

Les déclarations gouvernementales se suivent et se ressemblent. Exploration d'une chronique textuelle

Les corpus de discours politiques émis successivement sur d'assez longues périodes présentent souvent un aspect particulier du point de vue lexical: les déterminations du vocabulaire par les conflits, idéologies et approches partisans semblent subordonnées à un phénomène de proximité lexicale entre discours, liée à la proximité temporelle. Ce constat a suscité plusieurs approches en vue de comprendre et de disséquer en quelque sorte ce phénomène chronologique. André Salem propose une extension des méthodes d'analyse des "spécificités" au cas chronologique. Pour chaque forme lexicale suffisamment attestée dans le corpus, il peut définir soit les périodes regroupant plusieurs parties où elle apparaît trop ou trop peu souvent que ne le laisserait supposer une répartition moyenne, soit les moments d'accroissement ou de décroissance sensible de l'emploi de cette forme. L'immense intérêt de cette méthode réside dans la possibilité d'étudier finement comment une forme apparaît et disparaît dans un vocabulaire en évolution. Cependant, de même que la méthode des spécificités se complète souvent par des analyses plus globales des corpus, il faut pouvoir définir des méthodes globales d'analyse de la "chronogénéricité" des séries textuelles. Dans ce domaine, les travaux de Dominique Labbé ne peuvent être passés sous silence, car son approche du renouvellement met en évidence, dans ces séries, l'existence d'un vocabulaire général et de vocabulaires spécialisés qui s'introduisent, selon les nécessités du moment, dans le discours politique.

Nous nous proposons d'exposer ici une tentative de décomposition de l'aspect chronologique basée sur des méthodes de classification automatique à partir d'une chronique textuelle composée de l'ensemble des déclarations gouvernementales belges (*dg* dans la suite de ce texte) prononcées entre 1944 et la fin de l'État national, soit 1992. Le corpus comprend les trente-huit déclarations faites à cette époque devant le Parlement. Dans le système politique belge de type bicaméral, la formation du Gouvernement se déroule suivant un processus qui doit plus à la coutume parlementaire qu'au droit constitutionnel proprement dit. Cependant, quel que soit le processus poursuivi par le Roi pour composer un gouvernement - consultations, nomination d'informateurs, d'un formateur, désignation de l'équipe ministérielle - la dernière étape consiste en l'obtention obligatoire de la confiance des deux Chambres du Parlement, le Sénat et la Chambre des représentants. Cette confiance n'est accordée que sur la base d'une déclaration faite au nom du

[°] Haute école "Paul Henri Spaak", Bruxelles, Belgique.

Gouvernement par le Premier ministre¹ devant chacune des Chambres. Cette déclaration contient les éléments du programme du nouveau cabinet et permet aux parlementaires d'exprimer un jugement politique sur l'équipe qui se présente à eux.

Le système électoral du pays durant la période étudiée est de type proportionnel; cela rend relativement difficile la constitution de gouvernements homogènes mais favorise la constitution de coalitions; la scission des partis nationaux - sociaux-chrétiens, libéraux et socialistes - en partis homogènes linguistiquement (Parti socialiste et Socialistische partij, Parti de la Liberté et du Progrès et Partij voor Vrijheid en Vooruit, Parti social-chrétien et Christelijke Volkspartij) a encore renforcé cette tendance à la dispersion des partis représentés au Parlement. La conjugaison de ces éléments avec la création de partis dits "communautaires" aboutit à la constitution de cabinets pouvant comporter jusqu'à six partenaires. Dans la période qui nous occupe, il n'y eut en fait que quatre gouvernements homogènes, tous sociaux-chrétiens entre 1950 et 1954, et en 1958. Il faut aussi souligner qu'à l'exception de deux courtes périodes (1945-1947 et 1954-1958), le ou les partis sociaux-chrétiens ont été en permanence au pouvoir, seuls ou associés soit aux socialistes soit aux libéraux. La Belgique est donc gouvernée au centre, car c'est là que se situe la famille sociale-chrétienne, parfois au centre gauche avec les socialistes, parfois au centre droit avec les libéraux.

C'est cette césure gauche/droite, certes édulcorée par les nécessités du pouvoir et les compromis inhérents à la constitution de coalitions, que nous pensions *a priori* retrouver dans l'étude lexicométrique des déclarations gouvernementales. Mais l'analyse du vocabulaire² des trente-huit déclarations prononcées devant le Parlement depuis la seconde guerre mondiale nous a surtout permis de mettre en évidence la conjonction de deux sources de structuration des choix lexicaux: une évolution chronologique du vocabulaire et une certaine opposition entre discours volontariste et discours de gestion. L'aspect chronologique est dominant et se marque dans les analyses factorielles des correspondances (AFC) du tableau des formes.

Les déclarations gouvernementales se suivent et se ressemblent

La structuration du corpus par le temps se manifeste tout d'abord massivement dans la bipartition de l'ensemble des déclarations visible sur le premier plan factoriel

¹ La déclaration dont il existe deux versions, l'une néerlandaise, l'autre française, est lue devant les deux Chambres généralement par le Premier ministre ou par un autre ministre important du Gouvernement, en alternant le français et le néerlandais. Les deux textes ont la même valeur politique.

² L'étude du vocabulaire a été conduite suivant les normes de "Saint-Cloud", c'est-à-dire sur les formes graphiques, sans lemmatisation. Nous n'envisageons dans cet article que l'étude des formes simples sans tenir compte ni des segments répétés (Lafon et Salem 1983, Salem 1987) ni des quasi-segments (Bécue 1993).

issu de l'analyse des correspondances du tableau des formes simples [voir Figure 1]³. Les déclarations antérieures à 1965, c'est-à-dire celles qui ont été prononcées de 1944 (Pierlot-dg 00) à 1961 (Lefevre-dg 16), se regroupent toutes dans la partie droite de l'axe horizontal tandis que les déclarations les plus récentes se situent toutes dans la partie gauche. On sait combien, pour les historiens de la vie politique belge, il est une césure importante de cette vie, qu'ils situent au début des années 1960. Els Witte et Jan Craeybeckx⁴, par exemple, ne distinguent que deux grandes périodes dans l'après-guerre belge: de 1944 à 1960 et de 1961 à 1987.

D'une certaine façon, cette coupure se justifie par de nombreux éléments politiques. 1961 voit s'exacerber les tensions linguistiques, qu'on ne nomme pas encore communautaires, entre Wallons et Flamands; 1959 est l'année de la mise en application de la paix scolaire qui fige les positions entre les partisans de l'école laïque et ceux de l'école catholique et rend possible, entre autres, la création d'un nouveau parti libéral (PLP) sur la base d'une ouverture du parti libéral à une partie du centre droit catholique; 1960 marque aussi la prise de conscience du redéploiement industriel qui s'opère globalement de la Wallonie vers la Flandre, des régions de vieille industrialisation vers des régions, il y a peu, encore rurales. Évidemment, *a posteriori*, nous savons bien que ces années-là marquent aussi le début d'une période de croissance économique dont les effets sur les plans sociaux, politiques, éducatifs et culturels ne seront pas à négliger. Il paraît logique de retrouver une césure d'une telle importance dans le vocabulaire des déclarations gouvernementales. Cependant, nous devons constater que le premier gouvernement de la nouvelle période de l'histoire de la Belgique (dg 16 - Lefèvre-Spaak - 1961) voit sa déclaration continuer à se rattacher à l'ensemble précédent. Comme si ce gouvernement, qui inaugure les "Golden sixties", le processus de transformation de l'État unitaire en État fédéral et le désengagement du système colonial (le Congo est indépendant depuis 1960, le Ruanda-Urundi le deviendra en 1965), faisait tout cela avec le vocabulaire et les préoccupations du passé, sans anticiper le parler politique à venir!

Au-delà de cette fragmentation des déclarations en deux blocs successifs, le temps structure plus fortement et de manière plus fine l'évolution du vocabulaire des programmes gouvernementaux. En effet, si l'on classe l'ensemble de nos déclarations suivant, d'une part, leur succession naturelle dans le temps et, d'autre part, suivant leur ordonnancement le long de l'axe 1 (horizontal) de l'AFC, nous pouvons mesurer la similitude entre ces deux ordres et, par exemple, calculer le coefficient de

³ Dans le tableau soumis à l'analyse, les 38 déclarations forment les colonnes, les 1355 formes constituent les lignes. Nous n'avons retenu que les formes dont la fréquence était au moins égale à 10 dans le corpus. Cinq déclarations ont été mises en "éléments supplémentaires" (représentés mais non efficaces dans le calcul), soit parce qu'elles sont trop courtes pour être considérées comme des exposés programmatiques, dans le cas de gouvernement de "transition", soit parce qu'elles ne traitent que d'un seul sujet, la réforme institutionnelle, et ne peuvent donc être considérées comme des exposés de politique générale. Les trois premiers axes factoriels représentent respectivement 8,56, 5,96 et 5,14% de l'inertie du tableau lexical (les % indiqués dans le graphique 1 sont à corriger).

⁴ J. Craeybeckx, E. Witte (1987), *La Belgique politique de 1830 à nos jours. Les tensions d'une démocratie bourgeoise*, Bruxelles, Labor (2e édition). Voir aussi X. Mabilie (1986-1992), *Histoire politique de la Belgique. Facteurs et acteurs de changement*, Bruxelles, CRISP.

corrélation de rang de Spearman⁵; une valeur de r_s de 0,814 indique une forte association entre les rangs des déclarations dans les deux séries, association qui s'intensifie lorsqu'on exclut du calcul quatre déclarations trop spécialisées ou très courtes (dg 26, dg 28, dg 34 et dg 36); nous obtenons dès lors un r_s égal à 0,938 [Annexes

Tableau 7 en annexe].

D'un point de vue heuristique, on doit aussi constater la forme caractéristique prise par la projection des points correspondant aux déclarations sur les plans formés par le premier axe avec respectivement l'axe deux et l'axe trois de l'AFC. Sur le premier de ces plans, le nuage épouse la forme d'une courbe du troisième degré en S couché [voir figure 1 en annexe] et sur le deuxième, on distinguera l'esquisse d'une parabole [voir figure 2 en annexe]. Pour faciliter la lecture, nous avons superposé à la figure 1 une courbe représentant un polynôme du troisième degré obtenue par ajustement des moindres carrés à partir des coordonnées des déclarations actives⁶ uniquement.

$$y_i = -6,48.x_i^3 + 0,28.x_i^2 + 0,73.x_i - 0,05$$

Les y_i sont les valeurs prédites des coordonnées sur l'axe 2 étant donné les x_i , coordonnées sur l'axe 1.

De même nous avons superposé une courbe représentant une fonction du second degré des coordonnées sur l'axe 1 obtenue par ajustement par les moindres carrés

$$z_i = -2,700901.x_i^2 - 0,01,191073.x_i + 0,1626254$$

Les z_i sont les valeurs prédites des coordonnées sur l'axe 3 étant donnés les x_i , coordonnées sur l'axe 1.

On sait que ces formes en S couché et paraboliques sont la représentation factorielle de tableaux qui présentent une structure avec une bande diagonale chargée, du moins si l'on ordonne les lignes et les colonnes de ces tableaux suivant l'ordre de leurs coordonnées sur l'axe 1 [Benzécri 1976, Salem 1988]. Comme cet ordre est proche de l'ordre chronologique dans lequel ces discours ont été rédigés, il y a là le signe indubitable d'une influence du temps sur le choix et l'emploi du vocabulaire.

Nous conviendrons dès lors qu'il existe une forte composante "chronologique" associée au premier axe de l'analyse factorielle. Cependant, l'examen attentif du

Annexes
Tableau 7 nous rappelle que cet effet n'est pas strict mais fonctionne plutôt par sous-ensembles. Aussi la coupure est-elle remarquable entre dg 15 et dg 16, dont le rang dans les deux séries - naturelle et factorielle - est identique, terminant bien la première des deux périodes articulées autour de 1960. La mise en évidence de cet

⁵ Si N est égal au nombre d'éléments à classer et si les d_i représentent les différences entre les rangs obtenus par les éléments dans les deux séries, alors le coefficient de corrélation de rang r_s est : $r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N}$

⁶ . C'est-à-dire l'ensemble des déclarations à l'exception des déclarations trop courtes ou trop spécialisées.

aspect chronologique amène à s'interroger sur la caractérisation de cet axe du "temps" par les formes associées.

Les méthodes d'analyse d'une série textuelle chronologique

Si l'étude des déclarations conduit à faire l'hypothèse d'une tendance à l'évolution chronologique dans l'emploi et la "création" du vocabulaire politique, c'est assurément du côté de la régularité de la répartition des formes que nous trouverons des indications sur la signification de cette dérive temporelle. L'examen des projections des points représentant les formes sur le premier plan factoriel nous donne une image de cette évolution, dont nous avons donné une première interprétation [Deroubaix-Gobin 1987]. Cette première approche, qui met clairement le phénomène "chronologique" en évidence, nous incite cependant (et surtout) à affermir nos conclusions et à affiner l'analyse plus qu'à fournir des assurances. L'analyse factorielle des correspondances joue ici pleinement son rôle de méthode heuristique.

La méthode utilisée repose sur une certaine modélisation floue de ce que peut être le processus chronologique sous-jacent. De la même manière que Dominique Labbé [Labbé 1990], nous supposons qu'il provient du renouvellement plus ou moins permanent du vocabulaire politique suivant les problèmes du moment, qu'il y a donc apport, "invention" de nouveaux mots et concomitamment disparition de certains d'entre eux. Mais ces mots ne forment pas tout le vocabulaire; il subsiste à leur côté un vocabulaire général du corpus comportant tous les mots régulièrement utilisés par tous les textes avec des fréquences assez régulièrement distribuées [Deroubaix 1992]. Mais ce modèle, nous pensons pouvoir le développer sans recourir à l'image des réservoirs multiples, l'un pour le vocabulaire général, les autres pour les vocabulaires spécialisés, mais simplement en cherchant à distinguer, parmi des classes de formes dont la fréquence d'emploi est similaire, celles dont le profil indique un vocabulaire "spécialisé" au sens de D. Labbé et celles dont le profil les rapproche d'un vocabulaire général. En choisissant de créer ces classes de vocabulaire comme partition⁷ de l'ensemble des formes lexicales, nous décomposons en quelque sorte le vocabulaire en autant de paquets qu'il y a de vocabulaires spécialisés, plus un consacré au vocabulaire général. Parmi les vocabulaires spécialisés, nous pouvons alors tenter de repérer ceux qui sont de type chronologique.

En pratique, la méthode que nous avons utilisée consiste à opérer une classification ascendante hiérarchique (CAH) sur les lignes du tableau lexical (c'est-à-dire sur les profils d'emploi des formes), à choisir une bonne partition et à consolider la partition par une méthode de "centres mobiles".

Le principe de la classification ascendante hiérarchique consiste à considérer tout d'abord la partition de l'ensemble des éléments à classer en autant de classes

⁷ Cela signifie qu'une forme lexicale appartient à une et une seule classe. Rappelons que nous ne retenons que les formes dont l'utilisation dans l'ensemble du corpus atteint au moins la fréquence 10.

qu'il y a d'éléments, soit N , chaque classe comportant un et un seul élément; à fusionner ensuite les deux classes qui soient telles que de toutes les paires de classes envisageables ce soient celles dont la fusion crée la classe la plus homogène⁸, la moins dispersée. En poursuivant le processus de fusion de proche en proche on obtiendra, à chaque itération, une nouvelle partition de l'ensemble initial, successivement en $N, N-1, N-2, \dots, 3, 2, 1$ classes. Pour opérer le choix d'une bonne partition parmi cet ensemble, nous prendrons comme critère la valeur de l'indice d'homogénéité et élirons une partition dont la fusion de deux classes induit la plus grande augmentation de l'inertie intra-classe, autrement dit fait perdre le plus en homogénéité des classes⁹. La partition en six classes correspond à ce critère. Cependant, la méthode de constitution de la partition par une CAH nous garantit que cette partition en six classes est la meilleure de toutes les partitions en six classes que l'on peut obtenir à partir de la partition en sept classes; elle ne nous assure nullement que cette partition est la meilleure en six classes.

Nous aurons recours alors à une autre technique de classification automatique afin d'améliorer la partition obtenue. Pour chacune des classes de la partition initiale en six classes (P_0) nous allons calculer le profil des centres de classe (qui est une moyenne pondérée des profils des formes incluses dans la classe). Nous allons engendrer une nouvelle partition en six classes (P_1) en affectant chaque forme à la classe agrégée autour du centre d'une des classes de la partition P_0 . En calculant les centres des classes dans la nouvelle partition P_1 , en affectant les formes aux classes agrégées autour des centres de cette partition P_1 pour créer une partition P_2 et en réitérant la procédure, nous finirons par arriver à une partition stable qui ne se modifie plus. On peut montrer¹⁰ que la partition obtenue est identique ou meilleure à la partition P_0 de départ.

La répartition du vocabulaire du corpus en six classes fournit une première indication sur le processus sous-jacent à l'aspect "chronologique" de la distribution du vocabulaire. En effet, nous voyons apparaître deux types de classes particulières: une classe "stable", relativement bien répartie dans les discours (et donc dans le temps), cinq classes caractérisées chacune par une fréquence d'emploi plus élevée dans une des périodes du corpus; les périodes pouvant se chevaucher (cf. Tableau 1). L'aspect croissant puis décroissant de ces fréquences se lit dans les six colonnes du tableau, à l'exception de la classe 1, que nous avons qualifiée de stable, et de la classe 3, dont la décroissance en bas du tableau est tronquée¹¹.

⁸ L'homogénéité de la classe, qui garantit que les éléments qu'elle comporte sont de profil semblable, est mesurée par le moment centré d'inertie d'ordre deux, la distance entre profils servant au calcul de ce moment étant la distance du χ^2 , dite aussi distance de Benzécri.

⁹ Ce type de méthode a fait l'objet de multiples exposés et applications. On en trouvera des exemples dans les ouvrages cités. Plus spécialement pour la CAH, on trouvera des algorithmes bien décrits dans (Chambon et Pinson 1980), dans (Roux 1985) et dans (Jambu 1978).

¹⁰ Cf. note 8.

¹¹ À proprement parler, la classe 3 et peut-être aussi la classe 1 ne correspondent pas au modèle unimodal, à la classe 6 manque la partie gauche croissante et à la classe 4 la partie droite décroissante, mais il s'agit aussi des deux classes caractéristiques du début et de la fin de la période étudiée; il est raisonnable de penser, au vu de l'allure générale des autres classes, que ces deux classes appartiennent bien au même type et que seul le découpage historique est la cause de la troncature de leurs courbes.

Tableau 1

Répartition du vocabulaire en six classes (% dans chaque déclaration)

Poids des formes (%)						
	Classe 1	Classe 2	Classe3	Classe 4	Classe 5	Classe 6
dg00	65,47	1,72	2,34	2,34	6,15	21,97
dg01	60,03	0,92	2,61	2,33	5,79	28,32
dg02	58,27	0,87	1,83	2,50	6,44	30,10
dg03	54,51	0,96	2,38	2,55	4,67	34,94
dg04	66,17	0,96	2,63	3,41	9,98	16,86
dg05	64,09	1,41	3,12	2,60	11,87	16,91
dg06	57,99	0,94	2,22	3,12	7,58	28,15
dg07	58,85	1,56	2,15	1,85	8,14	27,45
dg08	62,67	1,05	4,01	5,05	12,02	15,20
dg09	65,41	1,08	3,31	4,29	12,85	13,05
dg10	63,56	2,28	2,71	1,95	12,36	17,14
dg11	64,82	1,25	3,60	4,27	12,58	13,48
dg12	61,78	0,91	2,93	5,55	9,70	19,13
dg13	64,26	1,04	3,61	9,66	5,27	16,15
dg14	66,40	1,39	3,52	7,57	7,03	14,09
dg15	64,60	1,86	4,27	10,29	5,69	13,29
dg16	64,23	1,09	4,54	10,39	5,56	14,19
dg17	69,33	2,71	4,55	4,33	4,08	15,01
dg18	61,77	1,53	4,97	5,71	3,44	22,57
dg19	67,28	2,73	4,12	4,31	2,92	18,64
dg20	67,49	2,85	6,58	6,19	4,08	12,82
dg21	64,70	2,93	5,26	2,99	3,52	20,61
dg22	60,90	8,59	4,47	2,52	5,16	18,37
dg23	67,45	3,49	6,90	3,80	3,78	14,59
dg24	62,85	8,81	5,16	2,08	3,44	17,67
dg25	67,81	1,59	9,01	2,79	2,90	15,90
dg26	63,95	1,58	8,16	3,16	4,74	18,42
dg27	66,24	1,61	13,21	2,32	3,93	12,69
dg28	57,53	9,01	8,24	2,84	5,28	17,10
dg29	64,04	2,37	12,95	2,29	4,04	14,31
dg30	64,66	2,03	13,19	3,84	3,36	12,92
dg31	67,21	1,50	12,34	3,24	4,68	11,03
dg32	70,29	1,37	11,73	2,48	3,79	10,33
dg33	68,17	1,28	10,17	2,82	5,03	12,53
dg34	65,58	3,88	8,37	2,02	5,27	14,88
dg35	66,46	1,41	14,71	4,29	3,59	9,54
dg36	60,40	3,49	7,99	2,16	7,32	18,64
dg37	64,85	1,19	11,84	3,43	2,33	16,36
corpus	64,51	2,03	5,79	4,63	6,16	16,89

L'examen à l'œil nu de ce tableau nous permet de distinguer facilement les cinq classes en forme de cloche. Il importe cependant de formaliser la démarche, afin d'être à même d'observer des classes de type chronologique parmi un ensemble plus grand de classes.

Nous proposons la méthode suivante. La définition que nous avons donnée des classes chronologiques suppose, au sens strict, que les parties du corpus qui recourent plus fréquemment que la moyenne au vocabulaire inclus dans une classe de type chronologique soient connexes dans le temps. D'une manière générale, cela signifie que la droite du temps se subdivise en trois parties: celle qui contient des discours où le vocabulaire spécifique à cette classe est suremployé, et deux parties caractérisées par le sous-emploi, la première précédant la zone de suremploi, la

deuxième suivant cette zone. Si l'on représente les zones de sous-emploi par un signe "-" et les zones de suremploi par un signe "+", une classe chronologique, au sens strict, se dessine ainsi:

----- (...) -----+++++++ (...) +++++----- (..) -----

On peut évidemment essayer de définir la "chronologie" de manière moins forte en admettant aussi des patrons tronqués soit au début du corpus, soit en fin de corpus et accepter ceci:

----- (..) -----+++++++ (..) +++++

ou

+++++++ (..) +++++----- (..) -----

Dès lors que l'on accepte ces modèles de chronologie, nous dirons qu'une classe est chronologique si la succession des zones où le vocabulaire de cette classe est en excès et celles où il est en déficit définissent une bipartition ou une tripartition de l'axe du temps, bipartition ou tripartition ordonnées des déclarations gouvernementales.

En représentant nos six classes selon cette technique, nous obtenons les suites:

Classe 1 +---+-----+-----+-----+-----+-----+-----+-----+
Classe 2 -----+-----+-----+-----+-----+-----+-----+-----+
Classe 3 -----+-----+-----+-----+-----+-----+-----+-----+
Classe 4 -----+-----+-----+-----+-----+-----+-----+-----+
Classe 5 +---+-----+-----+-----+-----+-----+-----+-----+
Classe 6 +---+-----+-----+-----+-----+-----+-----+-----+
Classe 6 +---+-----+-----+-----+-----+-----+-----+-----+

Aucune des classes ne correspond exactement à un des patrons de classe chronologique; certaines (les classes 2, 3, 4, 5) s'en approchent d'assez près. Nous pouvons mesurer l'écart entre le patron chronologique le plus proche et la structure réelle d'une classe, en comptant le nombre de parties du corpus mal classées.

Ainsi, la comparaison de la classe 2 à une tripartition idéale donnerait:

Classe 2 : -----+-----+-----+-----+-----+-----+-----+-----+
Idéal en trois : -----+-----+-----+-----+-----+-----+-----+-----+
Mal classés : 1 11 1 1

Cinq textes apparaissent mal classés. Le nombre de mal classés nous servira de mesure pour la qualité de la partition idéale proposée. Une adaptation de la méthode de Fisher¹² de partitionnement sous contrainte d'ordre total nous permettra de trouver la meilleure des partitions de l'ensemble des déclarations en k classes suivant le critère du nombre des mal classés.

¹² Cf. note 14.

Le découpage optimal de la séquence des sur- et sous-emplois.

Méthode

Il n'existe pas de stratégie qui nous garantisserait dans le cas général de trouver la partition optimale sauf dans le cas particulier de recherche d'une partition en k classes connexes d'objets d'un ensemble dans lequel il existe un ordre à respecter. Cela signifie qu'il est possible de chercher une partition en k classes de l'ensemble des discours du corpus telle que chacune des classes ne soit composée que d'éléments connexes dans le temps tout en étant assuré que cette partition est la meilleure¹³. En effet, cette possibilité est la conséquence de la réduction très importante du nombre de partitions possibles. Sans la contrainte de contiguïté des objets dans chaque classe, nous savons que le nombre de partitions possibles de N objets en K classes, nombre de Stirling, a une croissance approximative de l'ordre de K^N ; sous cette contrainte, le nombre de partitions en K classes devient égal à $\binom{N-1}{K-1}$, soit un nombre généralement beaucoup plus petit¹⁴.

Pour que l'algorithme puisse être appliqué, il faut que l'indice d'homogénéité W d'une partition P soit la somme des indices d'homogénéité des classes de cette partition P . Autrement dit, une bonne partition est une partition où la somme des "disparités" à l'intérieur de chaque classe de déclarations est minimale.

Définissons notre critère des "mal classés" de la manière suivante: pour une classe donnée, la présence majoritaire de "+" correspond à un patron idéal de classe composé uniquement de "+"; réciproquement, une présence majoritaire de "-" correspond à un patron idéal de classe composé exclusivement de "-". Le nombre de "mal classés" est la somme des discordances entre le patron idéal de la classe et sa composition actuelle. Dans le cas d'une classe composée d'un nombre pair d'éléments et où il y aurait égalité du nombre de "+" et de "-", le choix du patron idéal est indifférent, le nombre de discordances étant évidemment égal à la moitié du nombre des éléments de cette classe. Ce critère varie donc de 0 correspondant à une classe identique à son idéal, c'est-à-dire à une classe strictement homogène, jusqu'à 0,5 multiplié par le cardinal de la classe (si ce cardinal est pair) ou à 0,5 multiplié par le cardinal de la classe moins 1 (pour une classe composée d'un nombre impair d'éléments). Une classe composée d'un seul élément est toujours homogène, le critère valant dans ce cas toujours 0.

Si nous nous intéressons aux n_1 premiers objets de l'ensemble à classer, leur partition en k_1 classes est nécessairement composée d'une partition optimale de n_2 objets ($n_2 < n_1$) en $k_1 - 1$ classes et d'une classe supplémentaire composée des $n_1 - n_2$

¹³ Ou du moins l'une des meilleures; car il existe parfois plusieurs partitions en k classes aussi bonnes l'une que l'autre.

¹⁴ L'algorithme que nous allons utiliser est dérivé de l'algorithme de partition optimale de Fisher; on trouvera une explication plus complète de son fonctionnement dans G. Celeux et al 1989, p. 148 et suivantes, ainsi que dans J.-L. Chandon et S. Pinson 1980, p. 176 et suivantes.

d'homogénéité que procure la possibilité de disposer d'une classe supplémentaire pour répartir les objets (les déclarations gouvernementales). À *priori*, la partition en onze classes est idéale; il s'agit évidemment d'une solution triviale.

Peut-on retenir une partition en deux ou trois classes comme un bon modèle? La partition en deux zones fournit 10 éléments mal classés. Or, sur 38 éléments, le maximum possible d'éléments mal classés est de 19. La création d'une classe supplémentaire permet cependant de faire un gain appréciable en passant de dix à cinq mal classés, soit un gain absolu de cinq mal classés ou un gain relatif de $(10-5)/10$ soit 50%.

La distribution des gains relatifs nous incite à retenir la partition en trois zones comme un bon modèle pour la répartition des excès et des déficits des déclarations gouvernementales du vocabulaire de la classe 2 confirmant le diagnostic de chronologicité précédent. Le

Tableau 3 reprend les mêmes éléments que le Tableau 2 mais pour toutes les classes de la partition en six classes de vocabulaire.

Tableau 3

Partition optimale des déclarations en fonction de l'emploi du vocabulaire des six classes d'une partition du vocabulaire

		Classe observée		
n de classes	patron idéal	écart	gain absolu	gain relatif
Classe 1				
1	+++++-----+-----+-----+-----+-----+-----	17	-	-
2	-----+-----+-----+-----+-----+-----+-----	12	5	29,41%
3	+-----+-----+-----+-----+-----+-----+-----	11	1	8,33%
4	-----+-----+-----+-----+-----+-----+-----	10	1	9,09%
5	+-----+-----+-----+-----+-----+-----+-----	9	1	10,00%
7	+-----+-----+-----+-----+-----+-----+-----	8	1	11,11%
9	+-----+-----+-----+-----+-----+-----+-----	7	1	12,50%
11	+-----+-----+-----+-----+-----+-----+-----	6	1	14,29%
13	+-----+-----+-----+-----+-----+-----+-----	5	1	16,67%
15	+-----+-----+-----+-----+-----+-----+-----	4	1	20,00%
17	+-----+-----+-----+-----+-----+-----+-----	3	1	25,00%
19	+-----+-----+-----+-----+-----+-----+-----	2	1	33,33%
21	+-----+-----+-----+-----+-----+-----+-----	1	1	50,00%
23	+-----+-----+-----+-----+-----+-----+-----	0	1	100,00%
Classe 2				
1	-----+-----+-----+-----+-----+-----+-----	11	-	-
2	-----+-----+-----+-----+-----+-----+-----	10	1	9,09%
3	-----+-----+-----+-----+-----+-----+-----	5	5	50,00%
5	-----+-----+-----+-----+-----+-----+-----	3	2	40,00%
7	-----+-----+-----+-----+-----+-----+-----	2	1	33,33%
9	-----+-----+-----+-----+-----+-----+-----	1	1	50,00%
11	-----+-----+-----+-----+-----+-----+-----	0	1	100,00%
Classe 3				
1	-----+-----+-----+-----+-----+-----+-----	15	-	-
2	-----+-----+-----+-----+-----+-----+-----	2	13	86,67%
4	-----+-----+-----+-----+-----+-----+-----	1	1	50,00%
6	-----+-----+-----+-----+-----+-----+-----	0	1	100,00%
Classe 4				
1	-----+-----+-----+-----+-----+-----+-----	8	-	-
3	-----+-----+-----+-----+-----+-----+-----	3	5	62,50%
5	-----+-----+-----+-----+-----+-----+-----	2	1	33,33%
7	-----+-----+-----+-----+-----+-----+-----	1	1	50,00%
9	-----+-----+-----+-----+-----+-----+-----	0	1	100,00%
Classe 5				
1	-----+-----+-----+-----+-----+-----+-----	13	-	-
2	+++++-----+-----+-----+-----+-----+-----	4	9	69,23%
4	+-----+-----+-----+-----+-----+-----+-----	3	1	25,00%
6	+-----+-----+-----+-----+-----+-----+-----	2	1	33,33%
8	+-----+-----+-----+-----+-----+-----+-----	1	1	50,00%
10	+-----+-----+-----+-----+-----+-----+-----	0	1	100,00%
Classe 6				
1	-----+-----+-----+-----+-----+-----+-----	17	-	-
2	+++++-----+-----+-----+-----+-----+-----	11	6	35,29%
4	+++++-----+-----+-----+-----+-----+-----	8	3	27,27%
6	+++++-----+-----+-----+-----+-----+-----	7	1	12,50%
8	+++++-----+-----+-----+-----+-----+-----	6	1	14,29%
10	+++++-----+-----+-----+-----+-----+-----	5	1	16,67%
12	+++++-----+-----+-----+-----+-----+-----	4	1	20,00%
14	+++++-----+-----+-----+-----+-----+-----	3	1	25,00%
16	+++++-----+-----+-----+-----+-----+-----	2	1	33,33%
18	+++++-----+-----+-----+-----+-----+-----	1	1	50,00%

Découpage optimal de la séquence des sur- et sous-emplois: interprétation

Conclure à la "chronologicité" d'une classe de vocabulaire à partir du

Tableau 3 n'est pas chose aisée, du moins sans construction de règles. Plusieurs questions méritent d'être posées. La première est sans conteste la question de la classifiabilité des patrons de "+" et de "-" associés à chacune des classes de vocabulaire. Y a-t-il vraiment une structure dans ces séries? Un test classique de statistique non-paramétrique¹⁵ nous permettra de dire si la suite de "+" et "-" étudiée peut ou non être considérée raisonnablement comme non-aléatoire. Nous testons l'hypothèse que le nombre de séquences observé est inférieur à celui que l'on devrait attendre. Le Tableau 4 nous indique les probabilités d'observer sous l'hypothèse d'une répartition aléatoire des "+" et des "-" ce nombre de séquences. L'hypothèse que les séquences associées aux classes lexicales 1 et 6 ne sont pas significativement trop peu nombreuses ne peut pas être rejetée; par contre, si l'on se donne un niveau de signification de 5%, cette hypothèse doit être rejetée pour les séquences associées aux quatre autres classes. Celles-ci contiennent donc probablement un ordre parmi les "+" et les "-". Un diagnostic d'existence d'une structure des sur- et des sous-emplois de vocabulaire peut donc être tiré de ce test, condition nécessaire au diagnostic de chronologicité mais condition non suffisante.

Tableau 4

Test des séquences. Résultats

	r	n ₁	n ₂	N	z	P
1	23	17	21	38	1,07	14,23%
2	11	11	27	38	-2,27	1,20%
3	6	15	23	38	-4,53	0,00%
4	9	8	30	38	-2,32	1,00%
5	10	13	25	38	-2,97	0,15%
6	20	17	21	38	0,07	47,21%

• r est le nombre de séquences homogènes; n₁ est le nombre de "+" ; n₂ est le nombre de "-" ; N est la somme de n₁ et n₂ ;

• $z = \frac{r - \left(\frac{2n_1n_2}{n_1 + n_2} + 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}}$ suit en gros une distribution normale réduite ;

• P est la probabilité qu'une variable normale réduite dépasse les valeurs z calculées (inférieure ou égale à z si z est négative, supérieure ou égale à z si z est positive).

Cette utilisation d'un test de séquence, pour logique qu'elle apparaisse dans ce contexte d'analyse du signe des écarts d'emploi du vocabulaire, n'aboutit pourtant qu'à poser un diagnostic qui ne prend pas en compte tous les éléments que fournit la méthode de classification optimale utilisée. En effet, nous disposons, pour chacune des classes de la partition en six des écarts, de son "profil" par rapport aux meilleurs patrons idéaux composés de 1, 2, 3... r séquences homogènes. Pour juger de la qualité de ces écarts, nous devrions en connaître la distribution, ce qui permettrait de

¹⁵ En l'occurrence, le "one sample runs test", cf. Sidney Siegel 1956, p. 52 et suivantes.

savoir si les séquences de "+" et de "-" qu'il est possible de répartir en trois zones homogènes et contigües en n'ayant que trois éléments mal classés sont très nombreuses ou rares, soit parmi l'ensemble des séquences composées de trente-huit éléments soit parmi l'ensemble des séquences composées de trente-huit éléments et comportant treize "+". Nous ne connaissons pas ces distributions. Pour pouvoir néanmoins juger de la qualité des écarts, nous avons calculé par simulation six tables de la manière suivante:

1 - Nous avons engendré un échantillon de 100 000 séquences et cinq échantillons de 10 000 séquences de trente-huit éléments chacun par tirage aléatoire avec remplacement. Le premier de ces échantillons est composé de séquences dont l'écart initial à une séquence homogène unique n'est pas fixé. Les cinq autres échantillons sont respectivement composés de séquences comportant 17, 15, 13, 11 et 8 "+", ce qui correspond aux écarts initiaux des six classes lexicales étudiées.

2 - Pour chacune des séquences de ces six échantillons, nous avons établi les différentes partitions optimales en 1, 2, 3,...38 classes et nous avons calculé à chaque fois les écarts.

3 - Enfin, nous avons dressé six tables qui reprennent les répartitions des séquences des six échantillons selon la taille de la partition optimale et la valeur de l'écart à cette partition optimale¹⁶.

Nous pouvons maintenant établir quelle est la proportion de séquences qui peuvent se diviser en deux zones homogènes en ayant seulement 6 mal classés. Pour émettre un jugement sur les valeurs trouvées dans le corpus divisé en six classes lexicales, nous avons calculé la fréquence cumulée des différentes valeurs de mal classés pour chaque taille de partition optimale.

Il y a 31 séquences qui se partitionnent en deux avec six mal classées et 40 séquences qui se partitionnent en deux zones homogènes avec 6 ou moins d'éléments mal classés (31 + 5 + 4), ce qui conduit à une proportion de 40/100 000. Nous pouvons calculer cette proportion pour les valeurs observées dans la classification en six ensembles du vocabulaire, nous l'appellerons P_1 .

Nous pouvons calculer de la même façon une proportion P_2 , en nous servant, cette fois, non du tableau général engendré par l'échantillon des 100 000 mais des tableaux construits sur l'échantillon de séquences de même nombre de mal classés par rapport à un patron homogène. La comparaison de deux partitions optimales pour une même séquence peut désormais se faire facilement: on désignera comme la meilleure partition celle pour laquelle P_1 ou P_2 atteint son minimum, puisqu'ainsi sera mise en évidence la partition dont la "probabilité" d'apparition était la plus faible.

¹⁶ Pour les tables, cf. Deroubaix 1996.

Tableau 5

Évaluation de la qualité des différentes partitions optimales

N de classes	de	écart	P ₁	P ₂	N de classes	de	écart	P ₁	P ₂
Classe 1					Classe 4				
1		17	55,4	100	1		8	0,046	100
2		12	26,1	18,5	3		3	0,002	0,58
3		11	46,2	37,6	5		2	0,013	2,08
4		10	58,8	53,4	7		1	0,028	2,87
5		9	63,9	59,3	9		0	0,035	2,89
7		8	82,9	80,5	Classe 5				
9		7	89,2	87,8	1		13	6,4	100
11		6	90,3	88,9	2		4	0,004	0,01
13		5	90,3	88,9	4		3	0,019	0,04
15		4	90,3	88,9	6		2	0,074	0,14
17		3	90,3	88,9	8		1	0,112	0,24
19		2	90,3	88,9	10		0	0,121	0,28
21		1	90,3	88,9	Classe 6				
23		0	90,3	88,9	1		17	55,4	100
Classe 2					2		11	13,2	8,56
1		11	1,2	100	4		8	17,6	13,44
2		10	5,7	65,0	6		7	40,3	35,12
3		5	0,096	0,91	8		6	55,5	51,59
5		3	0,116	0,93	10		5	61,0	57,73
7		2	0,280	1,93	12		4	62,0	58,75
9		1	0,353	2,20	14		3	62,1	58,77
11		0	0,367	2,20	16		2	62,1	58,77
Classe 3					18		1	62,1	58,77
1		15	22,62	100	20		0	62,1	58,77
2		2	0	0					
4		1	0	0					
6		0	0	0					

Y a-t-il des classes "chronologiques"?

Les indices P₁ et P₂ sont strictement concordants, en ce qui concerne la pose du diagnostic: leurs minimums coïncident parfaitement dans cet exemple. Nous pouvons aller au-delà du diagnostic du test des séquences pour les classes lexicales 1 et 6. En effet, en ce qui concerne la classe 1, nous voyons que la performance de la recherche de partitions optimales n'est pas extraordinaire; la partition en deux zones semble la meilleure. Ces deux classes ne déclenchent pas de diagnostic de "chronologicit ". Par contre, la classe 2 en offre un bel exemple: elle se divise en trois zones, avec seulement 5 mal class s; seules quelques s quences de nos  chantillons ont pu faire la m me chose ou mieux (de l'ordre d'une sur dix mille – P₁ – ou une sur onze mille – P₂). Trois s quences homog nes, c' tait la condition pour pouvoir qualifier une s quence de chronologique. La classe 3 est plus  tonnante: la division en deux zones fournit une division quasi parfaite de la s quence (2 mal class s); dans nos  chantillons aucune s quence n'a fait aussi bien !

L'aspect chronologique réside dans le déficit d'emploi du vocabulaire de cette classe durant la première partie du corpus, suivi d'un suremploi quasi permanent dans la deuxième partie. Au même type que la classe 2 appartient la classe 4 mais avec encore plus de netteté, les indices P_1 et P_2 étant très faibles pour la division optimale en trois zones. Enfin, la classe 5 se situe clairement dans le type de la classe 3: division en deux zones homogènes mais avec suremploi dans la première partie du corpus et sous-emploi localisé dans la fin.

Tableau 6

Diagnostic de "chronogicité"

	Chronogicité	Périodes
Classe 1	non	-
Classe 2	oui	dg00-16(-) dg17-23 (+) dg24-38 (-)
Classe 3	oui	dg00-22(-) dg23-38 (+)
Classe 4	oui	dg00-11(+) dg12-16 (-) dg17-38 (+)
Classe 5	oui	dg00-12(+) dg13-38 (-)
Classe 6	non	-

Étude "chronologique" de la partition du vocabulaire en six classes

Organisée autour des formes *et, de, la, à, l, les, dans, le, des, d, en, au, un, Gouvernement, une, du, s, par, sur, son, avec, pour, politique, cette, plus, économique, ce, aux, pays, ces*, la première classe se trouve ne pas montrer de caractère chronologique; la liste des formes ci-dessus constituée par les parangons de la classe¹⁷, indique pourquoi. En effet, outre les morphèmes grammaticaux qui sont utilisés par tout texte écrit en français (*de, le, à, un...*), on retrouve un vocabulaire que nous pourrions considérer comme la marque générale du discours politique gouvernemental (*Gouvernement, politique, État, Belgique, Parlement, mesures, économique...*). Cette classe est celle qui contient le plus de formes (371) et qui couvre la plus grande surface du corpus en terme d'occurrences (64,51% des formes de fréquence supérieure à 9). Elle fournit la toile de fond lexicale sur laquelle vont pouvoir se déployer les discours particuliers de chaque gouvernement. Toutes les déclarations gouvernementales ne recourent pas dans les mêmes proportions à ce vocabulaire. De manière assez étonnante, en ordonnant les déclarations gouvernementales suivant leur recours plus ou moins grand au vocabulaire banal des déclarations gouvernementales, apparait un ordonnancement droite/gauche. En

¹⁷ Les parangons d'une classe sont simplement les éléments de cette classe qui sont les plus proches du centre de la classe. On peut considérer qu'ils en sont les éléments les plus représentatifs.

effet, tous les gouvernements à direction socialiste (à l'exception du premier gouvernement Leburton de 1973) font un emploi déficitaire du vocabulaire banal. Le vocabulaire de cette classe n'est pas composé uniquement de "mots-outils" communs à la plupart des usagers de la langue mais aussi de nombreux mots appartenant à la sphère de la politique gouvernementale, les deux sous-ensembles fournissant le ciment de toute déclaration gouvernementale en Belgique dans la période considérée.

Le caractère chronologique de la deuxième classe se marque par un excès d'emploi des formes de cette classe dans une période qui court de la déclaration de Pierre Harmel (dg 17, 1965) à celle de Léo Tindemans (dg 23, 1974). Cette période se caractérise par l'utilisation d'un vocabulaire de réorganisation de l'État, sans qu'il soit encore vraiment question du démembrement de l'État mais plutôt d'un début d'autonomie culturelle et linguistique (*autonomie, conseils, cultures, langue...*), de la création d'organes régionaux (*agglomération*). Le vocabulaire de cette classe est encore utilisé en dehors de cette période, mais surtout dans des textes qui reviennent sur la réorganisation de la Belgique et particulièrement dans la déclaration de Léo Tindemans en 1974, quand, pour la première fois, sont associés à l'équipe ministérielle des ministres membres de partis "communautaires" ou dans la déclaration de Wilfried Martens de 1980 quasi entièrement consacrée à l'exposé d'une réforme transitoire de l'État. Le processus de réforme de la Belgique trouve ainsi à s'exprimer à travers un vocabulaire spécifique propre, qui émerge vers 1964, est employé de manière dense et constante de 1964 à 1974, puis revient sporadiquement à chaque nouvelle crise institutionnelle.

Parangons de la classe 2 classés par distance croissante du centre de la classe: *accord, conseil, Constitution, tenu, conseils, souci, culturel, article, élaborer, réalisations, matières, pris, autonomie, juillet, doute, perspective, mener, dernier, régionaux, parlementaire, ministres, concernant, culturels, conformément, culturelle, permet, agglomération, arrêté, partenaires.*

La classe 3 représente un type chronologique différent. En effet, l'utilisation des mots de cette classe découpe le corpus en deux zones, une période de sous-utilisation entre 1944 et 1973, une période d'utilisation en excès à partir de cette dernière année. Vocabulaire de la crise, de la crise de l'énergie, de la crise de l'emploi, de la crise budgétaire, les formes de cette classe montrent aussi la montée en puissance des problématiques de *sécurité*: sécurité publique et sécurité sociale. Le modèle de société basé sur la croissance et ce qu'il est convenu d'appeler le compromis "fordiste", c'est-à-dire un système de régulation sociale¹⁸ qui permet la redistribution d'une partie de la richesse sous forme de la Sécurité sociale et qui s'était développé grosso modo entre 1944 et 1972, se trouvent remis en cause au nom de la recherche de solutions au problème du chômage. On voit l'emploi devenir un thème important mais couplé à celui de la recherche d'accomplissement d'objectifs budgétaires et d'équilibre de la Sécurité sociale.

À ces problématiques centrales de l'après 1974, sont associés des ensembles de préoccupations qui sont dérivées directement de la "crise" énergétique comme la prise en compte de *l'environnement*, ou bien sont simplement le développement de

¹⁸ R. Boyer, *La flexibilité du travail en Europe*, Paris, La Découverte, 1986.

politiques longuement préparées comme l'intégration européenne, qui finit par fournir en fin de période un cadre aux préoccupations de libéralisation et d'affaiblissement de l'intervention publique dans le domaine social et économique. Le tout constitue le vocabulaire commun de la déclaration gouvernementale de l'après-crise pétrolière.

Parangons de la classe 3: *sociale, emploi, cadre, fiscale, grande, population, nouveau, coopération, également, confiance, base, niveau, européenne, entreprises, budgétaire, objectif, important, outre, publiques, Sécurité, réforme, promotion, manière, meilleure, solidarité, rôle, chacun, nombre, sécurité, décisions.*

La classe 4 est d'un type de classe "chronologique" particulier: le type négatif. La période mise en évidence (de 1954 à 1961) est caractérisée par un déficit d'utilisation des mots de cette classe. Il s'agit donc d'un vocabulaire qui, en usage au début de la période couverte par le corpus, subit une éclipse puis revient en excès dans les déclarations d'après 1965. On remarquera la dominance du lexique économique, planificateur (*programme*) et social. La période où se marque un déficit de ces formes est assez dominée par le parti libéral, qui est présent dans trois cabinets sur cinq ! Ici aussi, on constatera le retard du discours sur le travail politique: la coalition "travailliste" Lefèvre-Spaak de 1961¹⁹, mise en place après la grande grève de l'hiver 1960-1961, quoiqu'elle ouvre une nouvelle période politique, celle de la gestion de la croissance, conforme son discours à ceux des gouvernements précédents, y compris dans ses déficits (*travailleurs, social*).

Parangons de la classe 4: *secteur, part, national, programme, objectifs, autre, action, travailleurs, efficacité, scientifique, secteurs, investissements, augmentation, social, indépendants, budget, structure, impose, responsabilités, marché, hommes, assurance, dépenses, ans, agricole, étroite, ressources, pension, méthodes, organismes.*

Vocabulaire de remise en route d'une société et d'un État profondément perturbés par le traumatisme de la guerre, l'ensemble des formes de la cinquième classe renvoie au processus de normalisation de la situation institutionnelle, de reprise de la production et de mise en place d'une organisation sociale. Elles ont donc été utilisées préférentiellement durant la période de reconstruction de l'immédiat après-guerre et plus encore dans la période de stabilisation des institutions qui la suivit.

Le statut de l'enseignement (*enseignement, technique*) qui, avec la question royale, sera l'un des problèmes les plus aigus de la vie politique de cette période, se retrouve tout aussi naturellement inclus dans cette classe de vocabulaire.

Parangons de la classe 5: *nationale, toute, équilibre, situation, œuvre, conditions, intérêt, prises, organisation, fonctionnement, moyennes, poursuivra, services, technique, nécessaires, cours, loi, statut, enseignement, agriculture, étude, production, point, Chambres, améliorer, législation, fonds, financier, veillera, générale.*

Enfin, il reste une sixième classe de vocabulaire qui n'a pas déclenché de diagnostic de "chronogénéité" dans le système de détection. La lecture du tableau

¹⁹ Cabinet social chrétien/socialiste, certains journalistes appellent cette coalition "rouge-romaine".

des parangons de cette classe tendrait à conforter ce diagnostic: on y trouve, pour l'essentiel, des mots dont on attend l'utilisation dans tout discours (*que, il, est...*).

Cependant, la distribution de ces formes incite à nuancer ce diagnostic. À l'exception des dg 4 et dg 5, nous avons un excédent d'emploi bien marqué durant les années 1944 à 1949 (près de 30% de la surface des textes sont des occurrences de formes de cette classe); on ne retrouve plus ensuite une telle concentration, même si des "pics" se présentent encore. Nous rencontrons là une limite de notre méthode de diagnostic. Celle-ci est peut-être trop rude, à ne retenir comme information pertinente que le signe de l'écart à la moyenne et à ne pas prendre en compte la valeur de cet écart. Limite et non contre-indication car, si l'on sent bien qu'il y a dans cette classe de vocabulaire quelque chose de "chronologique", on perçoit encore que cela ne suffit pas à rendre compte du pourquoi de la constitution de cette classe intermédiaire entre les "chronologiques" vraies, où les excédents et les déficits sont clairement le fait de périodes déterminées, et la classe 1, qui regroupe le vocabulaire bien réparti dans le corpus. Nous sommes face à un ensemble de formes qui combine deux types d'utilisation: un emploi commun à tout le corpus, certains des mots de cette liste étant difficilement évitables dans un texte politique, et un sur-emploi dans certains discours, particulièrement au début du corpus. Le mélange des deux types d'emploi brouille le caractère chronologique que pourrait avoir le deuxième type d'emploi. Il s'agit d'un cas limite.

Parangons de la classe 6: *que, il, est, qui, qu, n, se, pas, ne, y, sans, même, être, mais, ont, problèmes, tout, a, effort, elle, toutes, présente, faire, fois, fait, si, elles, c, aujourd'hui, peut.*

Périodisation?

En somme, le découpage chronologique du corpus fait sens, mais il ne débouche pas vraiment sur une périodisation simple. C'est le glissement régulier vers de nouveaux vocabulaires, associés à de nouveaux problèmes politiques, qui engendre l'image chronologique vue sur le plan factoriel, et non une stricte périodisation.

Les périodes que l'on peut définir à partir du renouvellement du vocabulaire se chevauchent et reflètent la multiplicité des sources du changement politique. Du point de vue des gouvernants, la politique internationale, l'économie, les tensions sous-nationalistes, etc., chaque domaine apporte son propre renouvellement lexical à un rythme qui lui est propre. Certes la tâche d'un gouvernement consiste à unifier ces discours (et les actions qui en découlent) en un ensemble politique cohérent, et c'est, en théorie, à cela que devrait servir l'idéologie partisane. Mais si l'on se veut le "président de tous les Français" ou plus simplement le Gouvernement de quatre, cinq ou six partis belges à programmes divergents, cette tâche devient difficile. Le discours des gouvernants se présente alors comme consensus et non comme rupture franche avec les prédécesseurs, du moins dans les corpus de discours de régimes démocratiques. Les contraintes de l'action gouvernementale, l'inertie de

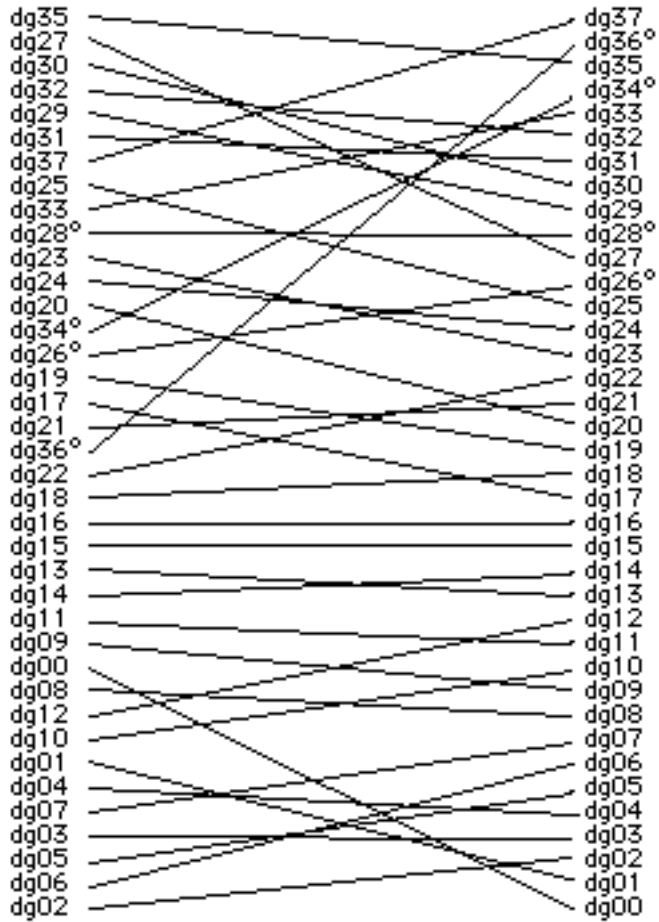
l'administration d'un État contribuent pour leur part au lissage de l'évolution. On le voit, la composante chronologique est la résultante de forces multiples.

Donc, le temps - on devrait dire plus proprement les diverses temporalités politiques - structure de manière serrée le discours gouvernemental, mais il le fait en conservant un certain flou, sous lequel transparaissent les préoccupations les plus politiques, flou qui nous autorise à trouver des ressemblances entre les discours contemporains de la fatalité marchande et ceux tenus par les gouvernements catholiques des premières années de la guerre froide. Par leur production à l'avant-scène politique de préoccupations budgétaires qui imposent une discipline à l'ensemble de la population mais surtout à ceux qui pouvaient bénéficier de la Sécurité sociale, les discours de ces deux périodes ont en commun une certaine tonalité conservatrice qui perturbe le simple déroulement de la ressemblance chronologique. De même, il importe de remarquer la similitude de tonalité entre les discours volontaristes du redémarrage de l'immédiat après-guerre et ceux du redémarrage économique et surtout sociopolitique des années 1960 et 1975 [Deroubaix 1997, Deroubaix-Gobin 1989].

La cité est l'œuvre de l'imaginaire social des hommes. Il serait vain de penser que cette production de la société, et des discours qui en sont inséparables, puisse un jour être résumée par l'étude d'une courbe, d'une série chronologique, serait-ce même celle du vocabulaire de ceux que l'on désigne erronément par "nos" gouvernants.

Annexes

Tableau 7. Ordre selon l'axe1 et ordre "naturel" chronologique.



- Déclarations mises en éléments supplémentaires

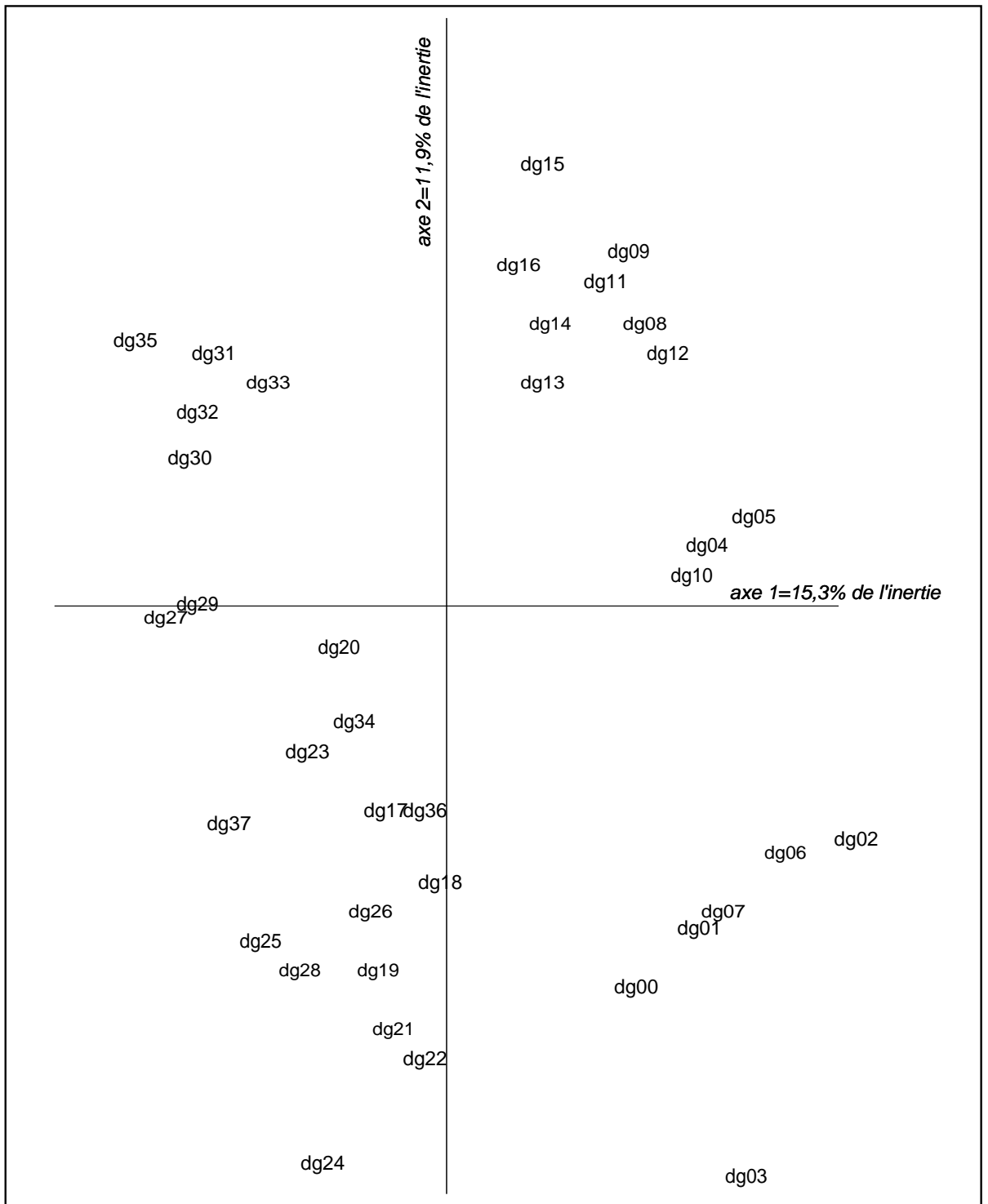


Figure 1. Plan 1 & 2 représentation des déclarations et "cubique" approchée

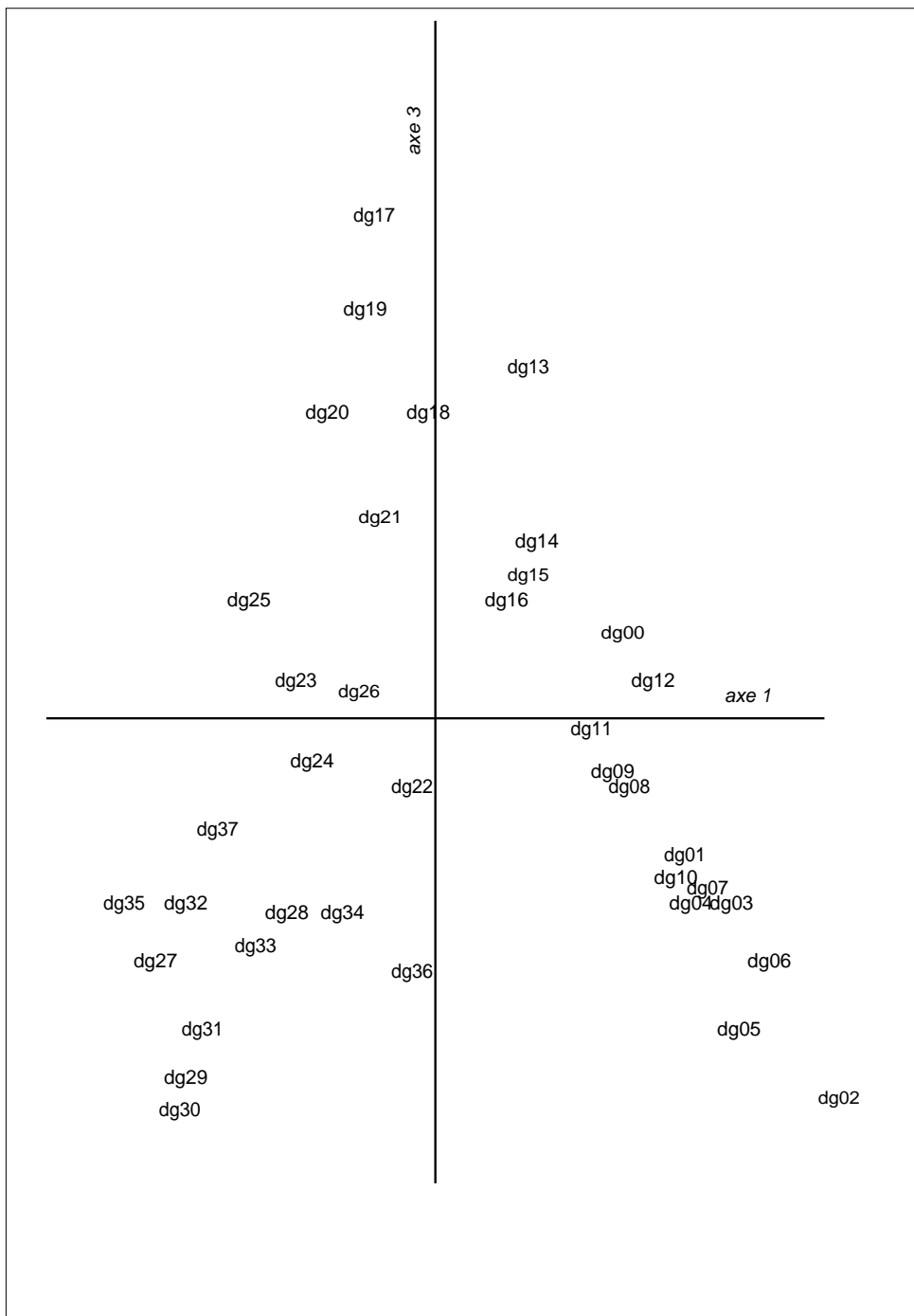


Figure 2. Plan 1 &3 représentation des déclarations et parabole approchée