.LeBlanc@nrc-cnrc.gc.ca

11. Dénomination anglo-saxonne, pour le programme de recherche des segments répétés développés initialement par A. Salem, 1987, *Pratique des segments répétés: essai de statistique textuelle*, Paris, Klincksieck.