Lexicométrie textuelle, sens et stratégie discursive

Serge Heiden, Maurice Tournier

LEXICOMETRIE TEXTUELLE, SENS ET STRATEGIE DISCURSIVE

Serge HEIDEN, Maurice TOURNIER

Laboratoire de lexicométrie politique
INaLF-CNRS, ENS, Saint-Cloud, France

Résumé: A partir du corpus des "face-à-faces" télévisés de deux élections présidentielles françaises, l'analyse lexicométrique des couples de mots en cooccurrence et ses représentations graphiques par ordinateur mettent en regard les principaux liens statistiques qui affectent les vocabulaires des quatre candidats (Mitterrand contre Chirac en 1988, puis Chirac contre Jospin en 1995). Leur comparaison montre l'importance de la stratégie discursive dans le choix et la mise en contexte des mots au sein des phrases.

Mots-clés: vocabulaire, texte politique, élections présidentielles 1988 1995, lexicométrie, cooccurrences, stratégie discursive

"Les mots n'ont pas de sens; ils n'ont que des emplois". La boutade de Wittgenstein pourrait mener au sens, si l'on connaissait pour un terme l'ensemble de ses emplois. Mais cet ensemble n'existe pas. Car les mots dépendent des "situations" d'emploi, causes et conditions de leur énonciation, et l'on sait bien que celles-ci varient à l'infini, dans le temps comme dans l'espace. La seule solution consiste à se saisir de l'une de ces "situations" où du texte s'énonce, précise, nettement circonscrite, datée, maitrisable dans ses aspects essentiels, puis à faire, à son propos, la somme des paroles à étudier, afin de comparer tous les emplois constatables, en entendant par là, en premier, les immersions des mots dans leur contexte proche. De cette somme, de cette exploration et de cette comparaison exhaustive des contextes, la recherche statistique des cooccurrences peut tirer une description objective. Nous met-elle sur une piste qui conduise au sens ?

Notre hypothèse de recherche le présuppose: une analyse lexicométrique des cooccurrences (dis-moi qui tu fréquentes...) construirait une image représentable et

peut-être représentative de l'usage fait des mots. Les liens statistiques seraient une résultante de surface des jeux de rapports à l'œuvre dans le profond du discours. Mais la recherche de ces liens (cooccurrences) exige un corpus de données textuelles homogène, c'est-à-dire dépendant de forts invariants énonciatifs. L'interprétation, nourrie de la compétence de situation propre au chercheur, fait ensuite le reste, dans la mesure où les constats sont suffisamment répétés et pertinents.

Nous avons choisi, pour donner une idée de la méthode utilisée, à savoir l'analyse des cooccurrences élaborée à Saint-Cloud, un exemple, certes trop restreint mais nouveau dans nos recherches : l'émission "Face à face", qui détient depuis plusieurs septennats une position-clé dans les élections présidentielles françaises. Nos quatre émetteurs, car il s'agit des deux duels télévisés opposant un candidat de gauche à un candidat de droite, s'appellent en 1988 François Mitterrand (texte de 9593 occurrences) et Jacques Chirac (9567 occ.), puis, sept ans plus tard, en 1995 Lionel Jospin (11836 occ.) et à nouveau Jacques Chirac (9917 occ.). Les invariants situationnels sont nombreux à garantir l'homogénéité du corpus : émission publique à une heure de grande écoute, placée entre les premier et le second tours de l'élection, où les deux candidats restés en lice, "face à face" mais séparés par des journalistes questionneurs, ajustent leur réponse - en disposant d'un temps identique de parole - sur les thèmes lancés par les journalistes (et préalablement choisis en accord avec eux). L'enjeu, on le sait, n'est rien moins que le vote définitif du dimanche suivant, qui les départagera. Le discours des présidentiables n'a pour finalité que de convaincre l'électorat-téléspectateur, dans un affrontement verbal où chaque mot et lien entre mots peuvent compter et se compter.

Nous allons donc les prendre aux mots, mieux encore aux mots des mots, puisque tel est l'objet de l'analyse des cooccurrences.

1- La méthode et l'outil d'analyse (1)

Une fois réalisés la partition du corpus à l'aide d'un codage des émetteurs et des dates, une opération de segmention du texte permet de repérer les unités en présence (**formes** graphiques entre délimiteurs) et de baliser les **phrases** (intervalles entre deux ponctuations "fortes"). C'est sur cette double segmentation et sur les fréquences constatées pour chaque forme dans chaque phrase de chaque émetteur que va s'appliquer le calcul des cooccurrences.

Mais, avant de procéder à ce calcul, une opération d'élagage (paramétrable, elle aussi) retire du vocabulaire certaines graphies jugées parasitaires ou de moindre intérêt pour l'analyse : ponctuations et autres signes non alphabétiques, mots-outils, chiffres, hapax, etc. L'esprit de cet élagage est de restreindre les calculs effectifs à une population réduite de formes considérées a priori comme "lexicales". Par contre, les calculs de la probabilité de cooccurrence théorique sont effectués, eux, sur l'ensemble des occurrences alpha-numériques, qui "fabriquent" le texte.

Nous avons choisi la phrase comme espace séquentiel de rencontre des formes lexicales. Pour chaque forme donnée, nous dressons la liste des formes apparaissant à sa gauche ou à sa droite à l'intérieur de cet espace, ce qui donne un nombre de rencontres en couple appelé **co-fréquence** (CF). Sur la base de la Fréquence totale F d'une forme A (F_A) et sur celle d'une forme B (F_B), de leur cofréquence (CF_{AB}) et du nombre de phrases composant le texte analysé (P), nous calculons une estimation de la **probabilité** a priori (p) (voir Lafon, 1984:129-199)(1) que ces deux formes se rencontrent au moins le nombre de fois que l'on constate effectivement dans le texte, ainsi que la **distance moyenne** (d_m), ou nombre d'occurrences interposées, entre les occurrences de A et celles de B dans l'espace des phrases où A précède B.

L'affichage des résultats est assorti d'un nouvel élagage (toujours paramétrable), qui élimine les couples ayant certaines propriétés liées au calcul :
- soit une probabilité de cooccurrence trop importante (supérieure, dans la présente expérience, à p = 5% ou 5.0e-2),
- soit une cofréquence CF trop faible (élagage plus grossier que sur p),
- soit une fréquence F trop faible pour l'une des formes (élagage encore plus grossier)
- soit une distance moyenne d_m trop importante ou trop faible, selon que l'on s'intéresse aux couples les plus soudés ou les plus éloignés (variable indépendante de p).
L'esprit de cet élagage est de limiter le volume des données affichées et de faciliter le parcours et la lecture dans l'espace de cooccurrence construit, le but étant d'établir une liste homogène de couples en forte "attirance".

Nous proposons quatre moyens de visualiser ces attirances.

1) Liste des couples sélectionnés.

Nous produisons la liste de tous les couples de formes en cooccurrence, avec un couple par ligne présentant le cooccurrent de gauche suivi par celui de droite (selon la séquence de phrase), leur fréquence respective dans le texte, leur cofréquence, la probabilité théorique liée à cette cofréquence, enfin la distance moyenne qui les sépare dans les phrases du texte. Le nombre de couples composant la liste dépend des seuils d'élagage établis à l'entrée. Elle peut s'ordonner selon divers paramètres : l'ordre inverse des probabilités, celui des cofréquences, l'ordre hiérarchique des fréquences, l'ordre alphabétique des cooccurrents gauches puis celui des cooccurrents droits, ou de la distance moyenne.

A	B	F_A	F_B	CF	p	d_m	A	B	F_A	F_B	CF	p	d_m
clin	oeil	5	5	5	1.4e-13	1.0	essais	nucléaires	10	10	8	2.2e-12	2.1
permettez	moi	15	129	12	4.3e-13	0.0	services	publics	9	9	6	2.7e-12	0.0
Giscard	D-Estaing	8	8	8	4.8e-13	0.0	impartialité	Etat	7	46	7	2.7e-12	3.3
général	De-Gaulle	23	13	12	5.3e-13	0.0	Georges	Pompidou	6	7	5	3.4e-12	0.0
deux	ans	85	71	25	6.9e-13	0.1	Jacques	Chirac	29	53	26	4.8e-12	0.0
durée	travail	18	33	11	7.4e-13	2.4	petites	moyennes	9	6	5	5.0e-12	1.0
partage	sentiment	5	10	5	8.6e-13	2.8	mètres	carrés	4	4	4	5.1e-12	0.0
compte	tenu	37	15	12	1.0e-12	0.0	crimes	délits	4	4	4	5.1e-12	1.0
hommes	femmes	20	13	9	1.2e-12	2.0	veut	dire	26	153	19	5.5e-12	0.1
nous	devons	290	13	13	1.2e-12	0.2	premier	ministre	63	55	42	5.7e-12	0.0
cinq	ans	17	71	12	1.5e-12	0.0	permettez	dire	15	153	12	5.7e-12	4.1
président	république	32	42	21	1.8e-12	2.4	monsieur	Jospin	177	35	33	6.7e-12	0.7
service	public	19	15	8	1.9e-12	0.0	monsieur	ministre	177	55	24	6.8e-12	3.3
Sécurité	sociale	8	27	8	1.9e-12	0.0	voudrais	dire	40	153	20	7.1e-12	1.8
immigration	clandestine	17	9	9	2.2e-12	0.1	monsieur	Mitterrand	177	47	45	7.5e-12	0.0

Figure 1. Liste des 30 premiers couples de formes en cooccurrence dans l'ensemble du corpus hiérarchisés dans l'ordre inverse de la probabilité (341 couples ont été sélectionnés avec F et CF > 3 sur un total de 45843 analysés).On observe que le haut de la liste est occupé par des formants de lexies ou de stéréotypes.

2) Lexicogramme direct associé à un "pôle".

A gauche et à droite au-dessous d'une forme choisie, appelée "pôle", nous présentons deux ensembles de formes, celui de ses cooccurrents à gauche et celui de ses cooccurrents à droite. Chaque ligne comprend par deux fois des informations de même

type : la forme cooccurrente, sa fréquence totale F, sa cofréquence avec le pôle CF, la probabilité de cette cofréquence p, la distance moyenne entre cooccurrents d_m. Le nombre de cooccurrents à gauche et à droite dépend des seuils d'élagage imposés en sortie, au niveau de chaque paramètre, et peut aussi se limiter en nombre. Leur hiérarchisation dépend, comme pour la liste, du paramètre privilégié.

Français
(40)

cooccurrents gauches					cooccurrents droits				
	F	**CF**	**p**	**d_m**		**F**	**CF**	**p**	**d_m**
je	542	23	7.258e-04	9.5	attendent	3	3	3.949e-05	3.7
élu	10	3	3.924e-03	11.0	éliront	2	2	1.191e-03	5.5
dire	86	8	4.999e-03	13.8	restent	2	2	1.191e-03	10.5
confiance	4	2	6.810e-03	1.0	septennat	2	2	1.191e-03	11.0
eh-bien	13	3	8.634e-03	15.7	veulent	7	3	1.240e-03	0.3
vient	5	2	1.108e-02	7.0	république	27	5	1.679e-03	14.0
autrement	6	2	1.623e-02	12.5	favorables	3	2	3.488e-03	10.0
esprit	8	2	2.889e-02	1.0	président	24	4	7.664e-03	9.0
vais	9	2	3.628e-02	14.5	proposition	5	2	1.108e-02	6.5
voudrais	22	3	3.673e-02	15.7	lieu	6	2	1.623e-02	13.0
président	24	3	4.584e-02	15.0	chances	7	2	2.219e-02	5.5
					mandat	7	2	2.219e-02	25.0
					réforme	9	2	3.628e-02	2.5

Figure 2. Lexicogramme de Français *dans le face-à-face de 1995 (2). (F et CF ≥ 2, p ≤ 5.0E-2)*

3) Lexicogramme récursif des connexions associées à une "source".

L'opération de recherche des cooccurrents peut être récursive, c'est-à-dire reproduite sur chaque cooccurrent pris à son tour pour pôle(3). L'hypothèse se redouble ainsi: un pôle "attire", pour entrer dans le sens, des voisins de phrase, qui eux-mêmes, pour entrer dans le sens, attirent leurs propres voisins. Au lieu de s'arrêter au premier niveau de cooccurrence, pris dans les voisinages immédiats du pôle de départ (appelé "source"), une itération poursuit la recherche des cooccurrents, passant d'un pôle à l'autre. Car la source, en fait, n'appelle pas autour d'elle des formes isolées mais des systèmes secondaires de cooccurrence. Pour rendre compte de cette connexion entre systèmes, le calcul obéit de nouveau à une série de règles décidées par le chercheur en fonction de son objectif, ces paramètres d'itération ne changeant pas durant tout le temps d'une expérience. Se construisent alors des chemins, des cycles, des impasses qui constituent un graphe connexe jusqu'à saturation(4).

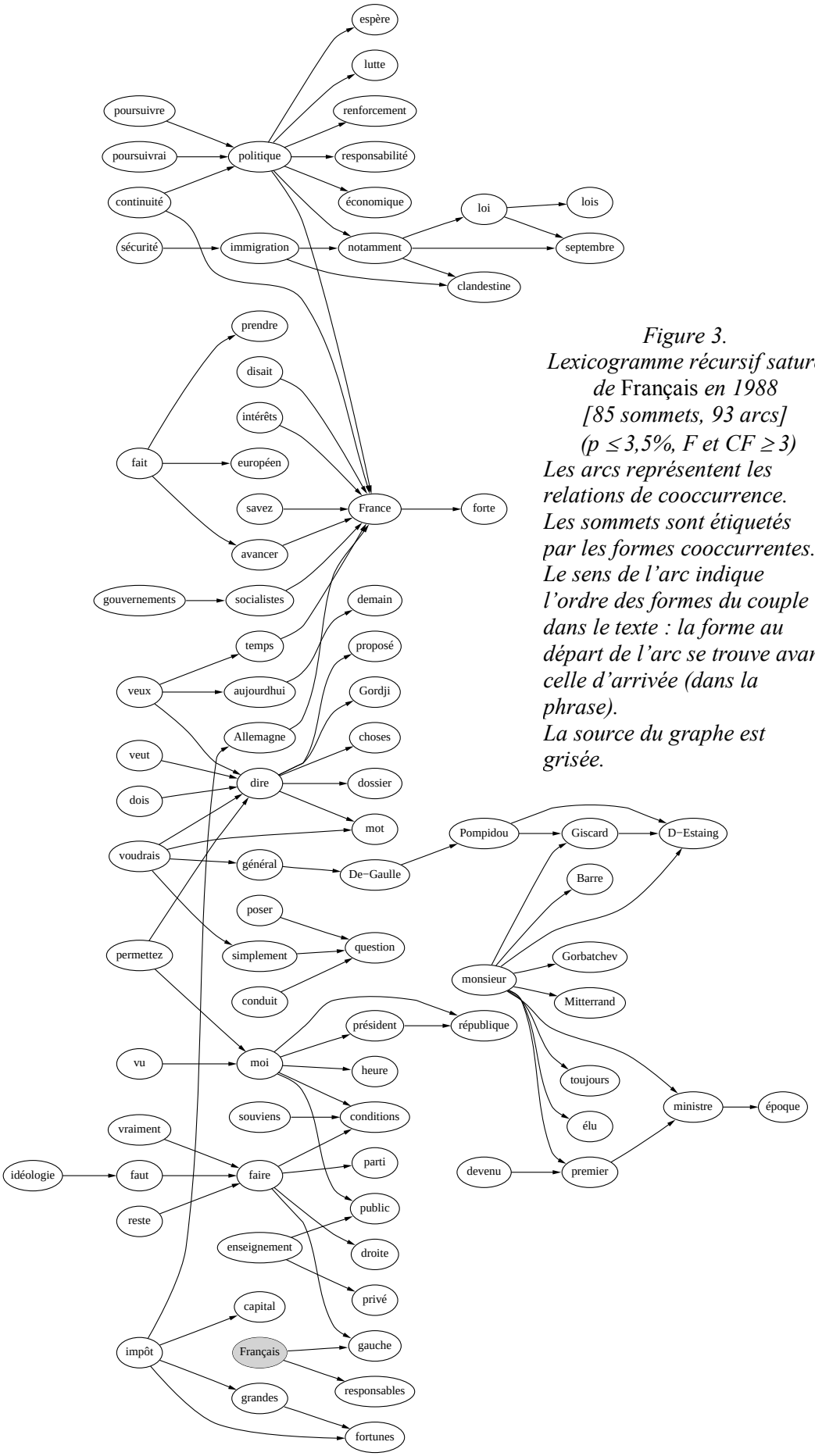

Figure 3.
Lexicogramme récursif saturé de Français *en 1988*
[85 sommets, 93 arcs]
($p \leq 3,5\%$, F et CF ≥ 3)
Les arcs représentent les relations de cooccurrence. Les sommets sont étiquetés par les formes cooccurrentes. Le sens de l'arc indique l'ordre des formes du couple dans le texte : la forme au départ de l'arc se trouve avant celle d'arrivée (dans la phrase).
La source du graphe est grisée.

4) Lexicogrammes récursifs non associés à une source.

L'ordinateur, possédant l'ensemble des données et des paramètres, est capable de fournir à la demande et automatiquement tous les graphes de connexions, quels qu'en soient les sommets. Peut se construire ainsi, selon le niveau d'exigence des règles de construction, le paysage mouvant d'un vocabulaire organisé en réseaux plus ou moins connexes, qui révèlent une certaine structure statistique interne aux textes.

Ces formes de visualisation dévoilent le système général des "attirances" statistiques entre unités lexicales. De ces attirances aux rapports d'habitudes et de ces rapports à la stratégie discursive, voire à la sémantisation, peut-être n'y a-t-il qu'un pas... Dans la présente recherche, nous n'évoquerons que des représentations des types 2 et 3. Notre hypothèse d'interprétation se retrouve face aux lexicogrammes directs des pôles *France* et *Français* et au lexicogramme récursif de la première personne (*je, j'*), que nous allons prendre pour objets de commentaires.

2- Le face-à-face des emplois

Le duel des présidentiables peut se jouer à quatre (les émetteurs, dont les prestations tournent chacune autour des 10.000 occurrences) ou à deux (les deux émissions à sept ans de distance) - le texte des journalistes étant, lui, provisoirement occulté. Dans les deux cas, le pôle de départ des lexicogrammes sera choisi parmi les formes du texte à la fois les plus fréquentes et les mieux réparties. Présentons d'abord le résultat de lexicogrammes de *France* et de *Français,* pôles qui représentent ce dont on parle et ceux à qui l'on parle, avant de nous intéresser au parleur lui-même.

1) Cooccurrents de *France* et *Français*

La ventilation des occurrences de ces deux formes entre les quatre émetteurs montre une distribution relativement équilibrée. Seul le nom *France* s'y découvre un petit peu suremployé par Mitterrand et sousemployé par Jospin. Tous les autres emplois sont de type banal, c'est-à-dire à probabilité supérieure à 5%(5). Les occurrences des formes de l'adjectif ne changent pas la physionomie du tableau:

	Corpus 40913 occ	Mitterrand 9593 occ	Chirac1 9567 occ	Chirac2 9917 occ	Jospin 11836 occ
France	93	29 (+2)	21	26	17 (-2)
Français	81	20	21	18	22
Française	1	1	0	0	0
Françaises	2	1	0	1	0
français	22	7	7	5	3 (-2)
française	11	6 (+2)	1	1	3
françaises	4	3 (+2)	0	1	0
totaux	214	67	50	52	45

Figure 4. Ventilation des sous fréquences du champ lexical France.

NB: En ce qui concerne la nomination directe de la France, au surenchérissement de Mitterrand s'oppose, chez Jospin - bien que celui-ci dépasse son "modèle" de 2243 occurrences - un déficit de 12 occurrences. Nous ne nous occuperons ici que des noms fréquents, *France* et *Français* (avec majuscule initiale).

Aux règles communes suivantes : F et CF >= 2, p < 5%, classement : dans l'ordre inverse de la probabilité, les lexicogrammes de *France* (les duellistes de chaque émission étant réunis) dégagent en 1988: 17 cooccurrents à gauche contre 9 à droite, et en 1995 14 cooccurrents à gauche contre 7 à droite. Ce déséquilibre entre les expansions gauches et droites est d'autant plus évident qu'il se produit dans cinq lexicogrammes sur six : cela signifie que la forme *France* a tendance, dans les discours étudiés, à se situer dans la seconde partie de la phrase; elle n'a donc pas pour fonction habituelle de "commander" la phrase ; elle représente plutôt un point d'aboutissement. Le recours aux contextes le montre : "... confiance en la France." clot les propos de J. Chirac.
Il en est tout à fait autrement de *Français*. Un déséquilibre en nombre de cooccurrents existe aussi pour ce pôle, mais cette fois au profit des expansions droites, dans 5 lexicogrammes sur 6, et beaucoup plus fortement en 1988 (5 cooc. contre 20) qu'en 1995 (11 contre 13). *Français* joue souvent les rôles de sujet ou de terme d'adresse, positionnés plutôt en tête. Les contextes là-aussi le montrent : "les Français" est 11 fois sur 22 sujet de verbe dans le texte de Jospin.

Sur le plan "sémantique", on remarque une égale situation au sein de l'époque : malgré la présence d'*histoire* chez Mitterrand, *France* attire dans ses voisinages : *maintenant, moment, aujourd'hui*, *temps*, *nouvelle, nouvelles, heure,* un certain vocabulaire du projet, avec : *avancer, continuité, politique, engagé, volonté, mesures, souhaite, pense, fera, veux, propositions, faudrait,* enfin un souci international permanent, avec : en 1988, *Allemagne, voix, concurrents, concurrence, Europe* et, en 1995, *Forpronu, pays, simulation, nucléaires.* L'économie n'est pas absente des lexicogrammes mais, alors qu'en 1988 il s'agit, surtout chez Mitterrand, d'*entreprises,* d'*intérêts* et d'*affaires,* ainsi que de l'*Europe* "unie et forte", en 1995 c'est de problèmes concrets, d'aménagement urbain, par exemple, dont parlent les deux candidats (*bureaux, mètres carrés, villes, logements*).

En ce qui concerne *Français,* les liaisons avec les mots porteurs de problèmes internationaux ou économiques disparaissent au profit du vocabulaire proprement politique de prise à témoin ou d'appel au rassemblement de l'électorat. On trouve, particulièrement à droite de ce pôle, en 1988 : *responsables, gauche, cohésion, confiance, droite, ministre, majorité, propositions,* et en 1995 : *confiance, élu, président, réforme, chance, proposition, mandat, septennat, république, pouvoir, changement,* à quoi s'ajoutent, selon l'émetteur, à gauche du pôle des verbes comme : *crois, voudrais, souhaite,* et à sa droite : *attendent, éliront, restent, veulent, peuvent, voulez, (sont) sensibles, favorables...* La permanence des mêmes mots, dans ce domaine, semble ici plus forte que dans les entourages de *France,* nom davantage soumis à la conjoncture présente.

Avons-nous appris quelque chose sur le sens pris par *France* et *Français*, dans les situations de face-à-face et dans la bouche de nos présidentiables ? En fait les enseignements apportés par les lexicogrammes se réduisent presque à la stratégie de discours où les deux mots sont impliqués : situation dans le temps présent et l'espace géopolitique pour *France,* appels à l'électorat (surtout Chirac) ou invocation de l'électorat (surtout Jospin) pour *Français.* Aller plus loin serait délicat, vu le petit nombre d'occurrences des deux pôles choisis.
Prenons maintenant un pôle à très haute fréquence, qui se situe à l'origine même de la parole émise : le pronom énonciateur *je-j'.*

2) Connexions autour de la première personne

Le genre "Face-à-face" engendre, de la part des candidats, un discours à la première personne. Avec le pronom sujet qui lui correspond, nous pouvons faire l'hypothèse que nous atteindrons les mots-clés du message lié directement à l'expression ou à l'implication de soi. En prenant les formes *je-j'* (réunies dans le lemme *JE*) pour source, quelles configurations riches en suites de connexions allons-nous découvrir ? C'est comme une pierre jetée dans l'eau du sens et qui irait, de ricochet en ricochet, jusqu'à épuisement. Que rencontre-t-elle au cours de cette itération ?
Le tableau de la ventilation des formes pronominales (des 1e et 2e personnes) donne dans notre corpus la distribution fréquentielle (et les probabilisations) suivantes:

	Corpus 40913 occ	Mitterrand 9593 occ	Chirac1 9567 occ	Chirac2 9917 occ	Jospin 11836 occ
je	800	171 (- 2)	196	182	251 (+2)
j'	228	64 (+2)	55	51	58
moi	129	25	38 (+2)	33	33
me	98	27	35 (+3)	17 (-2)	19 (-2)
m'	82	24	22	16	20
personnellement	4	0	0	0	4
mon ma mes	118	22	46	14	36
	1459	333	392	313	421
moi...je		12	23	22	19
nous	290	61	90 (+3)	71	68
vous	592	163 (+3)	186 (+6)	127 (-2)	116 (-8)

Figure 5. Ventilation des formes pronominales et possessives.

NB: Quatre observations: 1- Les "challengers" (Chirac1 face à Mitterrand, puis Jospin face à Chirac2) accentuent la projection personnelle (392 occ. contre 333, 421 contre 313), comme pour se pousser sur le devant de la scène. Dans ce domaine, l'effort de Jospin pour affirmer son "je" est visible. 2- Corrélativement, la diminution de cette auto-affirmation chez Chirac d'une élection à l'autre confirme cette interprétation : le premier ministre de la cohabitation, ancien challenger de Mitterrand, n'a, en 1995, plus besoin d'insister tellement sur le "je". 3- Le "nous" semble, en revanche, correspondre à une stratégie plus chiraquienne que socialiste. 4- La prise de parole en 1988 s'adresse à un "vous" direct (l'adversaire ou, parfois, les journalistes), davantage que celle de 1995, avec une réserve très visible sur l'emploi du "vous" de la part de Jospin.

Le lexicogramme du pôle *JE* chez les quatre présidentiables fait, quant à lui, surtout apparaitre sur sa droite un très grand nombre de cooccurrents et tout particulièrement, de verbes. Un tel constat était attendu, vu le fonctionnement antéposé du sujet en français. Un seul cooccurrent important sur la gauche répond présent au niveau des règles, mais chez Chirac et chez Jospin seulement : le pronom *moi.* Sur 104 occurrences de *moi* chez eux, 64 sont dans l'antécédance du JE. Mitterrand fait exception : la séquence *moi je* n'atteint chez lui que 12 occurrences.

Moins évidente, si l'on prend les verbes à la première personne du singulier, est la ressemblance qui se dégage entre les quatre hauts de liste des cooccurrents droits.

	Mitterrand	Chirac1	Chirac2	Jospin
JE →	(235)	(251)	(233)	(309)
	veux	crois	crois	pense
	pense	voudrais	voudrais	crois
	vais	dis	veux	propose
	crois	réjouis	dis	veux
	voudrais	sais	propose	voudrais
	dis	souhaite	dirais	peux
	peux	veux	vais	dis
	ferai	considère	partage	dirais
	aime	vois	peux	sais
	espère	constate	connais	vais
	souviens	poursuivrai	mets	souhaite
	estime		reviens	ferai
	voulais		trouve	reviens
	parle		rappelle	réponds
	conteste			
	15	11	14	14

Figure 6. Cooccurrents verbaux du sujet JE.

Dans les sept premiers cooccurrents, en effet, quatre reviennent à quatre reprises: *veux, crois, voudrais, dis...* Neuf autres reviennent à 2 ou 3 reprises. Plus rares sont les verbes à la première personne qui caractérisent à ce niveau l'un des émetteurs ; ils figurent quasi tous d'ailleurs dans la seconde moitié des listes : *aime, espère, souviens, estime, voulais, parle, conteste* chez Mitterrand (le plus "original"), *réjouis, considère, vois, constate, poursuivrai* chez Chirac 1, *partage, connais, mets, trouve, rappelle* chez Chirac 2, *réponds* chez Jospin (le plus "consensuel").
On trouve parmi ces formes : soit des opérateurs discursifs, déclaratifs (*dis, pense*), constatifs (*vois, sais*), optatifs (*souhaite, espère*), performatifs (*ferai, conteste*); soit des actifs (*aime, réjouis, partage, mets...*) ; soit des modalisateurs d'opérateurs discursifs à l'infinitif : *veux, peux, voudrais, vais* sont presque tous suivis de *dire* (lui aussi en tête des listes). Ces constats (proches pour certains du "Français fondamental" des années cinquante) montrent des candidats très soumis aux usages communs qui excèdent tout style individuel. A peine peut-on remarquer la priorité donnée aux "Je pense" et "Je ferai" à la fois par Mitterrand et Jospin et la primauté, permanente, du "Je crois" chez Chirac.

Mais qu'entrainent ces verbes à leur suite, quelles grappes de connexions ? Cette question mène au lexicogramme récursif de JE. Obéissant aux règles de construction $p < 5\%$, F et CF >= 3, environnement : la phrase, le graphe poursuivit jusqu'à saturation atteint 89 sommets et engendre 116 arcs chez Jospin, atteint 78 sommets et engendre 101 arcs chez Chirac2. Pour la visibilité, nous nous contenterons donc des sous-graphes ci-joints, où le nombre de cooccurrents engendrés par chaque sommet ne peut excèder en cours d'analyse 5 à droite et à gauche, l'itération de recherche partant de la source JE.

Figure 7.
Lexicogramme récursif saturé du je *de Chirac2*
[58 sommets, 79 arcs]
(p ≤ 5%, F et CF ≥ 3, 5 meilleurs cooccurrents à gauche et à droite)

De taille à peu près identique (52 sommets chez Jospin, 58 chez Chirac), ces sous-graphes mettent en évidence des phénomènes très différents: des locutions de langue ou de discours (*veux dire, veut dire, mets (en) cause, premier ministre, logements sociaux, premier tour, moi je...),* des usages conversationnels (*permettez moi , eh bien moi* chez Chirac, à *Monsieur Jospin* dit par Chirac, Jospin a l'habitude de répondre par *Jacques Chirac* - mais des tiers comme *Juppé* et *Balladur* attirent *"Monsieur"*), des thèmes fondamentaux (le "noeud" européen chez Jospin, celui du "il faut" chez Chirac ; les stéréotypes "impartialité --> Etat" chez Jospin, "créer --> emplois" chez Chirac), des tactiques discursives ("différence" en antécédance à "Jacques" et à " Chirac", eux-mêmes opposés à "moi", chez Jospin ; nombreux adverbes d'insistance gravitant autour de *voudrais, crois* et *(il) faut* chez Chirac), des noeuds argumentatifs (cinq arcs autour de *propose* et quatre autour de *pense* chez Jospin, six arcs autour de *voudrais* et cinq autour de *crois* chez Chirac), et surtout des

liaisons à distance que les proximités textuelles ne permettent pas de repérer: arcs entre *moi* et *France* ou entre *je* et *Français* chez Chirac, entre *nous* et *pays* ou entre *je* et *débat* chez Jospin, etc...).

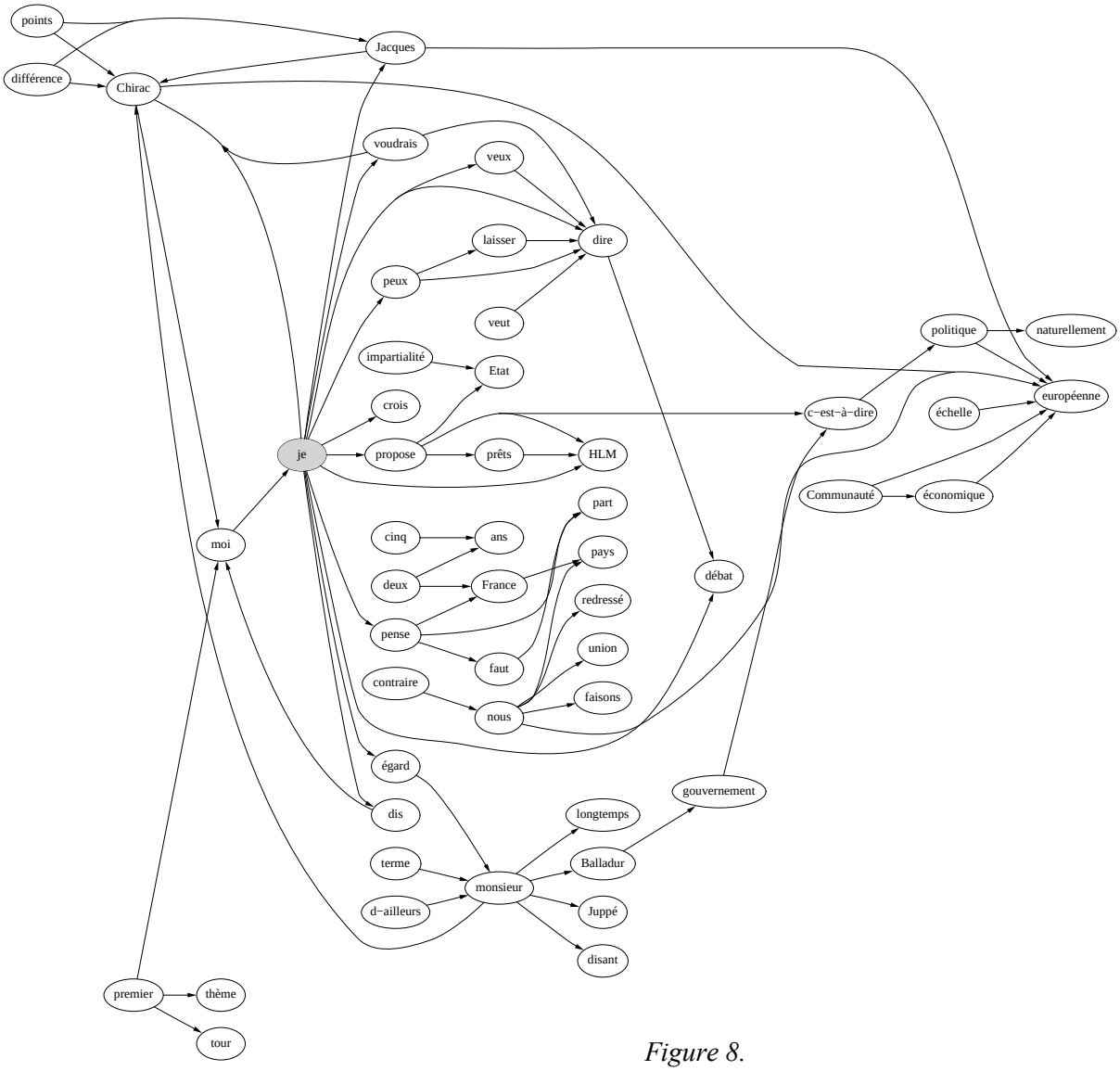

Figure 8.
Lexicogramme récursif saturé du je *de Jospin*
[52 sommets, 74 arcs]
($p \leq 5\%$, F et CF ≥ 3, 5 meilleurs cooccurrents à gauche et à droite)

Le "sens" des mots est-il concerné par ces simples constats statistiques ? Au niveau de ces lexicogrammes réduits, les entourages font surtout apparaitre des stratégies dans la mise en place des mots. Constater que dans l'ombre de *moi* il y a *France ,* que dans celle de *je* il y a *Français* et que se forme entre ces quatre sommets un cycle fermé, d'aspect tautologique n'est pas sans incidence sur la présentation de soi faite par J. Chirac devant les téléspectateurs; constater que *moi* et *je* sont impliqués dans une *différence* par rapport à *Jacques Chirac* ne l'est pas non plus pour L. Jospin. Ne saisit-on pas là, chez le premier, une volonté affichée de présidentiable incarnant déjà l'ensemble des Français et s'identifiant à la France et, chez le second, une tactique affichée de challenger s'opposant au premier (circuit *moi je Chirac*) et prenant place

face à lui (circuit *moi je dis*). En politique, la « posture discursive » est la première marche du sens.

Conclusion

Ces stratégies de mise en place de soi à travers la mise en place des mots sont – on vient d'en avoir un idée – étudiables par des automates. La lexicométrie des cooccurrences est, dans ce domaine, un bon outil , parce qu'elle permet, à partir de purs constats et calculs sur la surface textuelle, d'accéder jusqu'au projet du discours : quelle image donner de soi, comment entrainer l'électeur, afin que celui-ci acquiesce et agisse (donc vote) dans le sens désiré. L'automatisme de la méthode rend compte des automatismes du texte, décelables dans les systèmes de récurrences. Mais ces derniers , à leur tour, ne correspondraient ils pas aussi à des automatismes de notre mémoire et à ses réflexes mentaux, et les atteindre n'est il pas la visée de toute propagande ?

Notes:

(1) Méthode mise au point par Pierre Lafon, au laboratoire de Saint-Cloud. Se reporter à sa thèse parue sous le titre : *Dépouillements et statistiques en lexicométrie,* Genève-Paris, Slatkine-Champion, 1984, p. 129-199.

(2) Le soulignement dans le lexicogramme d'une forme cooccurrente, de sa fréquence et de sa cofréquence avec la forme pôle indique au chercheur qu'il peut accéder immédiatement, par lien hypertextuel, respectivement au lexicogramme de la forme soulignée, à sa concordance dans le texte ou à celle du couple.

(3) Afin d'obtenir des graphes plus proches de la sémantisation, nous avons systématiquement éliminé, non du calcul mais des représentations graphiques, tous les couples formés avec une forme-outil (une liste de 330 a été "sortie" de la recherche), ainsi que tous les couples formés avec un numéral.

(4) Les "lexicogrammes récursifs" sont dits "saturés" lorsque plus aucun "sommet" - cooccurrent pris pour pôle - ne répond aux règles de cooccurrence et de connexion fixées à l'origine de la recherche.

(5) Sur le calcul de cette probabilité (méthode dite "des spécificités"), voir P. Lafon, *ibid.,* p. 45-85. Les signes + ou - signifient qu'il y a suremploi ou sousemploi statistique de la forme en question. Le chiffre qui suit indique l'ordre de la petitesse de la probabilité : 2 ≅ 5%, 3 ≅ de l'ordre du millième, 6 ≅ de l'ordre du millionième.