



HAL
open science

Les déclarations gouvernementales sous la Ve République (1959-1997)

Dominique Labbé

► **To cite this version:**

Dominique Labbé. Les déclarations gouvernementales sous la Ve République (1959-1997). Jean-Louis Autin et Laurence Weill. Le Droit figure du politique. Etudes offertes au professeur Michel Miaille., Université de Montpellier I, tome I, p. 843-865, 2008. halshs-00365678

HAL Id: halshs-00365678

<https://shs.hal.science/halshs-00365678>

Submitted on 12 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les déclarations gouvernementales sous la Ve République (1959-1997)

Conférence prononcée devant le troisième cycle de science politique de la
Faculté de Droit de Montpellier (3 décembre 1998)

Dominique Labbé
Institut d'Etudes Politiques de Grenoble
(dominique.labbe@iep.grenoble.fr)

Résumé :

Présentation des mesures de distance intertextuelle, des classifications automatiques et arborées : méthodes, intérêt et fiabilité des mesures. Application aux discours inauguraux des Premiers ministres de la Ve République. Ces textes ne se partagent pas selon la coupure entre droite et gauche mais selon la conjoncture politique du moment et selon la solidité plus ou moins grande des majorités parlementaires.

Manuscrit du texte publié dans : AUTIN Jean-Louis et WEILL Laurence (Eds).
Le Droit figure du politique. Etudes offertes au professeur Michel Miaille.
Montpellier : Université de Montpellier I, 2008, tome I, p. 843-865.

Toute citation doit se faire à partir de l'ouvrage.

Avant-propos

Le langage – écrit ou oral – est le principal véhicule de la communication et mais aussi le vecteur privilégié des prénotions et des idéologies. C'est la raison pour laquelle Michel Miaille s'est toujours intéressé à la critique du langage. En 1975, il a bien voulu participer au jury de notre thèse d'Etat qui portait sur ce thème¹. Depuis, il a suivi avec attention nos travaux sur le langage, nous invitant régulièrement à intervenir dans le troisième cycle de science politique de la Faculté de Droit de Montpellier, troisième cycle qu'il a longtemps animé. Ainsi, le 3 décembre 1998, nous y avons présenté des "travaux en cours" sur les déclarations gouvernementales devant le parlement sous la Ve république.

A cette occasion, nous avons exposé, pour la première fois en public, nos recherches sur la mesure des proximités et des distances entre textes. Ces méthodes faisaient l'objet d'une recherche collective associant notamment : Jean-Guy Bergeron (Université de Montréal), Pierre Hubert (Ecole des Mines de Paris), Cyril Labbé (Université de Grenoble II), Xuan Luong (Université de Nice), Denis Monière (Université de Montréal) qui nous ont autorisé à présenter les résultats de nos recherches collectives.

Durant le processus de mise au point d'un procédé nouveau, l'exposé en séminaire de recherche est la première étape avant les colloques et les congrès puis la publication définitive dans une revue à comité de lecture. Cette première étape permet de tester l'intérêt de la démarche, la clarté de l'exposé, de répondre aux éventuelles objections...

Après cette conférence de Montpellier, la méthode a été présentée devant deux autres séminaires² puis, en 2000, devant deux congrès internationaux³ et, enfin, publiée dans trois revues scientifiques⁴. Depuis lors, elle a été largement débattue et elle est maintenant utilisée par de nombreux chercheurs à travers le monde.

Nous reproduisons ci-dessous le texte de cette conférence, texte qui était resté inédit de même que les données concernant les déclarations gouvernementales sous la cinquième république⁵.

Nous avons gardé à ce texte la dimension technique voulue par M. Miaille qui, selon ses propres termes, souhaitait faire pénétrer l'auditoire dans "l'atelier du politiste".

(Grenoble, août 2007)

¹ Dominique Labbé. Essai d'approche de l'idéologie au travers du discours politique. Le cas du PCF de 1962 à 1968. Thèse d'Etat pour le doctorat en sciences politiques. Paris : Fondation des sciences politiques, 1975. Publiée sous le titre Le discours communiste aux Presses de la Fondation des sciences politiques en 1977.

² Les données concernant les déclarations gouvernementales ont été également présentées, le 12 février 1999 à l'Ecole normale supérieure de Fontenay-Saint Cloud, devant un séminaire animé par Benoît Habert.

³ Labbé Dominique et Monière Denis (2000), "La connexion intertextuelle. Application au discours gouvernemental québécois", Martin Rajman et Jean-Cédric Chapellier (eds), Actes des 5^e journées internationales d'analyse des données textuelles, Lausanne, Ecole polytechnique fédérale, vol 1, p 85-94 ; Bergeron Jean-Guy et Labbé Dominique (2000). "L'évaluation de la négociation raisonnée par les acteurs. Une analyse lexicométrique". (Communication au XVI^e Congrès international de l'Association internationale des sociologues de langue française, Québec, juillet 2000). Reproduit dans Bernier Colette et Al. Formation, relations professionnelles à l'heure de la société-monde. Paris-Québec : L'Harmattan - Les Presses de l'Université Laval, 2002, p. 239-252. Ces textes peuvent être consultés en ligne sur les archives ouvertes du CNRS.

⁴ Labbé Cyril et Labbé Dominique. "Inter-textual Distance and Authorship Attribution". Journal of Quantitative Linguistic. December 2001, Vol. 8, n°3, pp. 213-231. Labbé Cyril et Labbé Dominique. "La distance intertextuelle". Corpus, 2003-2, p 95-118. Labbé Cyril & Labbé Dominique. "A Tool for Literary Studies: Intertextual Distance and Tree Classification". Literary and Linguistic Computing. 21-3, 2006, p 311-326. Textes consultables en ligne sur les archives ouvertes du CNRS.

⁵ Ces données sont extraites d'un corpus plus vaste constitué avec Denis Monière de l'Université de Montréal dans le but de comparer les discours gouvernemental au Canada, en France et au Québec. Voir Labbé Dominique et Monière Denis. Le discours gouvernemental. Paris : Champion, 2003.

Comment mesurer les proximités ou les dissemblances entre deux textes ? Une réponse précise à cette question est du plus haut intérêt pour les politologues et, plus généralement, pour les sciences sociales. Par exemple, en ce qui concerne les gouvernements de la Ve République, peut-on repérer, par le calcul, quelques grandes familles unies par des discours proches – sinon par des matrices idéologiques communes – et opposées les unes aux autres ?

Nous allons esquisser une réponse en utilisant les 25 déclarations gouvernementales – "les déclarations de politique générale" – faites par les Premiers ministres devant le parlement (voir en annexe, la liste de celles-ci) depuis les débuts de la Ve République.

Deux méthodes sont combinées.

En premier lieu, les "distances" qui séparent chaque texte de tous les autres sont mesurées à l'aide d'un calcul qui va être expliqué en détail. En second lieu, on tentera de répondre à la question de savoir s'il existe quelques groupes plus ou moins homogènes au sein de cette population de 25 textes.

I. Distances et proximités entre textes

Il s'agit de mesurer la distance entre deux textes comme on le fait entre deux objets quelconques. L'unité de mesure est ici le mot. La distance entre deux textes – que nous proposons de nommer : "distance intertextuelle" – est le nombre de mots différents que contiennent ces textes. Plus ce nombre est élevé, plus la distance est grande.

1.1 la distance intertextuelle

La "distance euclidienne" entre deux points A et B est la longueur du segment de droite unissant ces deux points. Cette distance présente les propriétés suivantes :

- identité : $d_{(a,a)} = 0$ (la distance d'un point à lui-même est nulle ; lorsque $d_{(a,b)} = 0$, les points A et B sont confondus) ;
- symétrie : $d_{(a,b)} = d_{(b,a)}$ (le résultat est le même que la mesure soit effectuée à partir de A vers B ou de B vers A) ;
- inégalité triangulaire : $d_{(a,b)} \leq d_{(a,c)} + d_{(c,b)}$ (le chemin direct entre deux points est toujours plus court qu'en passant par un troisième point, l'égalité n'est possible que si C appartient au segment AB).

Si la mesure adoptée ne présente pas toutes ces propriétés, il s'agit d'une mesure de "similarité" et non d'une véritable "distance". On remarque au passage que les distances dites "à vol d'oiseau", qui sont toujours données comme exemples de "distances euclidiennes" ne présentent ces propriétés que sur de petites échelles. Au-delà de quelques dizaines de kilomètres, du fait de la courbure de la terre, ces distances sont des arcs de cercle et non plus des droites. Elles ne peuvent donc pas être qualifiées de "distances euclidiennes".

L'indice qui servira à mesurer la distance entre deux textes A et B doit donc présenter ces 3 propriétés.

Pour bien comprendre le calcul présenté ci-dessous, il faut se souvenir qu'un texte est la succession d'un certain nombre de **mots** – dont le nombre total donne la **longueur** (voir en annexe les caractéristiques des différentes déclarations gouvernementales) – ces mots étant issus d'un **vocabulaire** nécessairement plus restreint puisque certains **vocables** (ou "mots différents") sont employés plusieurs fois dans le texte. Par exemple, "le", "les", "la", "l'" – et leurs équivalents avec une majuscule initiale - sont les différentes **formes** sous lesquelles l'article ou le pronom "le" apparaissent dans un texte. "le, article" et "le, pronom" sont des

vocables (ou "entrées de dictionnaire"). Chacune des **occurrences** de ces deux vocables – sous les formes "le", "la", "les", "l'" – sont des mots du texte.

Ces caractéristiques sont notées ainsi :

- le vocabulaire des textes A et B est constitué de V_a et V_b vocables différents ;
- chacun de ces vocables est présent un certain nombre de fois dans les textes A et B. Soit F_{ia} et F_{ib} les **effectifs** d'un vocable i respectivement dans les textes A et B.

- N_a et N_b : nombre de mots dans A et B (longueurs) ; $N_a = \sum F_{ia}$ et $N_b = \sum F_{ib}$.

Considérons d'abord le cas de textes de longueurs égales ($N_a = N_b$), la **distance** absolue entre eux sera le **nombre de mots différents** contenus dans ces deux textes :

$$D_{(a,b)} = \sum_{i \in A, i \in B} |F_{ia} - F_{ib}|$$

Et la distance relative sera égale à :

$$(1) D_{(a,b)} = \frac{\sum_{i \in A, i \in B} |F_{ia} - F_{ib}|}{N_a + N_b}$$

Nous baptisons cette mesure "**distance intertextuelle**" pour indiquer qu'elle porte sur l'ensemble des mots des deux textes.

Le calcul en valeur absolue – les deux barres verticales au numérateur indiquent que l'on neutralise les signes - est indispensable pour respecter la propriété de symétrie.

La formule (1) présente toutes les propriétés d'une distance euclidienne dans le cas particulier de deux textes de longueurs égales (ce qui est assez rare : on voit dans l'annexe 1 que les 25 déclarations ont des longueurs différentes). Pour mesurer la distance entre textes de longueurs inégales, il est proposé de réduire le plus long (B) à la longueur du plus petit (A).

Soit B' cette réduction de B à la longueur de A. Tout vocable présent F_i fois dans B aura une fréquence attendue (espérance mathématique) dans B' (c'est-à-dire dans A) égale à :

$$E_{ia(u)} = F_{ib} * U_{(a,b)} \text{ avec } U_{(a,b)} = \frac{N_a}{N_b}$$

D'où l'on tire la taille du texte réduit B' :

$$N_{b'} = \sum_{i \in B'} E_{ia(u)} = N_a$$

Dans la formule (1), on remplace les termes F_{ib} par $E_{ia(u)}$ et N_b par $N_{b'}$. La distance intertextuelle devient :

$$(2) D_{(a,b)} = \frac{\sum_{i \in A, i \in B} |F_{ia} - E_{ia(u)}|}{N_a + N_{b'}}$$

Les résultats de la formule (2) varient uniformément entre 0 (même vocabulaire, même fréquence d'emploi pour chaque vocable) et 1 (aucun mot en commun). Ces résultats sont très faciles à interpréter : une distance de 0.5 signifie que les deux textes partagent la moitié des mots qui les composent respectivement, etc.

Toutefois, certaines limitations doivent être mentionnées.

Premièrement, la formule (2) ne neutralise pas totalement l'influence des différences de longueur entre les textes. D'une part, dans les textes très courts, quelques mots rares peuvent peser d'un poids exorbitant dans le calcul (en l'état actuel de nos expériences, il apparaît que le calcul ne doit pas être appliqué à des textes de moins de 1000 mots). D'autre part, le vocabulaire d'une langue est limité (le vocabulaire politique l'est encore davantage !) Par conséquent, plus les textes s'allongent, plus leurs vocabulaires ont tendance à converger. Cette tendance est assez lente et les tests montrent que l'on peut sans inconvénient comparer des textes compris dans une échelle de longueur inférieure à 1/10, à condition d'être prudent lorsqu'on approche cette limite et de considérer que les textes les plus longs peuvent être légèrement "avantagés" par rapport aux textes les plus courts. Etant donné les longueurs très proches des déclarations gouvernementales, ces limitations ne les concernent pas.

Deuxièmement, les effectifs observés sont des entiers alors que les espérances mathématiques sont des fractions, ce qui introduit des décimales et une certaine incertitude dans le calcul (du fait du poids prédominant des vocables de faible fréquence).

Au total, les résultats de (2) sont compris dans une marge d'incertitude maximale de $\pm 5\%$ lorsque le procédé est appliqué à des textes dont les différences de longueurs sont inférieures à 1/5. Les différences de longueurs entre les déclarations gouvernementales étant nettement inférieures à cette échelle, cette formule peut donc être appliquée sans problème.

Ces réserves admises, que mesure l'indice de la distance intertextuelle ?

La formule (2) a été mise en œuvre sur un très grand nombre de corpus différents et les résultats ont été étalonnés en suivant les techniques usuelles en science de la vie (notamment les expériences dites "en aveugle" : échantillons tirés au hasard et anonymes).

Ces expériences ont montré que l'indice de la distance intertextuelle enregistre l'influence de 4 facteurs : l'auteur, le thème, le genre et l'époque. Le dernier facteur se comprend aisément : la langue est un organisme vivant dont le composant sémantique (le "lexique") évolue constamment. Il est donc nécessaire de comparer des textes contemporains afin de neutraliser l'influence de ce facteur "chronologique". L'influence du "genre" est surtout clair quand on compare l'oral à l'écrit. Il se manifeste également entre textes appartenant à des registres assez fortement codifiés (textes scientifiques, juridiques, administratifs, journalisme...). Il est donc nécessaire de comparer des textes appartenant à un même genre afin de pouvoir mesurer l'influence des thèmes – les idées - et des auteurs, ce qui est notre but ici.

Enfin, ces expériences ont abouti à l'étalonnage d'une échelle des distances. Pour des textes de la longueur des déclarations gouvernementales (supérieure à 4.000 et inférieure à 15.000 mots), on observe que :

- les valeurs inférieures à 0,20 – moins d'un mot sur 5 est différent - ne se rencontrent que entre des textes contemporains écrits par un même auteur et portant sur un même thème ;

- les valeurs comprises entre 0,20 et 0,25 peuvent signaler un auteur unique dans un même genre - mais à des époques ou sur des thèmes différents – ou, plus rarement deux auteurs traitant, à la même époque, dans un même genre, un même thème. Dans ce dernier cas, il est également probable que l'un des deux auteurs se soit "inspiré" de l'autre ;

- les valeurs comprises entre 0,25 et 0,35 concernent parfois un même auteur dans un même genre mais, dans ce cas, il y a un changement important de thème. La plupart du temps, ces distances signalent des auteurs différents mais contemporains et traitant des sujets proches dans un même genre ;

- au-dessus de 0,35, des textes d'un même auteur appartiennent à des genres différents. Si le genre est le même, alors les auteurs sont différents.

Naturellement, il ne s'agit pas de "seuils critiques" mais de graduations sur un continuum qui doivent être interprétées comme telles.

I.2 Les premiers ministres de la Ve

La formule (2), appliquée aux 25 déclarations gouvernementales, aboutit à un tableau carré trop grand pour être reproduit (il comporte 25 lignes par 25 colonnes – soit 625 cases)¹. Du fait de la propriété d'identité, la première diagonale de ce tableau est nulle. Du fait de la propriété de symétrie, la partie supérieure à cette diagonale est égale à la partie inférieure. Il y a donc 300 distances différentes.

La moyenne de ces 300 distances est égale à 0,288 avec un écart-type de 0,025 (tableau I). L'égalité des trois valeurs centrales (moyenne, médiane et mode) suggère que les valeurs se distribuent autour de la moyenne selon la célèbre courbe de Gauss, ce que vérifie l'histogramme des fréquences². Autrement dit, la dispersion autour de la moyenne est très faible : les trois quart des valeurs sont comprises dans un intervalle de $\pm 8,8\%$ autour de 0,288 (0,263 – 0,313) et 95% des distances sont comprises dans un intervalle de $\pm 16,6\%$ autour de 0,288 (0,233 - 0,338). Le corpus des déclarations gouvernementales est très homogène. Ces textes appartiennent à un même genre et la plupart sont unis par de fortes proximités (chronologiques et thématiques).

Tableau I. Principales valeurs caractéristiques et comparaison avec C. de Gaulle et F. Mitterrand.

	Déclarations gouvernementales	De Gaulle	Mitterrand
Distance moyenne	0,288	0,278	0,265
Ecart-type	0,025	0,027	0,025
Variation relative (%)	8,8	9,7	9,5
Médiane	0,288	0,273	0,263

A titre de comparaison, le tableau I donne les valeurs observées sur deux autres corpus politiques contemporains à auteur unique mais avec une diversité de genre (de Gaulle et Mitterrand³). Ces valeurs sont à peine inférieures à celles observées sur les déclarations gouvernementales. Comme il y a un seul auteur d'un côté et 15 de l'autre, on peut déjà conclure à l'existence probable d'une sorte de "moule" – caractéristique de la déclaration gouvernementale - dans lequel se seraient coulés les Premiers ministres (sinon tous, du moins la quasi-totalité d'entre eux). Nous reviendrons sur ce problème en conclusion.

Considérons ensuite les distances les plus petites qui signalent des couples très proches (tableau II ci-dessous).

¹ Le tableau VI ci-dessous en offre un exemple sur une échelle réduite.

² Figure non reproduite.

³ De Gaulle : déclarations, conférences de presse et entretiens radio-télévisés entre juin 1958 et mars 1969 (79 textes, 201 927 mots) ; Mitterrand : déclarations, conférences de presse et entretiens radio-télévisés entre juin 1981 et mars 1988 (68 textes, 305 215 mots). Pour une présentation de ces corpus : Labbé Dominique. Le vocabulaire de F. Mitterrand. Paris : Presses de la Fondation nationale des sciences politiques, 1990.

Tableau II. Les 30 couples les plus proches

N°	Textes A	Textes B	Distances
1	Mauroy (1981)	Chirac (1986)	0.226
2	Juppé (1995)	Jospin (1997)	0.228
3	Messmer (1972)	Messmer (1973)	0.229
4	Balladur (1993)	Juppé (1995)	0.230
5	Chirac (1986)	Juppé (1995)	0.232
6	Balladur (1993)	Jospin (1997)	0.237
7	Mauroy (1981)	Juppé (1995)	0.239
8	Chirac (1986)	Balladur (1993)	0.241
9	Mauroy (1981)	Balladur (1993)	0.243
10	Chaban (1969)	Barre (1977)	0.243
11	Chaban (1969)	Chirac (1974)	0.244
12	Messmer (1973)	Mauroy (1981)	0.245
13	Pompidou (1966)	Pompidou (1967)	0.246
14	Chirac (1974)	Chirac (1986)	0.246
15	Barre (1978)	Mauroy (1981)	0.246
16	Cresson (1991)	Balladur (1993)	0.247
17	Chirac (1986)	Cresson (1991)	0.247
18	Mauroy (1981)	Jospin (1997)	0.248
19	Chirac (1974)	Mauroy (1981)	0.250
20	Pompidou (1967)	Chaban (1969)	0.250
21	Barre (1988)	Balladur (1993)	0.252
22	Cresson (1991)	Juppé (1995)	0.253
23	Barre (1977)	Mauroy (1981)	0.253
24	Barre (1977)	Chirac (1986)	0.254
25	Fabius (1984)	Barre (1988)	0.254
26	Barre (1977)	Balladur (1993)	0.254
27	Messmer (1972)	Mauroy (1981)	0.255
28	Chirac (1986)	Jospin (1997)	0.255
29	Chirac (1974)	Barre (1977)	0.255
30	Barre (1978)	Juppé (1995)	0.256

La plus petite distance (0.226 que l'on arrondit à 0.23 pour tenir compte de l'incertitude signalée ci-dessus) sépare le discours de P. Mauroy (1981) et celui de J. Chirac en 1986. Elle est suivie par :

- le couple formé par A. Juppé (1995) et Lionel Jospin (0.228 que l'on arrondit aussi à 0.23) ;
 - les deux déclarations de P. Mesmer (1972 et 1973) : 0.229 arrondie à 0.23 ;
 - les discours de E. Balladur (1993) et d'A. Juppé (1995) : 0.23 ;
 - les discours de J. Chirac (1986) et d'A. Juppé (1995) : 0.23 ;
- Etc.

Certaines de ces proximités remarquables sont logiques puisque l' "auteur" est théoriquement le même. C'est particulièrement le cas de Messmer 1972 et 1973 (puisqu'il est de notoriété publique que P. Messmer apportait un soin particulier à l'élaboration de ses discours) ; Pompidou 1966 et 1967 ; Chirac 1974 et 1986...

Mais G. Pompidou et J. Chirac ont prononcé 4 déclarations gouvernementales, soit 12 couples possibles dont seulement deux apparaissent dans ce tableau. De même, P. Mauroy en a prononcé 3, soit 3 couples différents dont aucun n'apparaît dans ce tableau. D'après le modèle exposé ci-dessus, deux explications sont possibles (et peuvent être cumulées) : ces Premiers ministres n'ont pas fait appel aux mêmes collaborateurs pour écrire certains de leurs discours ; ils n'ont pas traité les mêmes thèmes (ce qui signifierait un tournant politique).

D'autres couples semblent également logiques puisque les locuteurs appartiennent au même camp politique et se succèdent d'assez près dans le temps : Balladur (1993) - Juppé (1995) ; Chirac (1986) - Juppé (1995) ; Chirac (1986) - Balladur (1993) ou parce qu'on peut supposer qu'ils entretiennent certaines affinités au-delà de leurs divergences tactiques : Chaban (1969) et Barre (1977) ; Chaban (1969) et Chirac (1974), etc.

En revanche, près de la moitié des couples les plus proches sont inattendus et "marient" un Premier ministre de droite à un autre de gauche. C'est notamment le cas des deux premiers couples qui concernent les deux alternances avec "cohabitation" (1981-1986) et (1995-1997). D'après le modèle ci-dessus, deux explications sont possibles (et cumulables) : les mêmes collaborateurs ont écrit les textes en question ; ces textes traitent des mêmes questions dans des termes assez proches.

Avant de discuter ces deux hypothèses, voyons si les couples les plus éloignés confirment ces tendances (tableau III).

Tableau III. Les 40 couples les plus éloignés

N°	Textes A	Textes B	Distances
270	Pompidou (1966)	Bérégoovoy (1992)	0.322
271	Bérégoovoy (1992)	Jospin (1997)	0.322
272	Messmer (1973)	Bérégoovoy (1992)	0.323
273	Barre (1967)	Bérégoovoy (1992)	0.323
274	Pompidou (1962B)	Fabius (1984)	0.324
275	Mauroy (1984)	Bérégoovoy (1992)	0.325
276	Mauroy (1984)	Juppé (1995)	0.325
277	Couve (1968)	Mauroy (1984)	0.325
278	Mauroy (1984)	Jospin (1997)	0.326
279	Pompidou (1966)	Mauroy (1983)	0.326
280	Pompidou (1962A)	Bérégoovoy (1992)	0.326
281	Chaban (1969)	Bérégoovoy (1992)	0.326
282	Mauroy (1984)	Cresson (1991)	0.327
283	Couve (1968)	Bérégoovoy (1992)	0.327
284	Pompidou (1962B)	Mauroy (1984)	0.328
285	Debré (1959)	Bérégoovoy (1992)	0.329
286	Pompidou (1962A)	Mauroy (1983)	0.329
287	Messmer (1974)	Bérégoovoy (1992)	0.329
288	Pompidou (1967)	Mauroy (1984)	0.331
289	Pompidou (1967)	Bérégoovoy (1992)	0.334
290	Debré (1959)	Chirac (1987B)	0.334
291	Debré (1959)	Mauroy (1984)	0.334
292	Couve (1968)	Barre (1978)	0.335
293	Barre (1978)	Chirac (1987B)	0.336
294	Barre (1978)	Fabius (1984)	0.337
295	Couve (1968)	Mauroy (1983)	0.341
296	Pompidou (1962B)	Bérégoovoy (1992)	0.343
297	Mauroy (1983)	Bérégoovoy (1992)	0.343
298	Barre (1978)	Mauroy (1984)	0.347
299	Debré (1959)	Mauroy (1983)	0.349
300	Barre (1978)	Bérégoovoy (1992)	0.360

On s'attend à trouver ici des couples opposant des locuteurs de camps opposés (droite/gauche). C'est effectivement le cas pour 23 d'entre eux et, dans une majorité des cas, le temps ajoute ses effets à ceux d'une possible distance idéologique. Mais 4 couples opposent des socialistes entre eux : n° 271 (P. Bérégoovoy – L. Jospin) ; n° 275 (P. Mauroy – P. Bérégoovoy) ; n° 278 (P. Mauroy – L. Jospin) ; n° 282 (P. Mauroy – E. Cresson). Et 3 couples distants opposent des Premiers ministres de droite : M. Debré – J. Chirac (n° 290) ; M. Couve

de Murville – R. Barre (n° 292) ; R. Barre – J. Chirac (n° 293). Dans certains cas, on peut invoquer le temps écoulé entre les deux hommes (par exemple, M. Debré et J. Chirac) mais cette explication ne tient pas dans la majorité des cas. Il faut donc admettre une sérieuse différence de thèmes entre ces discours tenus par des hommes appartenant à un même courant politique.

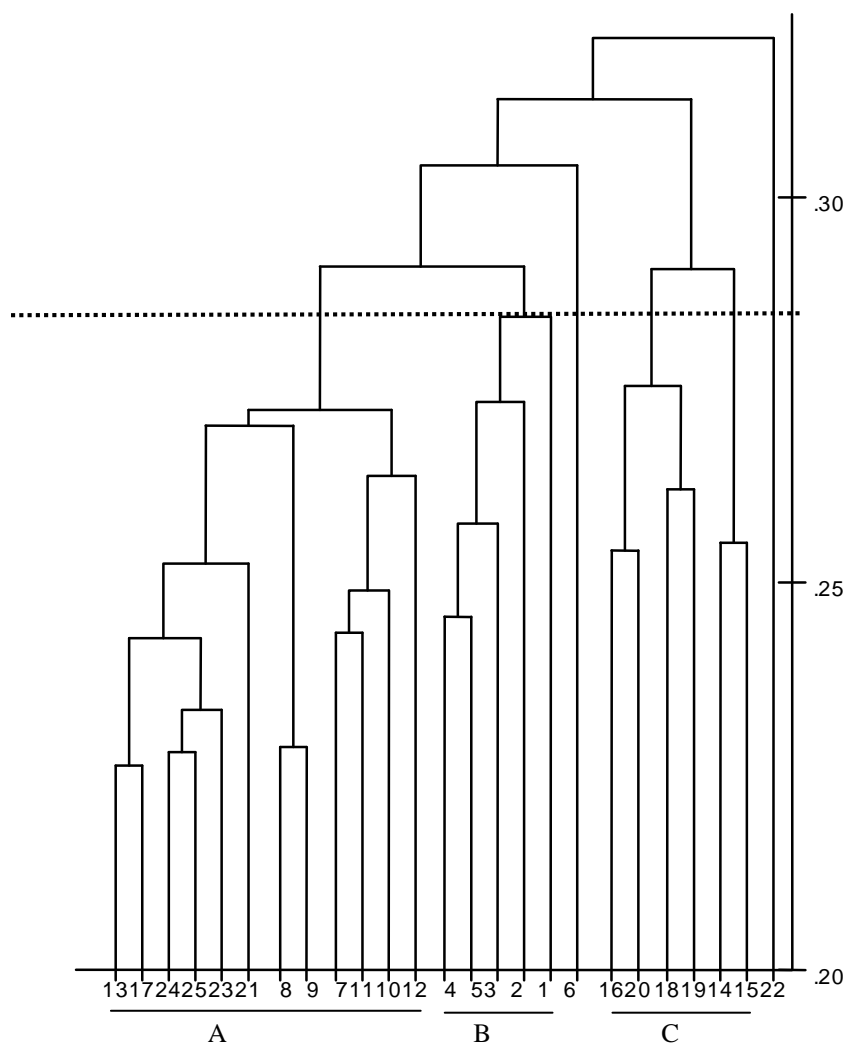
A ces constats, s'ajoute une remarque complémentaire. Dans le tableau 3, apparaissent un certain nombre d'orateurs ou de discours absents du tableau 1 (P. Bérégovoy, M. Debré, M. Couve de Murville). Le rapprochement des deux tableaux suggère donc que ces trois hommes sont relativement "originaux" par rapport au reste de la série... Pour vérifier cette intuition, on calcule, pour chaque texte, la moyenne des 24 distances le séparant de tous les autres. Cette mesure donne l'éloignement relatif du texte en question par rapport au centre de gravité du nuage formé par les 300 distances (tableau IV).

Tableau IV. Eloignement de chaque texte par rapport au centre de gravité de la population

Rang	Discours	Distance moyenne
1	P. Mauroy (1981)	0,266
2	E. Balladur (1993)	0,268
3	J. Chirac (1986)	0,269
4	J. Chaban-Delmas (1969)	0,273
5	P. Messmer (1972)	0,277
6	P. Messmer (1973)	0,278
7	A. Juppé (1995)	0,279
8	J. Chirac (1974)	0,279
9	R. Barre (1977)	0,279
10	M. Rocard (1988)	0,281
11	E. Cresson (1991)	0,285
12	J. Chirac (1987A)	0,286
13	G. Pompidou (1962B)	0,290
14	L. Jospin (1997)	0,290
15	G. Pompidou (1966)	0,291
16	G. Pompidou (1967)	0,292
17	G. Pompidou (1962A)	0,294
18	J. Chirac (1987B)	0,297
19	R. Barre (1978)	0,297
20	L. Fabius (1984)	0,299
21	M. Debré (1959)	0,299
22	M. Couve de Murville (1968)	0,305
23	P. Mauroy (1983)	0,308
24	P. Mauroy (1984)	0,309
25	P. Bérégovoy (1992)	0,319
	Moyenne	0,288

Les déclarations de P. Mauroy (1981), d'E. Balladur et de J. Chirac (1986) fournissent des sortes d'"archétypes" de la déclaration gouvernementale. A l'opposé, le même P. Mauroy – en 1983 et 1984 – P. Bérégovoy ou M. Couve de Murville ont produit des textes décalés. La hiérarchie suggère que les lignes de clivage séparent des types de situation politique et non des couleurs politiques. En haut du tableau : les Premiers ministres en début de mandat, assurés d'une large majorité et avec une situation économique et sociale plutôt favorable. A l'opposé, les Premiers ministres sans majorité assurée (P. Bérégovoy, M. Debré, R. Barre) ou faisant face à des désunions dans leur propre camp, voire à des crises politiques ou sociales :

Tableau V. Classification automatique des 25 déclarations gouvernementales



De gauche à droite :	Groupe B
Groupe A :	4. G. Pompidou 1966
13. P. Mauroy 1981	5. G. Pompidou 1967
17. J. Chirac 1986	3. G. Pompidou 1962B
24. A. Juppé 1995	2. G. Pompidou 1962A
25. L. Jospin 1997	1. M. Debré 1959
23 E. Balladur 1993	6. Couve de M. 1968
21 E. Cresson 1991	Groupe C :
8. P. Messmer 1972	16. L. Fabius 1984
9. P. Messmer 1973	20. M. Rocard 1988
7. J. Chaban-Delmas 1969	18. J. Chirac 1987A
11. R. Barre 1977	19. J. Chirac 1987B
10. J. Chirac 1974	14. P. Mauroy 1983
12. R. Barre 1978	15. P. Mauroy 1984A

En coupant le graphe suivant des seuils mentionnés ci-dessus, apparaissent les groupes des textes très proches, relativement proches, etc. Par exemple, la moyenne (trait pointillé sur le

graphe), isole trois classes (voir la légende placée sous le dendrogramme). Le classement n'obéit pas à une logique d'opposition entre deux camps. En dehors de quelques individus singuliers (le couple Mauroy-Chirac, M. Couve de Murville et P. Bérégovoy). Les gouvernements sont groupés selon une logique chronologique : à gauche les années 1969-97, au centre les années 1959-1967 et à droite les années 1984-88.

Cependant, la classification automatique présente un inconvénient majeur : lorsque deux textes sont fusionnés, les relations spécifiques que chacun entretient avec tous les autres sont remplacées par une moyenne qui peut être d'autant plus trompeuse que l'on s'élève dans l'arbre.

Pour pallier à ces inconvénients, plusieurs méthodes "multi-dimensionnelles" ont été imaginées.

La méthode la plus usuelle consiste à représenter l'ensemble des textes par des points dont les coordonnées dans l'espace sont déterminées par leur position relative par rapport à tous les autres. Ici les 25 textes forment un "nuage" de points comprenant 300 dimensions différentes. Par la méthode des "moindres carrés", on détermine d'abord le plan qui passe au plus près de tous ces points (et par le barycentre du nuage) puis l'on projette orthogonalement chacun des points sur ce plan, ce qui donne une représentation plane du nuage. Cette méthode a un inconvénient évident : un point sera d'autant plus fidèlement représenté qu'il sera proche du plan d'ajustement ; en revanche, les points les plus éloignés risquent d'être "mal" représentés.

La classification arborée ne présente pas de tels inconvénients.

II.2 Analyse arborée

L'analyse arborée est assez usuelle en biologie ou en génétique. Elle repose sur le théorème suivant : si tous les points d'une série sont séparés par des distances euclidiennes, il existe un "arbre" qui représente sans biais les positions respectives de chacun de ces points par rapport à tous les autres¹. La construction d'un tel arbre "parfait" exigerait que l'on examine toutes les combinaisons possibles alors que le nombre ces combinaisons augmente exponentiellement en raison de l'effectif de la série. Divers algorithmes ont été imaginés pour construire cet arbre sans avoir à examiner toutes ces combinaisons².

Pour présenter l'analyse arborée, examinons d'abord le cas de deux couples extrêmes et assez étranges (tableau VI) et l'arbre construit à partir de ce tableau élémentaire (tableau VII).

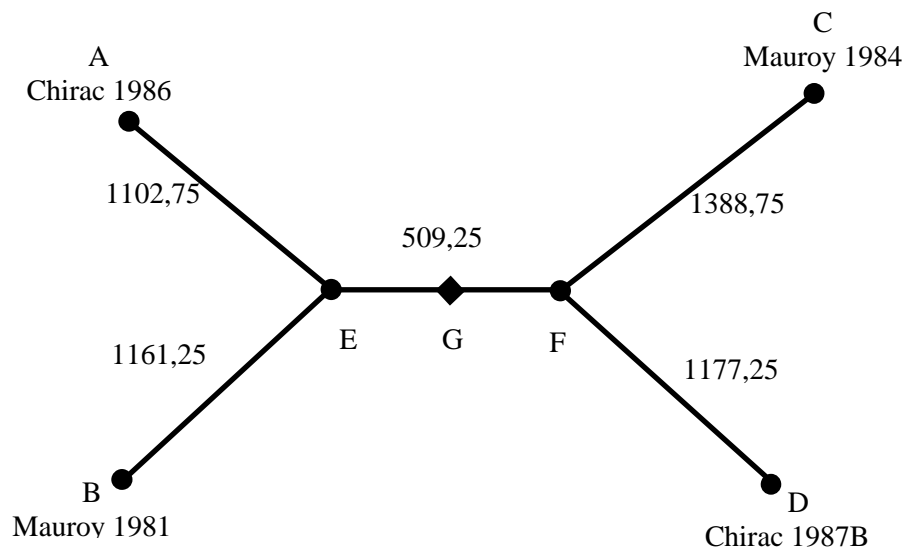
Tableau VI. Distances entre 4 déclarations gouvernementales (exprimées en "pour 10000 mots")

	P. Mauroy 1981	P. Mauroy 1984	J. Chirac 1986	J. Chirac 1987B
P. Mauroy 1981	-	2936	2264	2854
P. Mauroy 1984	2936	-	3124	2566
J. Chirac 1986	2264	3124	-	2783
J. Chirac 1987B	2854	2566	2783	-

¹ Pour la démonstration, voir : Luong Xuan. *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Université de Paris V, 1988. Concernant l'analyse arborée, voir également : Barthélémy Jean-Pierre et Guénoche Alain. *Les arbres et les représentations de proximité*. Paris. Dunod, 1988.

² Nous utilisons l'algorithme mis au point par X. Luong de l'Université de Nice (*op. cit.*). Le logiciel de calcul des arbres a été réalisé avec son aide.

Tableau VII. Classification arborée des 4 déclarations du tableau VI



Cette figure est l'arbre le plus petit que l'on puisse tracer (il faut au moins 4 points). A, B, C et D sont les "sommets" ou **feuilles** de l'arbre. Les feuilles A et B sont "adjacentes" (les plus proches) de même que C et D. Ces deux couples forment des groupes "opposés". Les points E et F matérialisent ces relations d'association et d'opposition : ce sont les **nœuds** de l'arbre. Ils ont été créés par l'algorithme qui a établi leurs positions relatives en calculant la longueur des segments de droite AE, BE, CF et DF (appelées arrêtes ou **branches**). Chacune des quatre feuilles est reliée par un **chemin** qui est composé d'au moins deux arrêtes.

Ces chemins permettent de visualiser la proximité et l'éloignement relatifs d'un texte par rapport aux autres. Par exemple, BEFC est le plus long chemin, ce qui rend visible le fait que la déclaration de P. Mauroy en 1981 est plus proche des deux de J. Chirac (1986 et 1987) que de la sienne propre en 1984. Certes, cette information est bien visible dans le tableau VI mais, dans une matrice de plusieurs milliers de cases, il y a fort à parier que ce genre d'information "contre-intuitive" sera perdue.

Le point G est la **racine** de l'arbre. Dans le graphe ci-dessus, elle est placée au milieu du **tronc** central parce que les points E et F ont été créés simultanément par l'algorithme. Habituellement, G est le dernier point placé. Il matérialise le "centre de gravité" du nuage des points.

Enfin, il faut remarquer que l'arbre est "connecté" - toute feuille est reliée au moins à une autre par une arrête - qu'il ne comporte aucun "circuit" - un seul chemin relie deux points quelconques - et qu'il est "valué" : tout segment a une longueur positive et proportionnelle à la valeur correspondante dans le tableau VI. Par exemple, il est facile de vérifier que la distance entre Mauroy 1981 et Chirac 1986 est bien égale à 2264 (1102,75 + 1161,25), etc.

Pour lire ces arbres, il faut se souvenir que la position absolue des feuilles et des nœuds n'a pas d'importance - Mauroy 81 pourrait aussi bien se situer en haut à droite, Chirac 86 en bas à droite, etc. - et que seuls comptent la longueur des chemins à parcourir pour aller de l'un à l'autre et le rattachement de certains textes à un même nœud qui attestent leur appartenance à un même groupe.

Comment estimer la qualité de la représentation et donc la confiance que l'on peut porter à telle ou telle partie de l'arbre ou à la figure entière ? Un indice simple permet de répondre à cette question : le rapport entre la distance intertextuelle et la taille du chemin correspondant sur l'arbre (tableau VIII ci-dessous).

Tableau VIII. Construction et qualité de l'arbre

Noeud	Chemins induits	Distances initiales	Distances arborées	Qualités des chemins (%)	Qualité du nœud
E	Mauroy81 - Chirac86	2264	2264,00	100,0	100,0
F	Mauroy84 - Chirac87B	2566	2566,00	100,0	100,0
G	Mauroy81 - Mauroy84	2936	3000,75	97,8	97,8
	Mauroy81 - Chirac87B	2854	2789,25	97,7	
	Chirac86 - Mauroy84	3124	3059,25	97,9	
	Chirac86 - Chirac87B	2783	2847,75	97,7	
Total		16527	16527,00	98,5	

Si la qualité des chemins entre couples de feuilles contiguës est toujours parfaite, il n'en est pas de même quand le chemin est plus complexe, c'est-à-dire ici pour les 4 chemins passant par le tronc central (AC, AD, BC, BD). Au total, pour ces 4 chemins, l'indice de qualité moyen est égal 97,8 % (ou encore : il pèse une incertitude globale de 2,2% sur ces 4 chemins). Au total, comme indiqué dans la dernière ligne, 98,5% de l'information totale contenue dans la matrice initiale est restituée par le graphe.

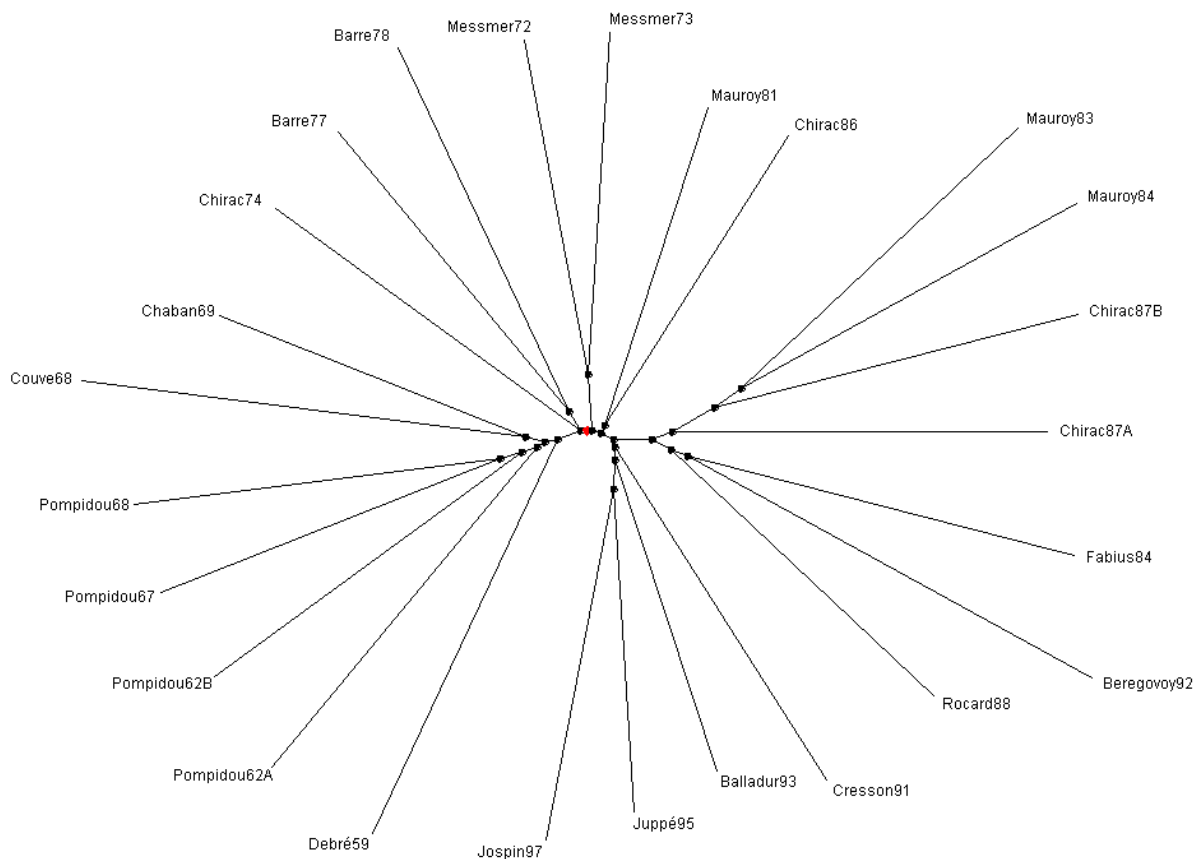
Ces légères discordances proviennent de deux sources :

- le procédé de calcul de l'indice de la distance intertextuelle (exposé ci-dessus) engendre une légère incertitude (il s'agit ici essentiellement du problème des décimales) ;
- l'algorithme de construction de l'arbre commence par mettre en place les feuilles terminales et réalise les arbitrages sur le tronc central.

Sous ces réserves, la classification arborée s'avère être actuellement la méthode la plus précise et la plus sûre pour représenter les proximités et les distances entre un grand nombre d'individus.

Voici l'arbre complet classant les 25 déclarations gouvernementales selon le même procédé.

Tableau IX. Classification arborée des 25 déclarations gouvernementales (1959-1997)



Avant de présenter l'interprétation de ce graphe, signalons que la qualité de l'arbre total est de 97,1%, que tous les nœuds ont une qualité supérieure à 96% et que 253 des 300 chemins ont une qualité supérieure à 95% ; 43 ont une qualité comprise entre 90 et 95% et 4 une qualité légèrement inférieure à 90%¹. On peut accorder toute confiance à ce graphe.

Trois conclusions s'imposent.

Premièrement, la forme "étoilée" du graphique souligne la relative "équidistance" de tous ces textes. Cela confirme les conclusions déjà tirées de l'examen direct des principales distances concernant l'homogénéité remarquable de ces discours.

Deuxièmement, la disposition des points obéit assez fidèlement à l'ordre chronologique. En partant du quart sud-ouest Depuis le discours de M. Debré en 1959 – situé au quart sud-ouest de l'arbre - les textes se répartissent chronologiquement en suivant les aiguilles d'une montre. Cette régularité est toutefois légèrement altérée par quelques couples particuliers - Messmer (1972-73), Mauroy (1981)-Chirac(1986) – et par un trio singulier : Fabius (1984), Rocard (1988), Bérégovoy (1992).

Troisièmement, dans le détail, on retrouve les principales classes issues de la classification automatique. A condition de laisser à part les individus singuliers signalés ci-dessus :

- le premier groupe contient tous les textes des onze premières années (1959-1969) ;
- le second groupe concerne les "années Giscard" (1974-1988) ;
- le troisième rassemble les années 1980, à l'exception de P. Mauroy (1981) et J. Chirac (1986)

¹ Tous ces nœuds relient Couve de Murville à Bérégovoy (89,6%), à Chirac87B (88,1%), à Chirac87A (87,0%) et à Fabius (86,6%). Tous ces chemins sont un peu plus longs sur le graphe qu'ils ne le sont dans la réalité mais aucune de ces distorsions ne remet en cause la position des feuilles sur l'arbre ni son architecture générale.

- le quatrième les années 1990 (si l'on excepte P. Bérégovoy, rattaché au groupe précédent).

Outre l'influence prédominante de la chronologie, le principe organisateur serait la conjoncture du moment.

Conclusions

L'analyse se poursuit par le calcul du vocabulaire caractéristique de chacun de ces groupes, l'analyse des structures de phrase et du style propre à chaque locuteur. On utilise pour cela des outils statistiques qui ont été déjà présentés lors de précédentes recherches¹.

Nous espérons avoir suggéré combien est prometteuse la combinaison de la distance intertextuelle avec les classifications par ordinateur, notamment l'analyse arborée. Naturellement, une analyse de qualité exige un traitement soigné des textes : correction orthographique et standardisation des graphies car il serait sans intérêt de prétendre mesurer finement un phénomène dont les observations auraient été faites sans aucun soin. Sous cette condition et dans les limites qui ont été indiquées, cette méthode fournira un outil puissant pour les chercheurs confrontés à un volume grandissant d'archives électroniques qui pourront être classées de manière objective.

La présente analyse aboutit à deux conclusions intéressantes.

Premièrement, sous la Ve République, le genre "déclaration gouvernementale" semble obéir à un moule assez rigide. En effet, des distances aussi faibles se rencontrent assez rarement dans les discours politiques tenus par des orateurs différents. De telles proximités surviennent dans certaines situations particulières : des genres très codifiés – par exemple, les circulaires administratives, les arrêts des hautes juridictions (conseil d'Etat, cour de cassation, conseil constitutionnel...) – ou des textes qui sont un peu des clones les uns des autres, comme par exemple les articles de presse portant sur un même événement et utilisant les mêmes dépêches d'agences, etc.

Il faut donc admettre que les Premiers ministres – en fait leurs collaborateurs – s'"inspirent" assez largement des déclarations de leurs prédécesseurs. Mais il faut sans doute aller plus loin. Le processus d'élaboration des discours ministériels emprunte la voie suivante : les directeurs d'administration centrale élaborent des notes de synthèse que recopient largement les collaborateurs chargés de la mise en forme du texte. Tous ces gens se ressemblent d'ailleurs beaucoup. Ils sont issus des mêmes milieux sociaux, des mêmes écoles, fréquentent les mêmes salons, ils ont les mêmes lectures, les mêmes sources d'information et ne changent guère d'affectation au gré des élections... Il n'est donc pas absurde de considérer que l'on se trouve en face d'une sorte de "locuteur collectif".

Deuxièmement, pour ce qui concerne les déclarations gouvernementales, l'axe droite-gauche n'est pas pertinent. La seule différence tient à la conjoncture politique et à la solidité plus ou moins grande du gouvernement.

L'opinion selon laquelle le clivage droite/gauche n'aurait plus de sens (au moins au niveau des politiques gouvernementales) est aujourd'hui largement répandue. Beaucoup de Français estiment que les hommes politiques parlent de la même façon et qu'ils disent tous la même chose. L'analyse des déclarations gouvernementales ne permet pas de leur donner tort, mais, naturellement, il faudrait des corpus plus vastes pour vérifier que cette homogénéité du discours existe effectivement à tous les niveaux de la vie politique et qu'elle n'est pas propre aux déclarations gouvernementales.

¹ Voir notamment : Labbé Dominique. Le vocabulaire de F. Mitterrand. Op. cit.

Annexe
Corpus des déclarations gouvernementales sous la Ve République (1959-1967)

Premier ministre	Date de la déclaration	Longueur (mots)	Vocables différents
M. Debré	janvier 1959	9 168	1 556
G. Pompidou	avril 1962	4 528	1 094
G. Pompidou	décembre 1962	5 569	1 227
G. Pompidou	avril 1966	6 960	1 381
G. Pompidou	avril 1967	5 328	1 132
M. Couve de Murville	juillet 1968	4 354	944
J. Chaban-Delmas	juin 1969	9 219	1 674
P. Messmer	octobre 1972	7 802	1 561
P. Messmer	avril 1973	7 123	1 465
J. Chirac	? 1974	8 766	1 539
Raymond Barre	avril 1977	7 291	1 424
Raymond Barre	? 1978	6 359	1 283
P. Mauroy	? 1981	12 136	1 974
P. Mauroy	? 1983	9 926	1 590
P. Mauroy	? 1984	5 910	1 192
L. Fabius	? 1984	5 536	1 169
J. Chirac	avril 1986	9 800	1 873
J. Chirac	avril 1987	8 171	1 412
J. Chirac	décembre 1987	7 532	1 472
M. Rocard	juin 1988	8 019	1 670
E. Cresson	mai 1991	5 693	1 253
P. Bérégovoy	avril 1992	5 708	1 248
E. Balladur	avril 1993	11 522	1 777
A. Juppé	mai 1995	13 590	2 046
L. Jospin	juin 1997	6 982	1 418
Total		192 992	6 632