

# L'analyse contrastive des réseaux de cooccurrence

## Le monde dans les discours des présidents de la Cinquième République

Jean-Marc Leblanc<sup>1</sup>, William Martinez<sup>2</sup>

<sup>1</sup>CEDITEC, Université Paris12 Val de Marne - leblanc.jeanmarc@free.fr  
<sup>2</sup>SYLED, Université de Paris 3 Sorbonne Nouvelle - will\_martinez@hotmail.com

### Abstract

This paper presents a series of lexicometric experiments aiming to describe phenomena of lexical co-occurrence in a corpus of New Year addresses by the French President of the Republic. Forming around the recurrent word *monde*, a co-occurrence network is shown to exist which can be represented with different degrees of precision by applying two programs: *Lexico3* and *Weblex*. At the first level of the study, binary co-occurrences describe phenomena of attraction and repulsion around the pole. Then by integrating the chronological characteristic of the corpus to the analysis, singularities can be pointed out in the use of the word *monde* by the consecutive speakers. Finally the exploration of co-occurrence networks shows that lexical evolution does not imply isolated units but rather complex lexical systems operating in the text.

**Keywords:** cooccurrences, networks of associated words, poly-cooccurrences.

### Résumé

Cet article présente une série d'expériences lexicométriques menées en vue de décrire les phénomènes de cooccurrence lexicale dans un corpus des messages de vœux des présidents de la République française. A partir du pôle récurrent *monde* s'élabore un univers cooccurrence qui est possible de représenter avec différents degrés de précision en exploitant deux logiciels : *Lexico3* et *Weblex*. Au premier niveau de l'étude les cooccurrences binaires décrivent les phénomènes d'attraction et de répulsion lexicale autour du pôle. En intégrant ensuite à l'analyse le caractère chronologique du corpus on relève les singularités de l'usage de cette forme chez les locuteurs consécutifs. Enfin, l'exploration de réseaux de cooccurrence montre que l'évolution lexicale concerne non pas des unités isolées mais des systèmes lexicaux complexes à l'œuvre dans le texte.

**Mots-clés :** cooccurrences, réseaux de mots associés, poly-cooccurrences.

### Introduction

Les allocutions de vœux des présidents de la République fournissent un matériau – bien que réduit – stable et homogène se prêtant particulièrement à la démarche lexicométrique, aussi bien du point de vue diachronique que contrastif. Ce n'est pas la parole présidentielle dans sa globalité que l'on saisit ici mais bien une manifestation particulière inscrite dans un genre discursif lié au rituel politique. Autre réduction : on ne s'intéresse ici qu'aux seules réalisations de la forme *monde* qu'on se propose d'examiner au moyen de différentes méthodes cooccurrence. Trois outils sont mobilisés : *Weblex* et les lexicogrammes récursifs, *Lexico3* et les cooccurrences spécifiques, ainsi que différents modules développés par Martinez [2003] comme la méthode des poly-cooccurrences.

### 1. Traduire sous forme de graphes de connexions des préoccupations différentes

On pose ici l'hypothèse que les emplois contrastés de la forme *monde* sont aptes à révéler le rapport à l'extérieur entretenu de façon différente par les locuteurs successifs et, partant de leur statut de chef d'État, de leur vision de la France. L'approche cooccurrence nous semble propre à établir des portraits croisés des présidents. Parmi les différentes méthodes, les

lexicogrammes récursifs apportent des éléments intéressants et nourrissent la réflexion sur l'interprétation qui peut être faite à partir de ces configurations.

Si la récursivité ne garantit pas nécessairement des co-présences réellement attestées au sein de mêmes contextes, on considère qu'elle permet en revanche de mettre en lumière de vastes réseaux et de prendre conscience de liens quasi neuronaux qui prennent naissance à partir d'une forme pôle. Chaque cooccurrent étant pris à son tour comme pôle, se dessinent de proche en proche des conceptions particulières, matérialisées par ces réseaux d'affinité. Les connexions qui se forment à partir de la forme *monde* qui nous occupe ici, révèlent selon les locuteurs, des perceptions très différentes de la France, de sa relation au monde, de son rôle. A partir de ces configurations, peut-on caractériser cette relation ? Est-elle sereine, assurée, posée, arrogante, trahit-elle une inquiétude, un pessimisme, un sentiment d'impuissance ?

Nous analysons ici un corpus de 42 000 occurrences constitué des allocutions de nouvel an des présidents de la cinquième République (1959-2001), doté de partitions *date* et *locuteur* au moyen desquelles nous contrasterons les usages. Nous calculons les lexicogrammes récursifs de la forme *monde* sur l'ensemble puis pour chacun des locuteurs et conservons les seuils et élagages par défaut<sup>1</sup>. Afin de ne pas surcharger les représentations nous choisirons de ne pas afficher d'étiquette d'arc. Enfin, la disposition n'est pas hiérarchique.

### 1.1. Configuration globale

Le premier lexicogramme nous fournit une synthèse des emplois sur l'ensemble du corpus. La forme *monde* y donne naissance à différents pôles d'attraction. Le pôle central autour de *monde*, représente les associations les plus syntaxiques (*partout, entier, capacité, tenir*). La symétrie des configurations peut-elle être prise en compte ? On se risquera à observer que la forme *monde* tient une position centrale et donne naissance à plusieurs nœuds qui correspondent à des axes sémantiques différents.

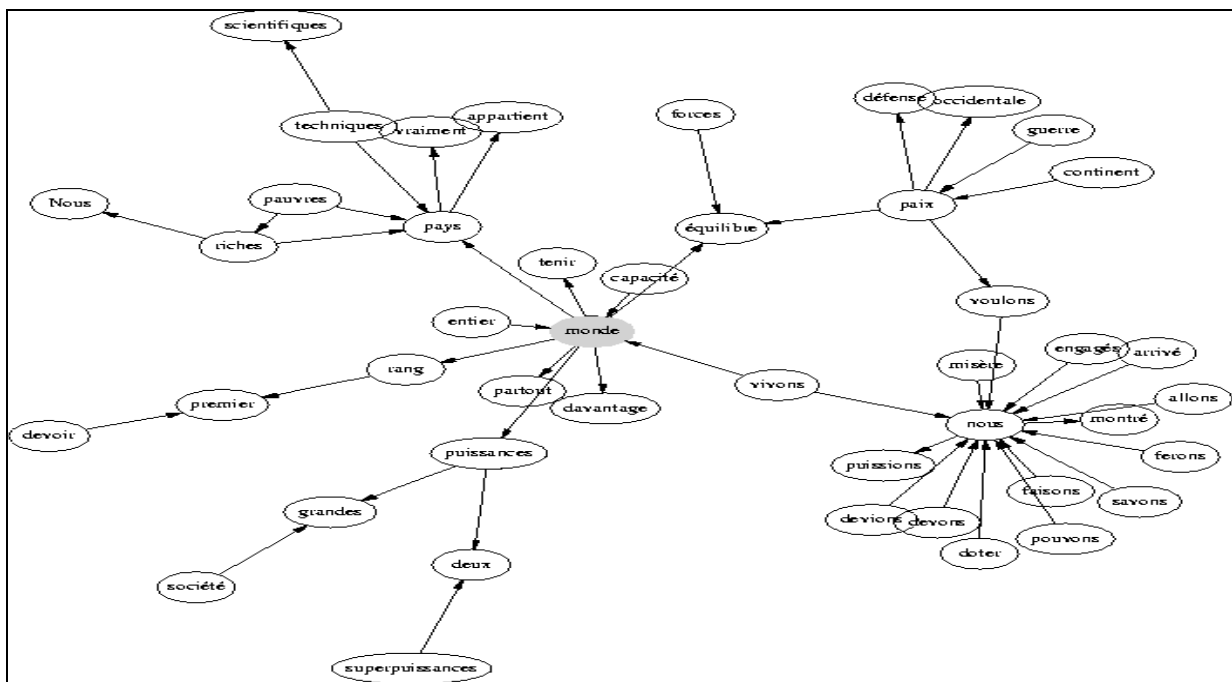


Figure 1 : Lexicogramme récursif sur l'ensemble du corpus

<sup>1</sup> Élagage des formes outils, numériques et hapax. Seuils : f 3, cf 3, p 5.0E-2, dm 1000.0.

- Un lien évident entre *nous* et *monde*, est matérialisé par le verbe *vivre* à la première personne du pluriel (*monde où nous vivons\**). C'est bien le rapport au monde qui est ici exprimé, dans une acception relativement large. Qu'en est-il de ce *nous*, doit on lui accorder une valeur de généralité ? Il n'est pas utile de commenter ici l'ensemble des cooccurrents de la marque personnelle qui n'entrent pas tous dans les mêmes contextes que *monde*, mais ceux-ci nous permettent de saisir globalement ses principales réalisations.

- A l'opposé de ce « *nous* », le pôle *pays* donne naissance à un nouveau nœud de cooccurrents.

Si toutes les cooccurrences ne sont pas attestées au sein des mêmes contextes, (par exemple *monde, pays techniques*), ces réseaux nous informent cependant du ou des sens dans lesquels sont employées les formes visualisées et orientent l'ensemble du lexicogramme. Ainsi le sens de la co-présence *monde* et *pays* est orienté par les cooccurrents de *pays* qui en sont de fait les qualifiants (*pauvres, riches* notamment). Nous saisissons ainsi les emplois les plus saillants du substantif et donc d'une partie des emplois de *monde*. C'est donc une certaine évocation du monde qui est faite ici, en particulier en termes d'équilibre, ou d'opposition entre pays riches et pays pauvres.

- Équilibre encore avec un troisième axe de sens qui se construit précisément sur la rencontre de *monde* et *équilibre*. Équilibre des *forces*, équilibre en termes de conflits internationaux (*guerre, paix*). *Monde* entre en cooccurrence avec *équilibre* qui à son tour appelle *forces* d'un côté et *paix* de l'autre. Quant au substantif *paix*, il appelle à la fois les lexies *défense, occidentale, guerre* et *continent* mais entre également en relation avec la première du pluriel, par l'intermédiaire d'un volitif (*voulons*).

- Un autre axe d'importance est tracé à travers les cooccurrents *puissances* et *superpuissances*, chacun donnant naissance à un embranchement différent. Ainsi explorons-nous une conception géopolitique du monde qui présente des affinités avec d'autres axes précédemment évoqués, construits quant à eux sur l'équilibre des forces. Voilà pour les nœuds principaux. On notera d'autres cooccurrents tel que *monde/rang/premier* dont il s'agira de vérifier en contexte s'il s'y construit une évocation de la France, et de sa suprématie ou tout autre chose. Le verbe *devoir* est-il cooccurrent de *rang* et *monde* ? On y verrait alors l'expression des obligations d'un pays qui se placerait parmi les premiers, aux premiers rangs mondiaux envers d'autres nations moins chanceuses, ou au contraire les actions nécessaires à son accession à de telles places. Peut-être encore s'agit-il du premier devoir de ces pays...

## 1.2. Le monde chez De Gaulle

Chez De Gaulle, le lexicogramme est singulièrement riche. (Figure 3). C'est que le pôle *monde* attire une autre forme, *France* à partir de laquelle un large réseau se forme. Il est assez intéressant de constater la disposition concentrique qui s'organise autour de *France*. Chez De Gaulle – si l'on peu accorder foi à cette disposition – c'est la *France* qui est au centre et non le *monde*. Le *nous* polysémique qui tenait une large place sur le lexicogramme global se précise chez lui pour se matérialiser en une entité centrale, la France. La première personne du pluriel qui donnait naissance à un large éventail de cooccurrents, n'est plus qu'un pôle périphérique parmi d'autres comme *monde*, pourtant pris comme pôle. Cette *France* entre en cooccurrence avec nombre de mots qui à leur tour forment de nouveaux nœuds. Ainsi, la France *peut* et *doit*, on lui assigne un *but*, mais est également évoquée par son action ou son état futur (*va*). Elle n'en est pas moins, ainsi que la première du pluriel, ancrée dans le présent (*maintenant, nous*). Elle est fortement associée au *je* du locuteur mais aussi à ses *enfants*. (Emploi métaphorique chez De Gaulle, *la France et ses enfants\**). C'est la France qui est directement liée à la *paix*, une paix qui entre en relation avec l'*Europe occidentale*, et l'*équilibre*. Une France liée à la *Politique*, politique *économique*, politique qui

se ferait dans l'*union*, politique comme *domaine*, politique de *défense*. Une France dont le rôle d'arbitrage semble s'exprimer dans l'association *France États*, mais aussi dans ses propres conflits (*France/Algérie*, mais une *Algérie Française*). Une France liée au rituel également (*Vive, République, année sérénité, année qui commence, année qui finit, année nouvelle*). Quant au monde, il ne donne lieu qu'à peu de connexions. Le monde *libre*, évoqué à trois reprises est le seul lien qui soit véritablement significatif.

### **1.3. Le monde, l'inquiétude, la crise chez Giscard**

Le discours de Giscard se distingue par la présence de nombreuses constellations et la multiplicité des référents. (Figure 3). En effet, autour de *monde, je, France, nous, français*, entraînent une manière de brouillage. Le lexicogramme ne semble pas s'appuyer sur des réseaux clairement identifiés. Ainsi ce monde a pour cooccurrent direct le substantif *problèmes*, tandis que *France* est liée aux *difficultés*. *Monde* et *vivons* évoquent l'environnement politique (*monde où nous vivons*) mais aussi la *paix*. La France dans le monde reste une préoccupation essentielle. La *France*, un *grand pays*, synonyme de *liberté*, doit *surmonter*, avoir *confiance* en *elle-même*. Pourtant, ce grand pays, n'est pas vraiment présenté comme un acteur.

### **1.4. Mitterrand : l'équilibre des puissances**

Chez Mitterrand, la configuration s'équilibre entre quatre pôles essentiels *monde, nous, Europe, pays*. (Figure 3). Ainsi, l'environnement du monde s'articule de façon harmonieuse : *nous*, la France, et ses préoccupations (*chômage, notre inflation*), tout ce qui relève de l'*économique*, mais aussi la *misère* qui nous entoure, qui *avons la chance de vivre dans un pays comme le nôtre* (attesté), qui *avons montré, (montré au monde ?)* qui devons *sortir* de telle ou telle situation tout ce qui fait ce *nous*, la France, et peut-être d'autres référents constitue un élément de ce monde avec l'*Europe*, l'*Europe des douze*, de la *communauté*, de la *défense*, de la *paix* qui est peut-être aussi un des moyens d'assurer l'*équilibre des puissances*, et puis, des *pays*, des *pays pauvres*, des *pays riches*, des *pays* qui cherchent ou qui jouissent de la *liberté*, mais aussi des *pays* qui ont une *histoire* (à moins qu'il ne s'agisse de *l'histoire de notre pays*). Une assez grande stabilité chez Mitterrand, un monde politique, un monde des puissances, dont l'équilibre est assuré par la France (ici *nous*) et par l'*Europe*. Mais pas de place semble-t-il pour ce « *monde où nous vivons* » d'acception générique.

### **1.5. Le monde de Chirac : adaptation, changement mutations**

Comme chez De Gaulle, le monde ne tient pas, chez Chirac, la place centrale du lexicogramme. (Figure 3). Mais alors que la *France* en était le centre chez De Gaulle, le *nous* s'y substitue chez Chirac. Ainsi, trois pôles essentiels sont ici en présence, *monde, nous* et *je*. La France figure en périphérie, et c'est le verbe *faire* qui semble lui donner naissance. *Faire ensemble, nous France, demain*, autant de repères spatio-temporels qui semblent caractériser depuis le début le discours chiraquien. *Monde pays, monde d'avantage*, voilà les seuls cooccurrents directs de notre forme pôle. Mais on y devine chez le dernier président la volonté d'agir sur le monde. Chez Chirac la *paix* ne *dépend* pas du monde ou de la France, mais de *nous* (le verbe *dépendre* figure d'ailleurs parmi les cooccurrents essentiels de la première du pluriel).

Différents pôles se construisent selon les locuteurs, pour lesquels nous venons de proposer – rapidement – quelques éléments d'interprétation. Peut-on affiner ces représentations et proposer d'autres mises en perspective ?

## **2. Monde, nous, France : un exemple d'expertise sous Lexico3**

### **2.1. Ventilation des fréquences**

Les méthodes cooccurentielles employées précédemment établissent l'existence d'un lien entre *nous* et *monde*, mais aussi entre *monde* et *France*. Sur la partition *locuteur* les courbes de fréquences relatives des lexies *monde* et *nous* sont inversées sur la quasi-totalité de la période : lorsque la fréquence d'emploi de *monde* s'accroît, celle de *nous* diminue et inversement. Doit-on y voir une forme de rejet de l'un par l'autre<sup>2</sup> ? En revanche le parallélisme de *monde* et *France* nous engage à penser que les deux formes pourraient apparaître conjointement. La forme *monde* ne serait donc pas systématiquement co-présente de la première personne du pluriel. Ce n'est donc pas toujours un rapport exprimé en termes de « nous et le monde » qui s'exprime ici. Les « destins » lexicométriques de *France* et *Monde* semblent en revanche étroitement liés, phénomène que les lexicogrammes récursifs de préfiguraient pas aussi nettement. La forme *France* n'étant clairement associée au monde que chez VGE et De Gaulle, occultée sur l'ensemble, discrète chez Chirac. Quant à la co-présence *nous* et *monde*, il convient de noter qu'elle est attestée au sein de huit phrases chez De Gaulle, ce que les lexicogrammes récursifs ne font pourtant pas apparaître, dix fois chez V.G.E, huit fois chez Mitterrand et douze fois chez Chirac. Peut-on pour autant affirmer que les formes *France* et *Monde* sont chrono-homogènes ? Sur une partition chronologique plus fine la tendance n'est pas aussi nette et on observe des accidents notables que les accroissements spécifiques et spécificités chronologique nous permettent de saisir plus finement.

## 2.2. Spécificités chronologiques et accroissements spécifiques

Les spécificités chronologiques montrent que les accroissements de *monde* (+3) et *France* (+3) se déroulent sur les mêmes périodes (1973-19881 pour le premier, 1973-1977 pour le second). La concentration de cette co-fréquence serait plus importante sur les années Giscard. L'indétermination chère à Chirac nous apparaît avec l'augmentation du syntagme *un monde* (+7) sur la période 1997-1998 et *une France* (+15) sur une période de deux ans, 1994-1996. Ce qui ne signifie pas que Chirac abandonne ensuite cet emploi mais qu'il y a un accroissement remarquable. La prise en compte des segments répétés – même s'il ne s'agit ici que de considérer les accroissements – permet d'observer que les emplois sont très différents entre les deux vocables choisis pour pôles. D'une part la France existe en position de sujet (*La France est*, +3 à partir de 1979), d'autre part la France s'inscrit dans deux types de segments *pour la France* et *pour que la France* qui expriment le moyen ou l'attribution, ce qui n'est pas le cas de la forme *monde*. On devine à la lecture des segments construits autour de *monde* des réalisations où la France est probablement associée (*dans le monde*, *au monde*, *du monde*), mais il convient de vérifier cette proposition en contexte. De même, une étude approfondie nécessiterait que l'on s'intéresse aux référents du *nous*, dont nous faisons mention précédemment mais aussi que l'on inclue dans nos mesures lexicométriques des segments comme *notre pays* qui peuvent être posés en équivalents de *France*.

Quant aux accroissements spécifiques, ils révèlent, plus ponctuels, des variations parfois décalées dans le temps entre les deux pôles. Un premier accroissement de *monde* à lieu en 1974 (+2), la même intensité est manifeste pour *France* mais l'année suivante, en 1975. Autre accroissement de *France* en 1977 puis 1978, les emplois de *monde* n'augmentent qu'en 1980. Enfin, *France* subit une nouveau « pic » en 1996, *monde* en 1998.

## 2.3. France et monde, contextes spécifiques

<sup>2</sup> Pas nécessairement si l'on considère que *nous*, n'est pas un cooccurent spécifique négatif de la forme *monde*. Fait intéressant, cette marque personnelle ne figure pas parmi les cooccurents essentiels de *monde*, d'après le calcul des cooccurents spécifiques, sur l'ensemble du corpus.

L'examen des contextes spécifiques<sup>3</sup> de *France* et *Monde* peut-il nous éclairer sur ce point ? Intéressons nous d'abord aux contextes de *France* et *Monde*. On notera chez Giscard quelques constructions remarquables : *ce qui est vrai du monde est vrai de la France...* (1977) ou encore « *Je suis conduit à penser tous les jours à la France dans le monde et à la France dans l'avenir* » (1978) qui pose un double positionnement, à la fois spatial et temporel. Les contextes de *France* et *monde* nous offrent de mettre en évidence une caractéristique commune à l'ensemble du corpus, le rang, le rayonnement de la France dans le monde.

« C'est pourquoi la **France** est respectée et estimée dans le **monde** » affirme Giscard en 1980 qui évoque encore cette même année le « rang de la France dans le **monde** » dont « nous sommes solidairement responsables ». Pour Mitterrand en 1982 : « Le rayonnement de la **France** est grand dans le **monde**, dans cette Europe qu'il faut construire, dans cet immense tiers-**monde** qui a confiance en nous. »

Giscard affirme en 1978 « Partout, je m'efforce d'affirmer l'image de la **France**, celle d'un pays ouvert sur l'évolution du **monde**, indépendant et fier mais fraternel, actif mais respectueux de la liberté et de la dignité des autres. »

Fort de cette image de la France, les présidents soulignent les obligations d'une nation que le monde entier prendrait pour exemple. Ainsi De Gaulle évoque en 1964 une conséquence évidente du « *redressement vis-à-vis du monde est que jamais tant d'hommes sur la terre n'ont tant attendu de nous ni éprouvé tant d'attrait pour la **France**.* », ou encore, en 1960 « *dans le monde, quelles conclusions désastreuses en seraient tirées quant à la capacité de la **France** d'assumer la responsabilité des affaires qui la concernent!* ». Là encore il mobilise le mode conditionnel pour exprimer une mise en garde.

Au-delà du rang et du rayonnement, les contextes spécifiques mettent en évidence le message, la voix de la France. Ainsi Giscard en 1978 s'exprime-t-il en ces mots : « *je souhaite que, [...], la **France** fasse rayonner son message de paix, de liberté, et de fraternité dans le monde.* »

Mitterrand quant à lui évoque en 1988, à propos du bicentenaire de la Révolution « ce message que la **France** a lancé au **monde** ». En 1981 encore : « nos autres objectifs [qui] seront principalement de donner à la communauté européenne des dix une volonté politique et de faire entendre la voix de la **France** parmi les peuples du tiers-**monde** ». En 1992 ce sont les soldats et plus généralement les Français de l'étranger qui sont promus messagers de la France : « vous penserez ce soir avec moi à ceux des nôtres, qui, partout dans le **monde**, en somalie, au Cambodge, en Bosnie, portent le message de la **France**, [...] ».

Ces énoncés qui participent de la fierté de l'appartenance nationale évoquent par ailleurs une supériorité française, entretenant un esprit de compétition mais aussi l'expression récurrente de la « *chance de vivre dans un pays comme le nôtre* ». Ainsi, Mitterrand en 1991 avance-t-il l'argument suivant :

« *Interrogez les chefs d'entreprise, les ingénieurs, les producteurs présents sur les marchés internationaux, eux qui peuvent comparer, qui savent que la **France** demeure l'un des quatre premiers pays du monde, qui constatent qu'à l'étranger on nous considère comme l'un des pays les mieux préparés pour le retour de la prospérité.* »

La France est ainsi à la fois un des pays les mieux placés au monde mais bénéficiant par ailleurs d'une image prestigieuse. Là encore il s'agit d'un fait récurrent à l'ensemble du corpus. Mitterrand encore, en 1990 évoque « *l'ambition de contribuer au succès de la **France*** »

---

<sup>3</sup> Nous utilisons ici le module d'extraction contextuelle *Contextes* développé par Martinez [2003] qui, à l'instar du programme proposé par Deroubaix [1996], permet de vérifier dans le texte les rencontres de différentes formes en hiérarchisant les extraits selon un ratio qui traduit la *densité informationnelle des contextes* par le rapport entre le nombre des formes recherchées et le nombre total des formes que compte l'unité contextuelle.

*qui reste, grâce à vous, l'un des premiers pays du monde.* ». Cette suprématie française, Giscard l'exprime différemment « *pour moi, la France c'est ce qu'il y a de meilleur dans le monde, à cause de son paysage et à cause de son peuple.* », (1974).

Le même Giscard affirme d'ailleurs ses ambitions pour la France : « *dans ma fonction, je suis conduit à penser tous les jours à la France dans le monde, et à la France dans l'avenir.* » (1978). Cet avenir est également appréhendé par Chirac :

« [...], pour que la France rayonne, pour qu'elle se déploie, pour que chaque français s'épanouisse dans le monde de demain, nous devons faire le choix de l'Europe, du mouvement et de la fidélité à l'idéal de la république. » (2001)

Ces contextes mettent en évidence une autre régularité, celle du monde dans lequel nous vivons. En 1975 Giscard revient sur la notion de rayonnement : « *aujourd'hui, dans le monde où nous vivons, compte tenu de notre dimension et, pour le moment<sup>4</sup>, de l'absence d'épreuve, le mot qui convient à la France est celui de rayonnement.* ». Mais il souhaite aussi, dans un énoncé plutôt polémique en 1976, que « *la France soit intelligente [...] pour sortir des idées étroites, des conceptions dépassées, pour bien comprendre le monde où nous vivons et pour choisir des solutions généreuses et courageuses.* » (1976).

Mitterrand en 1984 évoque la nécessité de l'union : « [...] dans le monde très dur où nous vivons, où l'on n'a rien pour rien, il faut que vous compreniez que la France a besoin de l'union de tous ses enfants. ». Pour De Gaulle, le monde s'inscrit sur fond de politique internationale : « *quant au monde où nous vivons, chacun voit jusqu'à quelle profondeur et avec quelle rapidité s'y modifient les relations internationales, et de quel poids, utile à tous, y pèse maintenant l'action de la France.* » (1966).

On citera enfin l'argument démographique cher à tous nos chefs d'État, illustré ici par De Gaulle : « *parmi les enfants qui sont venus récemment au monde, beaucoup verront un jour une France de cent millions d'habitants* », (De Gaulle, 1966).

#### 2.4. Contextes spécifiques de monde et nous

Les contextes spécifiques de *monde* et *nous* permettent de compléter cette rapide approche. La suprématie de la France y est à nouveau magnifiée, définissant par la même le référent de la première personne du pluriel. « *Nous sommes la quatrième puissance économique du monde* » s'enorgueillit Chirac en 1995, « *Ils témoignent du rang qu'occupe notre pays dans le monde et de sa capacité à prendre part au règlement des grands dossiers de la planète, à la place que nous avons héritée de la seconde guerre mondiale* » déclare Mitterrand en 1990. Ce type de déclaration n'est pas sans rappeler De Gaulle évoquant la capacité de la France à régler les problèmes aigus de l'univers, capacité précisément recouvrée grâce à une stabilité assurée par les institutions de la Cinquième République. Mitterrand en 1991 insiste sur le fait qu'à « *l'étranger on nous considère comme l'un des premiers pays du monde* », ou en 1993 énumère les grandes réalisations « *qui font de nous la quatrième puissance économique du monde* ». Notons par ailleurs que notre situation serait à en croire nos chefs d'État souvent « *enviée* » à l'étranger, fait, là encore, récurrent. Chez Chirac cette position dans le monde est souvent conditionnée à une nécessité. « *Pour qu'il [l'euro] nous permette d'améliorer notre rang et nos positions dans le monde, de grandes réformes [...] devront être engagées [...]* » (2001). De la même façon il évoque en 1999 une « *Europe qui nous permet de peser d'avantage dans le monde* », ou encore en 1995 « *si nous voulons être un pays qui compte dans le monde, nous devons, bouger, nous devons nous adapter.* ».

<sup>4</sup> On aura noté l'optimisme giscardien, « *pour le moment* ».

*Monde* est également synonyme d'environnement au sein duquel la France évolue. En 1998 Chirac évoque l'euro qui « **nous** apportera plus de stabilité dans un **monde** incertain », mais reste volontaire car en 1999 « dans un **monde** où rien n'est figé, l'avenir dépend de **nous** ». Giscard évoque en 1974 « la misère du **monde** qui **nous** entoure », mais aussi le « **monde** où **nous** vivons », en 1975, 1976 ou encore « le **monde** tourmenté où **nous** vivons » (1974), que l'on retrouve chez Mitterrand en 1984 « dans le **monde** très dur où **nous** vivons », cet environnement est toujours défini comme une donnée devant inciter à s'adapter, ou à s'unir. Chirac déplore d'ailleurs en 1996 « **nous** refusions de voir le **monde** se transformer autour de nous ». Chez Giscard en 1980 « vous savez que **nous** devons faire face à de grandes difficultés, parce que le **monde** change, qu'il est plus dur et moins prévisible qu'auparavant et qu'il peut devenir menaçant, et qu'en même temps **nous** voulons y maintenir notre rang [...] ». De Gaulle également évoque ce « **monde** où **nous** vivons » (1966), dans le contexte que nous avons déjà défini. Chirac évoque en 2001 « un **monde** où les foyers de haine et d'incompréhension sont nombreux », où « **nous** mesurons davantage combien il importe que l'Europe s'affirme [...] qu'elle participe en tant que telle aux évolutions du **monde** ». Giscard en 1974 affirme ne pas vouloir attrister les Français en leur « rappelant les difficultés et les risques réels du **monde** dans lequel nous vivons et dans lequel **nous** allons vivre l'an prochain ». Car c'est aussi l'évocation de périodes difficiles – passées ou à venir – qui s'exprime à travers cette co-présence. Mitterrand déclare en 1982 « comme la plupart des pays du **monde**, **nous** venons de vivre une année difficile », Giscard en 1977 évoque la nécessité d'avoir confiance « parce que **nous** traversons une période difficile, qui est une période d'évolution et d'adaptation dans le **monde** ».

Chez De Gaulle c'est le rôle international de la France qui est mis en avant. En 1968 il affirme que « les positions qui sont les nôtres [...], **nous** avons à les soutenir pour l'équilibre et pour la paix du **monde** ». Giscard évoque la politique étrangère en des termes moins volontaires : « **nous** vivons dans une de ces périodes où l'équilibre du **monde** repose sur la capacité de sang-froid de quelques hommes » (1979).

Les contextes spécifiques nous ont permis d'affiner quelque peu la première interprétation établie sur les lexicogrammes récursifs. *France* et *monde* sont deux cooccurrents essentiels dont il conviendrait d'approfondir encore les réalisations.

### 3. Cooccurrences spécifiques et poly-cooccurrents

Sur la base d'une segmentation du corpus des vœux présidentiels par le logiciel *Lexico3*, nous avons entrepris - grâce à un module informatique dédié<sup>5</sup> - d'analyser l'activité cooccurrentielle dans les contextes caractéristiques des 107 occurrences du pôle *monde*. Pour les expériences présentées ci-après l'exploration s'effectue dans les 101 phrases où apparaît le pôle (échantillon prélevé dans les 1 955 phrases du corpus) et n'est soumis à aucune restriction d'orientation ni limite de distance contrairement par exemple aux méthodes de collocation ou des segments répétés.

#### 3.1. Analyse des cooccurrences spécifiques

Pour appréhender les phénomènes cooccurrentiels autour du pôle *monde* nous avons analysé globalement les contextes d'apparition de cette forme. Dans l'échantillon de 3 445

---

<sup>5</sup> Le module *Coocs* exploité ici applique le modèle hypergéométrique qui est fondé sur la distribution en probabilité du nombre de rencontres de toutes les permutations possibles des formes étudiées dans l'hypothèse d'équiprobabilité. En adaptant le modèle à nos besoins, quatre paramètres sont nécessaires pour calculer un diagnostic en probabilité sur une cooccurrence :  $T$  correspond au volume en occurrences du corpus,  $t$  est le volume des contextes où apparaît le pôle,  $F$  est la fréquence du cooccurrent dans la totalité du corpus, enfin  $f$  est la fréquence du cooccurrent dans les contextes du pôle (la co-fréquence).



occurrences circonscrit (8% du corpus), le mode de calcul permet de relever à la fois les formes en suremploi et celles en sous-emploi. Le tableau 1 présente les 18 attractions et 8 répulsions ainsi détectées. D'emblée on constate que l'ensemble des cooccurrents négatifs définit clairement un champ sémantique cohérent et typique des discours de vœux : *je, année, français, mes, vœux, bonne, vive et votre*. En comparaison, les cooccurrents positifs forment un ensemble plus hétérogène à la fois sur le plan sémantique et grammatical : *vivons, pays, défense, rang, puissance, entier*, etc. Cette différence s'explique par le fait que l'exploration cooccurrentielle s'effectue à l'intérieur des bornes de la ponctuation phrastique ce qui restreint l'analyse aux seules phrases qui contiennent le pôle *monde*. Se trouvent donc exclues du comptage les phrases introductives des allocutions où foisonne le vocabulaire protocolaire des vœux. Dans l'extrait suivant on voit comment les deux classes de vocabulaire co-existent dans l'unité contextuelle du paragraphe tout en s'excluant au niveau plus précis de la phrase<sup>6</sup>.

F. Mitterrand, 1990

« *mes chers compatriotes, ce soir mes vœux tiendront en quelques mots très simples, ceux que vous emploieriez vous-mêmes quand vous vous direz "bonne année". que 1991 vous soit aussi heureuse que la vie le permet, que vous soient épargnées les grandes peines, la souffrance et la solitude, que vous vous sentiez solidaires, là où vous êtes, de ceux qui vous entourent et, d'une façon plus large, que vous ayez l'envie, l'ambition de contribuer au succès de la France qui reste, grâce à vous, l'un des premiers pays du monde. vive la république ! vive la France !* »

Tableau 1 : Cooccurrents spécifiques du pôle monde ( $Sp \geq 3$ ,  $F. min. \geq 5$ )

Cooccurrents positifs				Cooccurrents négatifs			
Forme	F	CF	Sp	Forme	F	CF	Sp
<i>dans</i>	351	61	+E08	<i>je</i>	344	17	-E03
<i>du</i>	281	52	+E08	<i>année</i>	205	6	-E03
<i>le</i>	742	104	+E07	<i>français</i>	142	4	-E03
<i>vivons</i>	11	7	+E06	<i>mes</i>	102	1	-E03
<i>tiers</i>	6	5	+E05	<i>vœux</i>	80	1	-E03
<i>pays</i>	122	22	+E04	<i>bonne</i>	76	0	-E03
<i>où</i>	78	17	+E04	<i>vive</i>	60	0	-E03
<i>tant</i>	35	10	+E04	<i>votre</i>	59	0	-E03
<i>défense</i>	20	7	+E04				
<i>rang</i>	12	6	+E04				
<i>puissances</i>	8	5	+E04				
<i>un</i>	306	38	+E03				
<i>si</i>	93	16	+E03				
<i>paix</i>	65	13	+E03				
<i>puissance</i>	19	6	+E03				
<i>équilibre</i>	16	6	+E03				
<i>entier</i>	6	3	+E03				
<i>message</i>	5	3	+E03				

**Guide de lecture du tableau 1 :** L'analyse des cooccurrences détecte les principales attractions autour d'un pôle en comparant, entre autres données, la fréquence globale de chaque cooccurrent ( $F$ ) avec sa co-fréquence ( $CF$ ) dans les contextes du pôle et fournit un indice de spécificité ( $Sp$ ) signalant son suremploi (+Ex) ou son sous-emploi (-Ex).

<sup>6</sup> L'unité contextuelle choisie pour l'analyse cooccurrentielle conditionne la prospection lexicale et détermine la quantité et le type de l'information rapportée : dans une fenêtre de 1 ou 2 mots autour du pôle, des associations plutôt d'ordre syntaxique, et dans une fenêtre plus large comme le paragraphe, des liens sémantiques. La quantité d'information livrée étant liée à la taille de la fenêtre d'exploration (Cf. Dubrocard [1995] qui mesure par un indice du Chi2 l'évolution de ce rapport dans des fenêtres de différentes dimensions), Reinert [1998] propose de définir *ad hoc* la meilleure 'unité de contexte élémentaire' pour l'étude d'un corpus donné.

### 3.2. Analyse des cooccurrences spécifiques par partie

La deuxième étape de l'analyse intègre au calcul des attractions lexicales la dimension contrastive du corpus. En effet ce dernier est jalonné de clefs méta-textuelles signalant l'évolution chronologique de cette compilation et la succession des différents présidents-locuteurs. Dès lors on peut adapter le paramétrage du modèle hypergéométrique<sup>7</sup> afin de détecter les formes en suremploi (ou en sous-emploi) dans les contextes du pôle *monde* chez un locuteur particulier comparé à l'ensemble des autres émetteurs (Tableau 2).

Tableau 2 : Cooccurents spécifiques du pôle monde chez chaque locuteur

De Gaulle				Giscard			
Forme	CF Glo	CF Loc	Sp	Forme	CF Glo	CF Loc	Sp
<i>jour</i>	4	4	+E03	<i>vous</i>	34	19	+E04
<i>les</i>	57	24	+E02	<i>je</i>	17	11	+E03
<i>en</i>	50	22	+E02	<i>liberté</i>	4	4	+E03
<i>ne</i>	10	6	+E02	<i>et</i>	99	37	+E02
<i>tout</i>	13	7	+E02	<i>des</i>	41	17	+E02
<i>y</i>	11	7	+E02	<i>dans</i>	61	23	+E02
<i>ou</i>	7	5	+E02	<i>france</i>	31	13	+E02
<i>chacun</i>	3	3	+E02	<i>son</i>	12	7	+E02
<i>tant</i>	10	6	+E02	<i>souhaite</i>	5	4	+E02
<i>moyens</i>	3	3	+E02	<i>problèmes</i>	5	4	+E02
<i>doute</i>	3	3	+E02	<i>moment</i>	5	4	+E02
<i>libre</i>	3	3	+E02	<i>vivons</i>	7	5	+E02
Pompidou				Chirac			
Forme	CF Glo	CF Loc	Sp	Forme	CF Glo	CF Loc	Sp
<i>logement</i>	2	2	+E03	<i>elle</i>	22	9	+E03
<i>le</i>	104	7	+E02	<i>nous</i>	60	18	+E03
<i>ne</i>	10	2	+E02	<i>un</i>	38	12	+E02
<i>pas</i>	9	2	+E02	<i>s</i>	17	7	+E02
<i>n</i>	9	2	+E02	<i>être</i>	5	3	+E02
<i>ses</i>	6	2	+E02	<i>où</i>	17	7	+E02
Mitterrand				<i>doit</i>	5	3	+E02
Forme	CF Glo	CF Loc	Sp	<i>ensemble</i>	2	2	+E02
<i>europe</i>	14	9	+E03	<i>car</i>	2	2	+E02
<i>tiers</i>	5	5	+E03	<i>davantage</i>	4	3	+E02
<i>la</i>	130	44	+E02	<i>devons</i>	4	3	+E02
<i>politique</i>	8	5	+E02	<i>doivent</i>	2	2	+E02
				<i>marche</i>	2	2	+E02
				<i>2000</i>	2	2	+E02
				<i>adapter</i>	2	2	+E02
				<i>patrimoine</i>	2	2	+E02
				<i>plein</i>	2	2	+E02

**Guide de lecture du tableau 2 :** L'analyse des cooccurrences par parties rapporte les principales attractions autour d'un pôle en comparant, entre autres données, la co-fréquence globale de chaque cooccurent avec le pôle (*CF Glo*) dans la totalité du corpus avec sa fréquence locale (*CF Loc*) dans une partie du corpus.

Le tableau 2 livre les cooccurents spécifiques positifs (seuil de co-fréquence :  $\geq 2$ ) qui ont été relevés chez les cinq présidents et laisse apprécier l'évolution de l'univers cooccurentiel du

<sup>7</sup> Pour ce calcul,  $T$  correspond au volume en occurrences des contextes où apparaît le pôle,  $t$  est le volume des contextes où apparaît le pôle chez le locuteur étudié,  $F$  est la fréquence du cooccurent dans la totalité des contextes où apparaît le pôle (soit la co-fréquence globale), enfin  $f$  est la fréquence du cooccurent dans les contextes où apparaît le pôle chez le locuteur étudié (soit la co-fréquence locale).

pôle dont les composantes varient en nombre suivant l'émetteur du discours : de 4 cooccurrents chez Mitterrand jusqu'à 17 chez Chirac. On notera aussi que l'analyse a dégagé cinq constellations lexicales distinctes qui ne partagent qu'un élément commun, le négatif *ne* chez Pompidou et De Gaulle. Ces résultats confirment des usages particuliers de *monde* chez chaque locuteur comme l'illustrent les contextes des parties 1.2 à 1.5 de cet article.

### 3.3. Analyse des poly-cooccurrences

Dans la dernière étape de cette analyse il s'agit d'étudier les attractions lexicales au-delà de la cooccurrence binaire. Les phénomènes de cooccurrences multiples et simultanées étant, de par leur nature plurielle et diffuse, plus complexes à étudier, il a fallu adapter la méthode de détection des cooccurrences spécifiques à la mise en évidence de réseaux cooccurrentiels. L'algorithme de prospection contextuelle employé dans cet objectif se distingue par deux caractéristiques principales : la récursivité et la profondeur de l'exploration contextuelle. Cette procédure est décrite schématiquement dans l'énumération ci-dessous où le signe '+' symbolise la notion de *cooccurrence chaînée* qui est à la base de la procédure.

- Etape 1 : Le pôle A a pour cooccurrents spécifiques les formes B, C et D
- Etape 2 : Le pôle A+B a pour cooccurrents spécifiques E et F
- Etape 3 : Le pôle A+B+E n'a pas de cooccurrent spécifique, alors l'exploration s'interrompt pour ce chemin
- Etape 4 : Le pôle A+B+F a pour cooccurrents spécifiques I et J
- etc.

Au fur et à mesure que l'exploration cooccurrentielle progresse, la poly-cooccurrence mise au jour se précise et l'espace contextuel exploré se réduit jusqu'à ne plus présenter de formes spécifiques. Au final, la méthode livre une représentation graphique du réseau d'associations lexicales détecté autour de la forme pôle. L'arborescence de la figure 2 reproduit le réseau détecté à partir de *monde* dans les discours du locuteur Giscard.

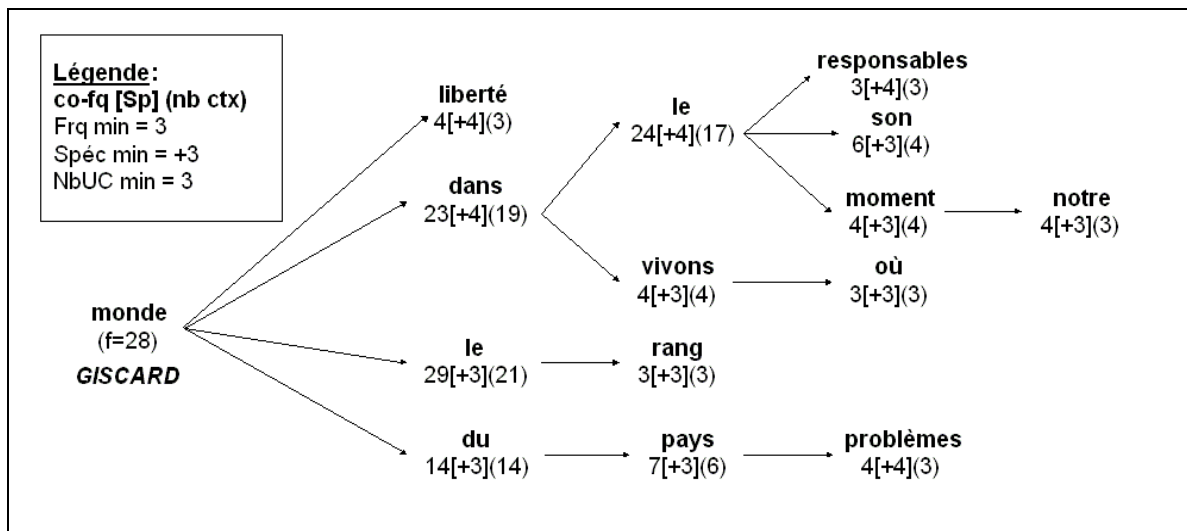


Figure 2 : Poly-cooccurrents autour du pôle monde chez V. Giscard

**Guide de lecture de la figure 2 :** Pour faciliter la lecture des chemins de cooccurrence nous employons une triple notation sous la forme de cartouches avec la syntaxe suivante : *co-fréquence [spécificité] (contextes d'apparition)*. Par exemple, chez Giscard à partir du pôle *monde* on signale une cooccurrence avec la forme *du* par une flèche notée 14[+3](14) qui indique que le cooccurrent apparaît 14 fois au côté du pôle avec une spécificité de +E03 dans un total de 14 phrases. En poursuivant la lecture à droite du graphique, on découvre que dans 6 de ces contextes réunissant *monde* et *du*, la forme *pays* apparaît 7 fois avec une spécificité de +E03.

Pour expliciter la procédure mise en œuvre reconstituons une partie du calcul à l'origine de l'une des 'branches' de l'arbre de la figure 2. Le tableau 3 montre les trois étapes de calcul nécessaires pour déterminer la poly-cooccurrence *monde+du+pays+problèmes* chez V. Giscard. L'analyse des six contextes d'apparition du pôle composé *monde+du+pays* (3<sup>ème</sup> colonne du tableau) révèle une cooccurrence spécifique avec la forme *problèmes* du fait de quatre rencontres dans trois de ces contextes.

Tableau 3 : Poly-cooccurents du pôle monde chez V. Giscard

Pôle : monde			Pôle : monde+du			Pôle : monde+du+pays		
Forme	CF	Sp	Forme	CF	Sp	Forme	CF	Sp
<i>dans</i>	23	+E04	<i>liberté</i>	12	+E04	<i>liberté</i>	12	+E04
<i>vivons</i>	5	+E04	<i>moment</i>	6	+E04	<i>moment</i>	6	+E04
<i>le</i>	29	+E03	<i>des</i>	38	+E03	<i>nord</i>	5	+E03
<i>du</i>	14	+E03	<i>pays</i>	12	+E03	<i>grands</i>	4	+E03
<i>son</i>	7	+E03	<i>autres</i>	6	+E03	<i>problèmes</i>	4	+E03
<i>moment</i>	4	+E03	<i>vivons</i>	5	+E03	<i>dans</i>	33	+E03
(...)			(...)			(...)		

On pourra vérifier ces résultats statistiques par l'extraction des contextes porteurs de la poly-cooccurrence *monde+du+pays* dont trois contiennent également la forme *problèmes* (voir ci-dessous).

V. Giscard, 1974-1980

- 1974 « vous avez vu que, dans les quelques mois qui ont suivi l'élection présidentielle et bien que la France ait choisi, pour la représenter, le plus jeune des dirigeants des grands **pays**, notre **pays** a affirmé à nouveau sa place parmi les nations responsables des grands **problèmes du monde**. »
- 1974 « je lui souhaite, dans le **monde** tourmenté où nous vivons, d'apparaître précisément comme un **pays** capable de proposer la conciliation aux autres et, en même temps, de rechercher avec réalisme la solution des **problèmes** qui se posent désormais à l'échelle **du monde** et qui sont donc des **problèmes** mondiaux. »
- 1978 « partout, je m'efforce d'affirmer l'image de la France, celle d'un **pays** ouvert sur l'évolution **du monde**, indépendant et fier mais fraternel, actif mais respectueux de la liberté et de la dignité des autres. »
- 1980 « mais c'est un des plus grands et, des plus vigoureux **pays du monde**. »
- 1980 « la force d'un **pays** se mesure ainsi à des signes politiques : la stabilité et l'efficacité de ses institutions, l'indépendance et la puissance de sa défense, son active diplomatie de paix, sa contribution à l'organisation **du monde**. »
- 1980 « vous, toutes et tous, vous êtes davantage responsables **du** destin de notre **pays** qu'à aucun autre moment de son histoire, parce que tout ce qui se passe dans le **monde**, connaissez aussitôt, parce que vous en savez plus sur les **problèmes** de notre temps que tous ceux qui vous ont précédés. »

#### 4. Quelques éléments de conclusion

Les lexicogrammes ont révélé autour du pôle *monde* des configurations cooccurentielles différentes selon les locuteurs et l'examen des contextes spécifiques a permis de mettre au jour des régularités et des particularités d'emplois chez les différents locuteurs. L'interprétation des lexicogrammes a permis d'engager une réflexion sur leur pertinence et les questions qu'ils peuvent soulever. Ainsi on a déterminé qu'un certain nombre de pôles articulaient l'environnement de *monde*, *nous*, *France*, *Europe*, *pays*...devenant eux-mêmes des centres d'univers cooccurentiels parallèles.

La méthode des cooccurrences spécifiques, quant à elle, a permis dans ses différentes étapes d'observer non pas les seules attractions lexicales autour de *monde* mais aussi les

formes que ce pôle rejette, puis de repérer l'évolution de cette activité cooccurentielle chez les différents locuteurs et enfin les poly-cooccurrences ont apporté un contrepoint à l'interprétation des lexicogrammes en traçant des *squelettes de phrases*, reflets de co-présences réellement attestées diminuant ainsi la part spéculative de l'interprétation.

Cet article compare deux méthodologies de la cooccurrence lexicale multiple qui, bien que fondées sur le même modèle statistique, conçoivent la mesure de cooccurrence de manière différente. Les lexicogrammes limitent l'analyse aux attractions mutuelles entre deux formes et, au-delà de ces relations binaires, ne garantissent pas la transitivité du réseau cooccurentiel. Les poly-cooccurrences, elles, élaguent les résultats par des seuils de fréquence et de spécificité, tout en contraignant à chaque étape de l'analyse l'aire contextuelle explorée. Pourtant, malgré leurs différences, une interrogation commune réunit ces deux méthodes : au-delà de la cooccurrence binaire comment mesurer la relation cooccurentielle ? Les lexicogrammes ouvrent l'exploration aux multiples interconnexions de la trame textuelle alludant sans cesse à de nouveaux pôles et à de nouveaux univers cooccurentiels tandis que les poly-cooccurrences resserrent le filet de la prospection sur un nombre de plus en plus petit de contextes et d'associations lexicales avérées autour du pôle initial.

## 5. Bibliographie

DEROUBAIX, J.-C. (1996). *Les déclarations gouvernementales en Belgique (1944-1992). Etude de lexicométrie politique*, Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3, sous la direction d'André Salem, Paris.

DUBROCARD, M. (1995). *César dans César : quelques remarques sur l'utilisation du mot Caesar dans les Commentaires sur la Guerre des Gaules. Influences des dimensions du contexte sur la détermination des cooccurrences significatives*, Travaux du Cercle de Linguistique de Nice, Nice.

HEIDEN S. (2004). *Interface hypertextuelle à un espace de cooccurrences : implémentation dans Weblex*, 7<sup>èmes</sup> Journées internationales d'Analyse Statistique des Données Textuelles (JADT 2004), Louvain-la-Neuve.

HEIDEN S., LAFON P. (1998). *Cooccurrences, La CFDT de 1973 à 1992, Des mots en liberté, Mélanges Maurice Tournier*, ENS Éditions, tome 1, Fontenay-aux-Roses.

LAFON P. (1984). *Dépouillements et Statistiques en Lexicométrie*. Slatkine-Champion. Paris.

LEBART L., SALEM, A. (1994). *Statistique textuelle*, Dunod, Paris.

LEBLANC, JM (2005). *Les vœux des présidents de la cinquième République (1959-2001). Recherches et expérimentations lexicométriques à propos de l'ethos dans un genre discursif rituel*, Thèse de Doctorat en Sciences du Langage, Université de Paris 12 Val-de-Marne, sous la direction de Pierre Fiala.

MARTINEZ, W. (2003). *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*, Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3, sous la direction d'André Salem, Paris.

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. Paris.

SALEM, A. (1998). *Approches du temps lexical. Statistique textuelle et séries chronologiques*, Mots n° 17, Ecole Normale Supérieure de Fontenay-Saint Cloud, Paris.

