

Autour du *Je* présidentiel

Jean-Marc Leblanc, Pierre Fiala

CEDITEC – Université Paris12 – Val de Marne – France
leblanc.jeanmarc@free.fr, fiala@univ-paris12.fr

Abstract

The functional comparison of various approaches to cooccurrence in some robust standard textual analyzers makes it possible to develop an experimental method in the courses of textual data processing intended for researchers in social sciences, which is too often absent from the field. Using a corpus of highly ritualized institutional speeches (New year addresses of the presidents of the French 5th Republic, 1958-2003), labelled or not, various functions to measure cooccurrence with LEXICO3, WEBLEX, HYPERBASE, ALCESTE are used in turn and compared. Different cooccurrence data are thus highlighted through figures, sorting, and graphic display, each leading to specific interpretations. The comparison shows the absolute necessity of in-depth experimentation in the analysis of textual data before any hermeneutic venture.

Keywords: cooccurrences, lexicogrammes, labeling, political discourse.

Résumé

La comparaison fonctionnelle de diverses approches des cooccurrences dans quelques analyseurs textuels standard robustes permet de développer dans les formations d'informatique textuelle destinées aux chercheurs en sciences humaines, une démarche expérimentale, qui fait trop souvent défaut dans le domaine. À partir d'un corpus de discours institutionnels fortement ritualisés (messages de vœux des présidents de la 5^e république, 1959-2001), étiqueté ou non, diverses fonctions cooccurentielles de LEXICO3, WEBLEX, HYPERBASE, ALCESTE sont successivement utilisées et comparées. Elles mettent en évidence, par les calculs, les tris, et les présentations graphiques des faits de cooccurrences différents, susceptibles d'interprétations complémentaires. La comparaison permet d'insister sur la nécessité absolue d'une expérimentation approfondie dans le traitement des données textuelles avant toute entreprise herméneutique.

Mots-clés : cooccurrences, lexicogrammes, étiquetages, discours politique.

1. Développer l'expérimentation et la comparaison en analyse textuelle

Trente ans après la parution des premiers travaux représentatifs du domaine, l'analyse des données textuelles poursuit en France son développement comme champ interdisciplinaire où se côtoient des disciplines diverses qui confrontent leurs hypothèses et leurs méthodes : statistique, linguistique, sociologie, histoire, communication ou philosophie politiques, histoire littéraire, analyse documentaire. Des ouvrages récents témoignent de la permanence du débat (Chateauraynaud, 2003), quand ce n'est pas des polémiques (Labbé, 2003), rendues malheureusement stériles parce que plus attachées à des promotions disciplinaires ou personnelles qu'au développement du domaine. Les progrès des interfaces permettent heureusement maintenant de comparer utilement les outils et les développements qui sont proposés pour traiter les problèmes.

Le renouvellement de l'intérêt pour l'analyse du discours et de la communication politiques, stimulé par les recherches sur l'argumentation et les mécanismes de persuasion, mais aussi les controverses sur l'utilisation des données quantitatives, ainsi que les utilisations sauvages des outils statistiques dans les débats politico-médiatiques nous ont incités à développer depuis quelques années une démarche de formation à la recherche où nous soulignons l'importance d'une plus large expérimentation. Pourquoi et comment comparer les textes ? Comment tester et optimiser des outils logiciels toujours plus nombreux, analogues sans être pourtant identiques, ni dans leur esprit ni dans leurs fonctionnalités ni dans leur résultat.

Le calcul des cooccurrences, ou mesure probabilisée des attirances entre formes dans un contexte donné, qui repose sur une notion bien décrite dans le domaine, a connu plusieurs types d'implémentations et de nombreuses exploitations ces dernières années. Mesure des cooccurrences binaires et recherche des lexicogrammes (Lafon, 1984 ; Tournier, 2003 ; Heiden, 2003), voisinage (Labbé, 1990), cooccurrences des formes spécifiques (Salem, Martinez, 2003). Cette problématique est propice à une démarche expérimentaliste visant à mettre en évidence les propriétés exploitables de chaque méthodologie.

2. *Je* dans le corpus des vœux présidentiels (1959-2001)

À l'instar de Labbé (1990) et Tournier (2003), nous examinons, le statut du *je* présidentiel, réduit ici à sa forme graphique, en explorant ses espaces cooccurentiels dans les messages de vœux aux Français sous la Cinquième République (1959-2001). Nous mobilisons différents outils statistiques dans une démarche expérimentale, cumulative et comparée afin de mettre en lumière les réseaux lexicaux mais aussi sémantiques et thématiques qui gravitent autour de la marque personnelle centrale dans ce genre institutionnel. Les réseaux de cooccurrence, saisis par ces approches différentes apportent un éclairage sur la construction de l'éthos présidentiel au sein d'une forme fortement ritualisée mais aussi sur les outils eux-mêmes

2.1. *Les paramètres du corpus*

Les interventions produites de décembre 1959 à décembre 2001 soit 43 discours représentent un volume textuel de 41125 occurrences pour 5203 formes.

Ces discours forment une série textuelle chronologique, mais ils sont abordés ici dans leur dimension synchronique. Nous menons l'expérience sur le corpus dans son ensemble, tout en restant attentifs aux emplois individuels lorsqu'il s'agit de retourner au texte

2.2. *Les marques énonciatives*

Les spécificités par locuteur des pronoms personnels et adjectifs possessifs mettent en lumière le contraste des profils énonciatifs qui caractérise ce discours pourtant très ritualisé.

À l'énonciation fortement personnalisée, centrée sur le *je* de Giscard (*je*, +5), mais multipliant aussi les marques en direction des Français (*vous*, +20), on opposera en premier lieu l'importante prise de distance produite chez le général de Gaulle par le rejet des marques de la première personne du singulier (*je*, -16) et de la seconde du pluriel (*vous*, -25), la prise en charge de l'énoncé étant assurée par un *nous* dont le référent est essentiellement la France. (*nous*, +2)

3. Approches des cooccurrences de JE

3.1. LEXICO3 : la cooccurrence par un chemin inattendu : les spécificités

L'outil « carte des sections » établit la distribution de *je* dans la linéarité du texte, délimité ici en paragraphes, dressant une « topographie textuelle » du corpus. (Lamalle, Salem, 2002). Le calcul des cooccurrences spécifiques met en évidence à partir des sections ainsi délimitées les formes spécifiques des paragraphes attestant la forme. La liste produite porte sur l'ensemble du corpus et ne présente que les formes dont l'indice de spécificité est supérieur à 2. Le seuil de probabilité est de 5 %, la fréquence minimale des formes considérées est de 2 occurrences (Tableau 1).

Les spécificités positives montrent qu'une forte proportion de verbes, gravite autour du référent du locuteur : verbes marquant la volition (*souhaite*, *voudrais*, *forme* [le vœu], *veux*), le jugement (*pense*, *crois*), factifs (*fais* +3), verbes d'état et auxiliaires (*suis*, *ai*), énonciatifs (*dis*), verbes marquant la connaissance (*sais*, *vois*), quelques infinitifs (*Dire*, *adresser*), constituent l'essentiel du système verbal restitué par la recherche des cooccurrences spécifiques. On note par ailleurs la présence d'un verbe conjugué à la troisième personne du pluriel : *vivent*

(+3), dont on trouve la réalisation dans de fréquentes adresses aux Français *qui vivent à l'étranger* (De Gaulle, 1967), *qui vivent dans la solitude* (V.G.E, 1978), *qui vivent dans la peine* (Mitterrand, 1986), *qui vivent dans la difficulté quotidienne* (Mitterrand, 1988)...

Ces messages sont particulièrement marqués par des verbes de « circonstance » (*souhaiter, adresser former*), par des volitifs et des verbes exprimant la connaissance. Cependant, cette classification sémantique établie a priori ne résiste pas longtemps à l'examen des contextes.

Une analyse approfondie indique que la forme *voudrais* est intimement liée au référent de l'interlocuteur, sur employée chez les locuteurs qui précisément multiplient les marques énonciatives en direction des Français. La valeur n'est donc que rarement purement volitive, les emplois étant essentiellement métadiscursifs, intervenant dans des annonces de plan où bien souvent le locuteur s'adresse à une certaine catégorie de Français (*Je voudrais d'abord exprimer ma sympathie à toutes celles et à tous ceux qui vivent ces derniers jours de 1999 dans l'épreuve*. [Chirac, 1999]). Giscard et Pompidou qui entretiennent un lien très étroit avec les Français emploient cette forme dans une modalité directive qui intensifie la relation (*Je voudrais que vous sentiez, que vous compreniez...*).

Les contextes de *veux* montrent une tendance vers des emplois métadiscursifs ou explicatifs (*Je veux dire*), même si la volition apparaît parfois chez Chirac et Mitterrand dans une faible mesure.

Parmi les verbes exprimant le jugement, *penser* intervient essentiellement dans des énoncés énumératifs (*Je pense aux artisans, je pense aux agriculteurs, je pense à certaines petites entreprises*) mais bien souvent affectifs et empathiques, liés à l'évènementiel (*Et je pense aussi à nos compatriotes de Toulouse...*[Chirac, 2001]) ou plus généralement destinés à adresser un geste en direction des Français les plus démunis, évocation qui devient systématique à partir de Pompidou. (*Je pense spécialement à ceux de nos aînés qui vont franchir seuls le cap du nouvel an*. [Chirac, 2000], *Je pense à celles et à ceux d'entre vous qui connaissent le deuil, les chagrins, le poids de la maladie et de la solitude, qui souffrent du chômage*. [Mitterrand, 1981]). Parmi les verbes exprimant la conscience et la connaissance, l'examen des contextes montre que la forme *sais* entre essentiellement dans des modalités allocutives. Les emplois soulignent l'empathie, plus particulièrement chez Chirac, souvent comme procédé argumentatif. Cette marque d'empathie introduit dans de nombreux cas chez Chirac une relance incitative et mobilisatrice (Tableau 2).

Le champ lexical du rituel, verbes, substantifs, formules d'adresse est largement représenté parmi les formes qui gravitent autour du *je* présidentiel : *compatriotes, chers, mer (compatriotes d'outre mer), Français, Françaises, vive, année, bonne, heureuse, vœux*.

Parmi les spécificités négatives – minoritaires et dont les indices se révèlent moins élevés que pour les sur-emplois – nous retiendrons le diagnostic porté sur la première personne du pluriel. (*nous, -7*). Ainsi, le schéma *je/vous*, dont on relève les traces parmi les pronoms personnels et adjectifs possessifs se construit par le rejet de la première personne du pluriel.

3.2. La notion d'environnement thématique HYPERBASE

C'est à partir de la recherche des contextes qu'il nous est offert d'étudier l'environnement thématique sous Hyperbase. Cette recherche thématique procède d'un calcul de spécificité particulier. Plutôt que de porter un jugement sur la fréquence d'apparition d'une forme dans une sous partie du corpus par rapport aux autres et à l'ensemble, on calcule les spécificités des mots qui se trouvent au voisinage d'une forme pôle. En d'autres termes, et nous empruntons cette formulation à Etienne Brunet : « Ce programme de repérage thématique fait le décompte de tous les mots situés dans le même paragraphe que le ou les mots-pôles et mesure leur spécificité, c'est-à-dire la plus ou moins grande attirance que le mot-pôle exerce sur eux. » Le choix des partitions n'est pas sans incidence sur la recherche thématique. Aussi adopterons-

nous la partition locuteur qui présente l'avantage de ne pas morceler le corpus par trop ainsi qu'un découpage en paragraphes qui constitue un échantillon contextuel suffisamment large et correspond à une division naturelle du corpus.

Les formes constitutives de « l'environnement thématique » de *je* sont classées selon leur rang de significativité, c'est-à-dire dans l'ordre décroissant de l'écart réduit, qui s'étend ici de la valeur 25,79 à 2. Contrairement aux cooccurrents spécifiques de Lexico3, aucun diagnostic n'est porté sur les sous emplois. La seconde colonne présente l'effectif de la forme considérée sur l'ensemble du corpus, la dernière colonne indique le nombre d'occurrences dans la sous partie, constituée à partir des contextes de *je*. Pour faciliter la lecture, et la comparaison avec les résultats produits par Lexico3 nous avons porté les indices de spécificité relevés par cet outil. (Colonne *Sp*) (Tableau 3).

L'essentiel des faits que nous avons abordés par l'entrée des cooccurrents spécifiques est restitué. Les premiers rangs sont identiques à ceux observés au moyen de la méthode de cooccurrents spécifiques, ce qui conforte nos premières conclusions : le rituel et la thématique des vœux prédominent. Parmi les verbes : *souhaite* dont on a déjà examiné les contextes et établi qu'il était essentiellement employé dans la formulation de vœux, de même *forme*, *adresser*, mais aussi *limiter* (*je ne voudrais pas me limiter à vous présenter mes vœux...*). Les usages de la première personne sont essentiellement conditionnés par le genre discursif y compris chez les locuteurs dont l'énonciation est peu personnalisée. Le *Je* est à la fois dialogique et familier, rituel et circonstanciel. On s'interrogera sur la procédure d'élagage appliquée sous Hyperbase – dont on maîtrise peu les paramètres – mais dont la présentation synthétique fait émerger les faits saillants du corpus mais aussi des données plus anecdotiques, qui apparaissent sous Lexico parmi de nombreuses autres formes, dotées généralement d'indices très faibles de spécificité. On soulignera cependant l'absence remarquable des formes *chers* (indice positif de 4 sous Lexico), *compatriotes* (+5), *France* (+3), *Vive* (+5) *bonne* (+5) qui entrent dans la réalisation de formules d'adresse ou qui appartiennent à la dimension rituelle du discours (*Vive la France, bonne année...*). Ces outils fournissent deux représentations des faits de cooccurrences, propres à orienter l'analyse dans des directions parfois différentes, d'où la nécessité de les interroger systématiquement.

3.3. Du calcul hypergéométrique aux lexicogrammes (WEBLEX)

Le calcul de cooccurrences implémenté dans Weblex repose sur le modèle développé par Pierre Lafon (1984). On se reportera à Heiden (2004) et Tournier (2003) pour des approfondissements méthodologiques.

« Le lexicogramme d'un mot s'interprète comme une synthèse des cooccurrents gauches et droits d'un mot, à l'intérieur de toutes les phrases où il apparaît » (Heiden, 2004). Le tableau 6 fournit ainsi une dimension supplémentaire qui ne pouvait être appréhendée avec les outils précédents et présente les principaux cooccurrents gauches et droits de la forme pivot *je* dans l'ensemble du corpus vœux au seuils de fréquence et co-fréquence minimale de 3 occurrences, probabilité de 5 % (soit 5.0^{-2}), distance moyenne 1000 occurrences. Pour chaque cooccurrent la fréquence totale de la forme dans le corpus (**F**), sa co-fréquence avec la forme pôle, c'est-à-dire le nombre de rencontres attestées (**CF**), le diagnostic de probabilité de la rencontre (**P**) et la distance moyenne (**d_m**) fournissent les indications quantitatives et statistiques de ces rencontres.

Nous constatons au premier regard un déséquilibre entre les cooccurrents gauches et les cooccurrents droits qui dominent très largement, ce qui ne saurait surprendre compte tenu de la nature de la forme pivot qui implique par essence de plus fréquentes relations sur sa droite que sur sa gauche. Les cooccurrents gauches restituent la dimension rituelle du discours. Cinq d'entre eux sont constitutifs de formules d'adresse (*mes chers compatriotes de métropole et d'outre mer*). La position du substantif « vœux » ne surprendra pas d'avantage : (*les vœux que*

je forme, que je vous adresse...) ni la présence de la forme *cœur* dont les contextes attestent la réalisation (*C'est de tout cœur que je...*).

Les verbes, généralement post-posés au pronom personnel figurent tout naturellement à droite du pivot. *Souhaite* se trouve au premier rang (tri par probabilités) comme c'était le cas avec les autres analyses. Les onze premières formes à droite sont des verbes dont la morphologie indique qu'ils sont fléchis à la première personne du singulier, et qui sont peu distants de la forme pivot. On notera que le temps est majoritairement le présent de l'indicatif. Après la catégorie verbale viennent des adjectifs et des substantifs qui ancrent le discours dans le présent, ou sont issus de la thématique des vœux : (*heureuse, vœux, soir, bonne, année, adresse*)...

Peu de termes politiques émergent de ces lexicogrammes, ainsi que nous l'avions déjà observé au moyen des autres outils cooccurrentiels : le *je* est essentiellement mobilisé par le genre discursif. Une dernière remarque concerne la distance moyenne des cooccurrents, plus importante sur la gauche de la forme pivot que sur sa droite où l'on observe quelques collocations. Qu'en est-il de la présence forte de l'interlocuteur que nous avons cru déceler dans les analyses précédentes ? Parmi les verbes, aucun ne semble être conjugué à la deuxième personne du pluriel, ainsi que nous l'avions déjà noté. On ne s'étonnera pas de l'absence des marques personnelles renvoyant à l'interlocuteur dont on a noté la forte spécificité sous Lexico, confirmée par Hyperbase. Nous avons en effet choisi de conserver l'élagage des formes outils du vocabulaire afin de porter un regard différent sur le corpus, en ne considérant que les seuls mots pleins. Une expérience menée sur ces mêmes lexicogrammes sans suppression de ces formes a cependant confirmé l'attraction importante des marques de la première du singulier et de la seconde du pluriel.

Le lexicogramme récursif (Tableau 4) associé à une forme révèle des réseaux de cooccurrences sur le principe d'un enchaînement de cooccurrents de cooccurrents. Partant de la source du lexicogramme (*je*) il en établit les principales attirances. À partir de ces nouvelles formes, d'autres cooccurrents sont mis en évidence selon le principe que chaque nouveau cooccurrent est à son tour pris comme source. La représentation qui en résulte synthétise l'ensemble des connexions qui prennent naissance autour de la forme pôle. Le lexicogramme récursif de la première personne du singulier, a été établi avec élagage des formes outils aux seuils précédemment adoptés.

La thématique des vœux ressort nettement de cette représentation. La connexion de verbes gravitant autour du *je* est très visible : *forme, adresser, sais, demande, dis...*, verbes qui nous sont désormais familiers. Dans cette configuration, l'ordre des termes n'est pas restitué bien qu'il soit possible de le faire. Cette perspective spatiale permet de saisir des faits qui ne seraient pas nécessairement apparus aussi clairement sur des listes hiérarchisées. Ainsi les formules d'adresse sont-elles matérialisées par les nœuds *départements, territoires, outre-mer*. Les qualifiants de *vœux, ardents* et *confiants* que le général de Gaulle adresse à la France sont clairement identifiés, de même que les emplois métadiscursifs (*je veux dire, je voudrais dire, je voudrais adresser*). On se gardera cependant d'interpréter trop hâtivement ces connexions. Suivons par exemple la « branche » qui relie le pronom personnel à *disais*. Le verbe est lui-même relié à *an* qui forme un nœud à partir duquel deux nouveaux arcs partent vers *fêter* et *nouvel*. Nul doute que le substantif *an* entre en cooccurrence avec le verbe et l'adjectif. Des réalisations comme « *fêter le nouvel an* » ne seront pas difficiles à attester. Pas de doute non plus sur la relation *je-disais* vérifiée en contexte. Il en va tout autrement de réalisations poly-cooccurrentes qui attesteraient *je, disais, fêter, nouvel an* dans un même contexte. Ceci tient au caractère de récursivité des lexicogrammes selon lequel se révèlent les principales formes cooccurrentes de la marque personnelle – dont *disais* – prises à leur tour comme sources – révélant le substantif *an* – jusqu'à saturation du lexicogramme – aboutissant à *fêter* et *nouvel*. Chaque nœud matérialise ainsi des paliers de cooccurrences dont les réalisations

n'entrent pas nécessairement dans les mêmes contextes. Il s'agit donc, bien évidemment, de ne pas considérer le chemin « artefactuel » que nous venons de retracer comme un schéma syntaxique, un « squelette de phrase » (Martinez, 2003).

3.4. Une utilisation particulière du tri croisé (ALCESTE)

L'expérimentation menée ici repose sur une utilisation un peu particulière d'Alceste. Il ne s'agit plus de faire émerger les structures saillantes du corpus, d'en identifier les classes d'énoncés ou « mondes lexicaux » (Reinert, 1993 et 1998) mais d'utiliser le tri croisé pour mettre en évidence les cooccurrents du *je*, sur l'ensemble du corpus.

Cette analyse produit une classification binaire selon la présence absence de la forme croisée avec le corpus au sein des énoncés. La totalité des unités de contexte élémentaire est prise en compte sur la base d'une classification ascendante, 100 % des U.C.E sont donc classées, contrairement aux analyses usuelles qui reposent sur le principe de la classification descendante.

Le tableau 6 présente les formes les plus caractéristiques des deux classes obtenues, ordonnées selon le Khi2 décroissant. Ces classes nous donnent confirmation de certains faits observés par ailleurs.

– Sur l'aspect pronominal, la dimension dialogique du discours est confortée, ainsi que le rejet de la première personne du pluriel: *vous* figure au second rang des formes significatives de la classe 1, les marques de la première du pluriel à un rang identique dans la classe opposée.

– Sur l'aspect énonciatif: Parmi les énoncés qui s'opposent au *Je* la variable *loc_dg* est la première forme significative de la deuxième classe. C'est-à-dire que la plupart des énoncés du locuteur de Gaulle se trouve classé dans la catégorie la plus éloignée du *je*. De Gaulle étant le locuteur qui use le moins de cette personne, ce résultat tend à conforter nos premières conclusions.

– Sur la chronologie: D'autres variables figurent parmi les premières classes constitutives de la classe 2. Les années 63, 65, 61, 68, 60, 66, 2001 et 2000 sont en effet dotées d'indices de Khi2 relativement importants. Ainsi cette classification binaire et somme toute assez brutale porte un éclairage chronologique sur les emplois pronominaux puisque, recoupant cette information par les spécificités par partie, il ressort que la forme *je* est sous employée dans les messages correspondants par rapport à l'ensemble du corpus. Ainsi résumons-nous notre démarche par ce va et vient entre les différents outils, les résultats de l'un nous engageant à interroger l'autre.

Dans la première classe, en revanche, la variable *Giscard* est la plus représentative. Nous avons constaté qu'il était celui chez qui la personnalisation du discours était la plus sensible. Ses messages par ailleurs sont ceux qui laissent la part la plus importante à la thématique des vœux. Ce fait se vérifie, là encore par la méthode des spécificités.

Cette thématique des vœux est précisément très représentée dans la première classe, à travers des substantifs : *vœux, année, soirée* à travers des adjectifs : *heureux, bonne* mais aussi des verbes : *souhait<, forme, adresser, présenter, échanger*. Le rituel et les formules d'adresse sont par ailleurs constitutifs de cette première classe : *Compatriote +, cher +, metropole +*.. On notera que le lexique est plus affectif dans cette même classe, tandis que la classe 2 recense un lexique plutôt politique, économique et social: *economie, développement, organisation, conflit...*

Cela nous amène à nous interroger sur la nature des relations entretenues entre les marques de la première personne et les différentes thématiques des messages. : Le *je* rejette-t-il dans ces

messages la dimension politique au profit d'un lexique plus circonstanciel ou rituel. Cette dimension politique serait-elle alors développée autour d'autres référents ?

4. Des résultats convergents, mais des mises en reliefs contrastées

Le croisement des approches et des outils statistiques dont la finalité première n'est pas nécessairement la recherche localisée de phénomènes cooccurentiels révèle des résultats convergents pour les faits les plus saillants. Cette démarche expérimentaliste et comparative a permis de réunir un faisceau sur lequel l'analyse interprétative peut s'appuyer plus fortement, que sur les faits révélés par un outil unique. Le croisement des perspectives valide par ailleurs la démarche lexicométrique sur un corpus de faible étendue tel que celui des vœux présidentiels. L'application de paramètres et de seuils, plus ou moins intuitifs mais souvent différents, le choix des fenêtres contextuelles, l'élagage ou non de certaines formes de vocabulaire, sont autant de facteurs qui incitent à confronter ces outils avant toute phase d'interprétation. Spécificités positives et négatives, topographie textuelle (Lexico 3), présentation synthétique (Hyperbase), ou ordonnée (Weblex lexicogrammes simples), réseaux d'affinités et récursivité (Lexicogrammes récursifs), partition binaire et classification ascendante (Alceste), fournissent des perceptions différentes et complémentaires des faits de cooccurrences.

Références

- Chateauraynaud F. (2003). *Prospéro – Une technologie littéraire pour les sciences humaines*. CNRS Editions.
- Heiden S. (2004). Interface hypertextuelle à un espace de cooccurrences: implémentation dans Weblex. In *Actes des JADT 2004*.
- Labbé D. (1990). *Le vocabulaire de François Mitterrand*. Presses de la Fondation Nationale des Sciences Politiques.
- Labbé D. (2003). *Corneille dans l'ombre de Molière. Histoire d'une découverte*. Les Impressions Nouvelles, Collection « Bâtons rompus ».
- Lafon P. (1984). *Dépouillements et Statistiques en Lexicométrie*. Slatkine-Champion.
- Lamalle C. et Salem A. (2002). Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels. In *Actes des JADT 2002*.
- Martinez W. (2003). *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de troisième cycle.
- Reinert M. (1993). Les « mondes lexicaux » et leur logique. *Langage et société*, vol. (66).
- Reinert M. (1998). Quelques interrogations à propos de l'« objet » d'une analyse de discours de type statistique et de la réponse « Alceste ». *Langage et société*.
- Reinert M. (1990). Système Alceste: Une méthodologie d'analyse des données textuelles. In *Actes des JADT 1990*.
- Tournier M. (2003). *De France à Je. La traversée des emplois Cooccurrences et connexions*. Des sources du sens, École Normale Supérieure Lettres Sciences Humaines Lyon, Collection Langages.

Annexes

Forme	Frq. Tot.	Fréquence	Coeff.	Forme	Frq. Tot.	Fréquence	Coeff.
je	344	344	51	j	88	51	3
souhaite	65	64	23	grandeur	7	7	3
vous	326	227	23	soir	42	28	3
voeux	80	62	10	adresser	7	7	3
mes	102	75	10	ma	20	15	3
sais	19	19	8	nom	30	21	3
voudrais	19	19	8	mon	29	19	3
pense	22	21	7	amis	11	10	3
suis	26	24	7	fraternité	19	14	3
forme	17	16	5	seuls	14	12	3
vive	60	42	5	france	302	150	3
heureuse	22	19	5	vivent	10	9	3
bonne	76	51	5	m	23	17	3
veux	11	11	5	vois	6	6	3
crois	11	11	5	ministre	6	6	3
que	677	336	5	famille	25	17	3
dire	48	35	5	très	27	18	3
dis	12	12	5	fais	6	6	3
compatriotes	62	43	5	françaises	41	27	3
ai	41	31	5	doivent	14	1	-3
année	205	110	4	quel	12	0	-3
vos	39	27	4	la	1397	546	-3
chers	55	37	4	algérie	21	3	-3
me	22	17	4	économique	46	10	-3
français	142	80	4	qu	313	108	-3
votre	59	38	4	peut	50	11	-3
chacune	25	20	4	europe	99	28	-3
mer	19	14	3	part	32	3	-5
				nous	655	217	-7

Tableau 1. Cooccurents spécifiques de « je » sur la totalité du corpus « vœux »

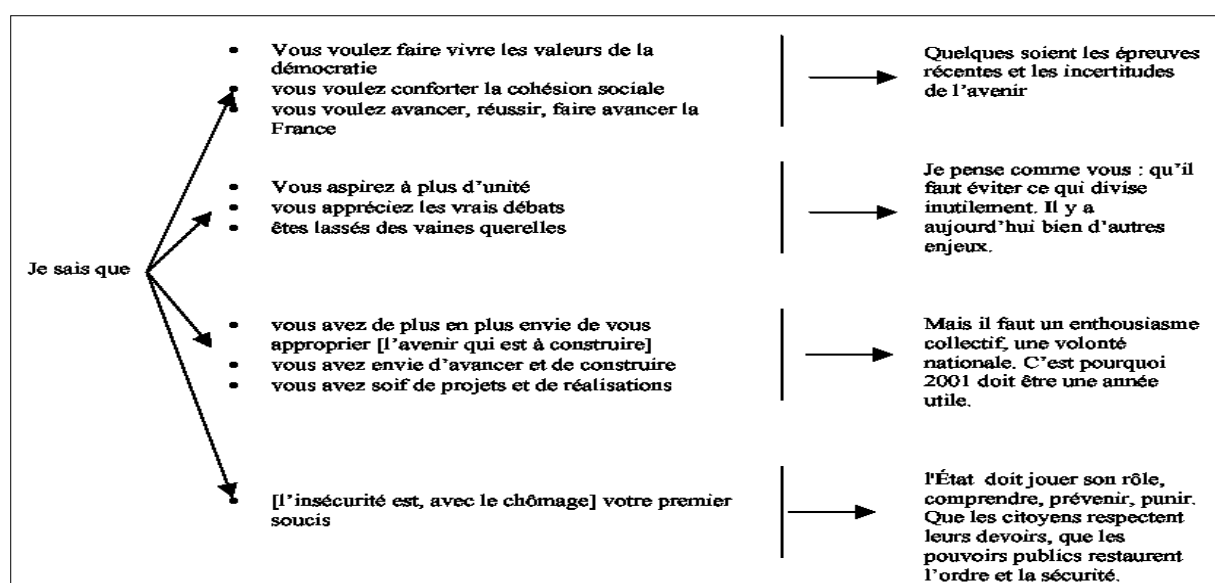


Tableau 2. Marques de l'empathie et relances incitatives chez J. Chirac

Ecart	Corpus	Extrait	Mot	Sp.	Ecart	Corpus	Extrait	Mot	Sp.
25.79	344	347	JE	51	2.74	4	4	LÉGISLATIVES	2
11.80	323	213	VOUS	23	2.74	4	4	AUGMENTÉ	2
11.06	65	65	SOUHAITE	23	2.71	29	17	MON	3
7.23	79	58	VEUX	10	2.65	11	8	RAISONS	2
6.46	19	20	VOUDRAIS	8	2.61	19	12	FRATERNITÉ	3
6.17	26	24	SUIS	7	2.60	127	58	AUX	1
5.99	22	21	PENSE	7	2.60	122	56	AVEC	2
5.98	19	19	SAIS	8	2.60	17	11	ADRESSE	2
5.56	48	35	DIRE	5	2.50	6	5	UNS	2
5.50	41	31	AI	5	2.50	6	5	MINISTRE	3
5.44	674	301	QUE	5	2.50	6	5	EXPRIME	1
5.15	17	16	FORME	5	2.50	6	5	DÉPARTEMENTS	2
4.75	12	12	DIS	5	2.50	6	5	AVAIS	2
4.55	11	11	VEUX	5	2.40	178	77	PAS	2
4.55	11	11	CROIS	5	2.40	33	18	GOVERNEMENT	1
4.34	25	19	CHACUNE	4	2.40	8	6	COMPRENDRE	2
4.25	59	36	VOTRE	4	2.35	202	86	ANNÉE	4
4.20	22	17	ME	4	2.35	18	11	AVEZ	2
4.17	101	55	MES	10	2.33	14	9	SEULS	3
4.03	42	27	SOIR	3	2.33	14	9	CHOSSES	2
3.68	30	20	NOM	3	2.22	41	21	FRANÇAISES	3
3.63	7	7	ADRESSER	3	2.17	90	41	AUSSI	2
3.59	3	4	LIMITER	2	2.16	21	12	POSSIBLE	1
3.55	101	52	CEUX	2	2.14	30	16	ABORD	2
3.52	39	24	VOS	4	2.13	37	19	AN	2
3.49	27	18	CŒUR	2	2.13	5	4	VRAIMENT	1
3.36	6	6	VOIS	3	2.13	5	4	VENU	1
3.36	6	6	FAIS	3	2.13	5	4	STRASBOURG	1
3.30	22	15	BONHEUR	2	2.13	5	4	RÉPONDRE	1
3.28	11	9	AMIS	3	2.13	5	4	PROGRESSER	1
3.08	25	16	FAMILLE	3	2.13	5	4	OFFRE	1
3.08	23	15	M'	3	2.13	5	4	KOWEÏT	2
3.07	5	5	MESSAGE	2	2.13	5	4	GARANT	1
3.07	5	5	MARS	2	2.13	5	4	FRAPPÉ	1
3.07	5	5	FIER	2	2.13	5	4	FIDÈLE	1
3.07	5	5	FAIBLES	1	2.13	5	4	ADRESSENT	1
3.07	5	5	DISAIS	2	2.12	19	11	MER	3
3.05	83	42	J'	3	2.09	17	10	ÉTRANGER	2
3.01	10	8	VIVENT	3	2.04	7	5	PRÉSIDENT	1
2.88	26	16	TRÈS	3	2.04	7	5	PARLER	1
2.85	22	14	HEUREUSE	5	2.04	7	5	MAJORITÉ	2
2.85	20	13	MA	3	2.04	7	5	FRATERNELLE	1
2.84	7	6	GRANDEUR	3	2.02	11	7	SOUFFRENT	1
2.84	7	6	DEMANDE	2	2.02	11	7	PROFESSIONNEL	2
2.77	142	65	FRANÇAIS	4	2.02	9	6	SOLITUDE	1
2.74	4	4	VAIS	2	2.02	9	6	RÉUSSIR	1
2.74	4	4	TIRE	2	2.02	9	6	DÉPIT	1
2.74	4	4	TIENS	2	2.02	9	6	DÉBUT	2
2.74	4	4	REÇU	2	2.00	24	13	TROP	2
2.74	4	4	QUICONQUE	2	2.00	24	13	BESOIN	1

Tableau 3. Environnement thématique de la forme « je ». Partition locuteur, contexte paragraphe

(346)cooccurrents gauches				je				cooccurrents droits						
	f	cf	p	d _m		f	cf	p	d _m		f	cf	p	d _m
chers	55	21	4e-06	7.3	souhaite	65	64	6e-57	0.2	faibles	5	4	2e-03	20.8
métropole	18	10	3e-05	11.1	pense	22	21	9e-18	0.0	salue	3	3	3e-03	0.7
compatriotes	62	20	1e-04	7.2	voudrais	19	19	3e-17	0.1	dirai	3	3	3e-03	1.0
françaises	41	15	2e-04	10.7	sais	19	18	4e-15	0.6	répète	3	3	3e-03	1.3
outre-mer	14	7	1e-03	8.6	forme	19	16	8e-12	0.0	souviens	3	3	3e-03	1.3
voeux	81	21	2e-03	4.9	crois	11	11	3e-10	0.1	assure	3	3	3e-03	14.0
coeur	27	9	8e-03	4.3	veux	11	11	3e-10	0.1	sûr	12	6	3e-03	1.0
vois	6	3	4e-02	17.0	dis	12	11	3e-09	0.7	rendre	9	5	4e-03	7.6
					dire	39	19	1e-07	3.3	promis	6	4	4e-03	2.0
					adresser	7	7	9e-07	6.9	espère	6	4	4e-03	5.8
					fais	6	6	7e-06	0.8	propose	6	4	4e-03	6.5
					heureuse	22	12	7e-06	13.5	dit	14	6	8e-03	3.8
					voeux	81	26	9e-06	7.3	fraternelle	7	4	9e-03	17.8
					soir	42	17	1e-05	6.8	demandé	4	3	9e-03	2.0
					bonne	75	24	2e-05	8.5	satisfaction	4	3	9e-03	12.0
					demande	7	6	4e-05	2.2	tire	4	3	9e-03	19.3
					disais	5	5	5e-05	0.8	nom	30	9	2e-02	6.4
					vois	6	5	3e-04	0.0	vivre	30	9	2e-02	13.1
					vais	4	4	4e-04	0.0	parle	8	4	2e-02	2.2
					année	202	42	7e-04	10.1	constate	5	3	2e-02	0.0
					adresse	17	8	9e-04	1.0	ardents	5	3	2e-02	3.0
					fier	5	4	2e-03	6.5					

Tableau 4. Lexicogramme de la forme je dans le corpus vœux. Seuils : f 3, cf 3, p 5.0E-2, d_m 1000.0

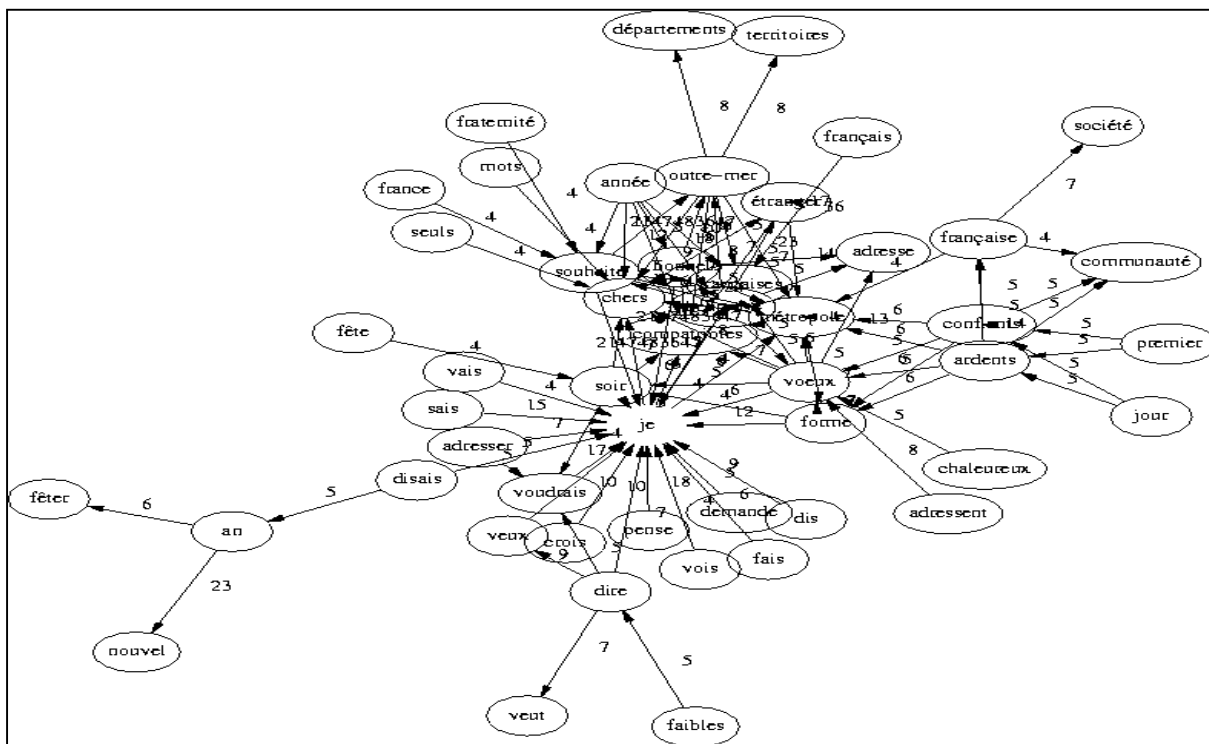


Tableau 5. Lexicogramme récursif de la forme je dans le corpus vœux.
Seuils : f 3, cf 3, p 5.0E-2, d_m 1000.0

Classe 1 : 255 U.C.E (27%)

Classe 2 : 714 U.C.E (73%)

Forme réduite	Khi2	Forme réduite	Khi2	Forme réduite	Khi2	Forme réduite	Khi2
je	969	vous	12,07	*loc_dg	52,38	trouver	4,1
vous	148,97	soir+	12,07	nous	14,84	ailleurs	3,97
souhait<	148,06	bonne+	11,34	econom+	10,6	leur	3,97
suis	60	redire.	11,25	année_1963	10,31	guerre+	3,93
dire.	45,77	tiens	11,25	devoir.	8,67	moderne+	3,73
mes	44,38	present+er	11,25	année_1965	7,29	solid+e	3,61
voeu+	38,95	echang+er	11,25	année_1961	7,27	rappor+er	3,61
dire+	38,14	metropole+	10,98	année_1968	6,24	acquérir	3,61
forme+	37,03	savoir.	10,9	année_1960	6,16	soi	3,61
me	27,54	m	10,52	moyen+	5,56	quel	3,61
année+	27,44	soiree+	10,12	developpement+	5,5	troubl+er	3,61
vouloir.	27,02	avais	10,12	*année_1966	5,12	scientifi<	3,61
heureux	23,87	coeur+	9,97	arme+	5,07	ensemble+	3,55
compatriote+	23,61	promettre.	9,86	cooperati+f	5,07	*année_2000	3,55
français+	23,31	parl+er	9,43	tech+n	5,07	*année_1962	3,44
adresser+	22,45	année_1974	9,34	peuple+	5,06	mondia+	3,44
cher+	19,82	mon	9,34	face+	4,71	moment+	3,38
ai	18,98	pas	9,32	organisat+ion	4,71	cas	3,24
nom+	17,59	ce	9,27	*année_2001	4,57	dehors	3,24
outre	15,11	ma	8,67	avons	4,52	telle	3,24
qu+	14,12	mer+	8,67	elle	4,46	multipli+er	3,24
enjeu+	14,07	recu+	8,43	jusqu+	4,34	monétaire+	3,24
adresse+	13,15	aim+er	8,43	plan+	4,34	but+	3,24
*loc_giscard	12,76	demand+er	8,09	conflit+	4,34	route+	3,24
propos+er	12,36	rappel+er	7,63	inflation<	4,34	aid+er	3,01

Tableau 6. Croisement de la variable « je » avec l'ensemble du corpus (Alceste)