



**HAL**  
open science

# Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique

Damon Mayaffre

► **To cite this version:**

Damon Mayaffre. Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique. Histoire. Université Nice Sophia Antipolis, 2010. tel-00655380

**HAL Id: tel-00655380**

**<https://theses.hal.science/tel-00655380>**

Submitted on 28 Dec 2011

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

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

**Université de Nice – Sophia-Antipolis**  
UMR 6039, *Bases, Corpus et Langage*

**DAMON MAYAFFRE**

**Vers une herméneutique matérielle  
numérique**  
Corpus textuels, Logométrie et Langage politique

Volume 1 :

**Mémoire de synthèse**

**Présenté en vue de l'Habilitation à diriger des recherches**

**Directrice Sylvie MELLET**

**30 avril 2010**

Jury :

**Paul CHILTON**

**Jacques GUILHAUMOU**

**Sylvie MELLET**

**François RASTIER**

**André SALEM**

**Jean-Marie VIPREY**

## Avertissement

Ce mémoire essaye de dire en peu de mots – moins de 100 pages – ce que j’ai pu développer ailleurs trop longuement – 1100 pages d’ouvrages ; 600 pages d’articles.

Les références bibliographiques ont été volontairement limitées pour réduire le feuilleté scientifique et l’écho intertextuel ; les notes de bas de page ont été autant que possible comptées afin ne pas multiplier les excursus ; le ton direct, libre, sans détour s’écarte en partie des précautions universitaires habituelles dans l’espoir d’égayer la lecture.

En tant que mémoire de synthèse, ce travail laisse le plus souvent de côté le corps des démonstrations pour ne dire que l’essentiel : la cohérence du parcours de recherche depuis 12 ans, les résultats obtenus, les convictions scientifiques naïves ou profondes. En complément, le lecteur sera renvoyé, au fil des pages, aux deux volumes d’articles et aux deux ouvrages joints au dossier.

## Remerciements

Mes remerciements vont d’abord vers Sylvie Mellet qui a été déterminante dans mon cursus intellectuel ces dernières années. Il m’apparaît chaque jour plus clairement la chance d’avoir été accueilli à *Bases, Corpus, Langages* sous sa direction et de travailler quotidiennement avec elle. Merci pour son soutien, son travail, ses conseils.

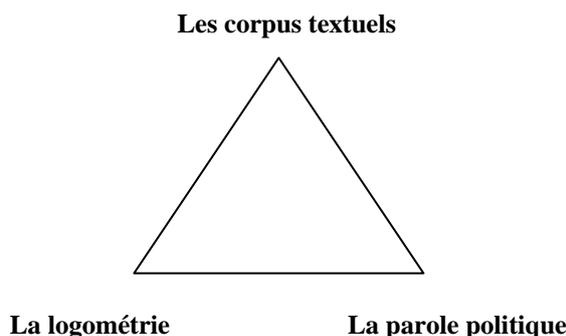
Mes remerciements s’adressent ensuite à Etienne Brunet ; une personne à part ; un être au-dessus.

Merci aussi à tous ceux qui m’ont aidé ou supporté ; aux collègues de BCL, de l’Université de Nice ou d’ailleurs ; à ma famille, Vanessa, Mélisande, Daphnis et Horace.

Nice, 1<sup>er</sup> mars 2010

## Introduction

Un objet (les corpus textuels numériques), une méthode (la logométrie), une finalité (l'analyse de la parole politique contemporaine) : le travail effectué depuis 12 ans, dont nous voulons ici faire la synthèse et souligner les perspectives de recherche, se trouve enfermé dans un triangle que l'on peut représenter ainsi :



Sans forcer le trait et donner au parcours une cohérence *a posteriori*, chacune des performances scientifiques que l'on a été amené à produire depuis une dizaine d'années se situe en effet, obstinément, quelque part dans le périmètre de cette figure géométrique ; plus ou moins attirée par l'un des sommets selon que l'angle théorique (réflexion sur les corpus numériques, les textes, les discours), l'angle méthodologique (réflexion sur la logométrie, la statistique textuelle, les logiciels d'ADT) ou l'angle pragmatique (réflexion sur le langage politique contemporain, le lexique et les thématiques des discours de l'entre-deux-guerres, la rhétorique et la grammaire présidentielles sous la V<sup>ème</sup> République) a été privilégié ; plus ou moins attirée, au contraire, par l'orthocentre selon que la mixité entre ces trois dimensions que l'on veut inséparables de la réflexion a été recherchée.

Nos références bibliographiques, ouvrages, articles, comptes rendus de lecture ou éditions de revue<sup>1</sup> se présentent donc classiquement sous forme de liste chronologique (page 2 et 3) mais plus pertinemment sous forme de figure (page 4).

---

<sup>1</sup> Ne seront pas listées les communications orales faites en colloque ou en séminaire : la plupart d'entre-elles ont fait soit l'objet de publications dans des actes, soit ont été reprises et développées dans des articles.

## OUVRAGES

- I-** *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres. Maurice Thorez, Léon Blum, Pierre-Étienne Flandin et André Tardieu (1928-1939)*, Paris, Honoré Champion, 2000, 800 pages.
- II-** *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la V<sup>ème</sup> République*, Paris, Honoré Champion, 2004, 292 pages.

## ARTICLES

- 1-** “L'Histoire mot à mot. Les mots des migrations dans le discours politique des années trente”, *Cahiers de la Méditerranée*, n°54, juin 1997, pp. 159-171.
- 2-** “Une méthode d'analyse de textes en Histoire : la lexicométrie”, *Historiens et Géographes. Bulletin de la régionale de Nice*, n°34, 1998, pp. 10-20.
- 3-** “Histoire politique et politique de l'Histoire”, *La Pensée*, n°319, juillet-septembre 1999, pp. 139-147.
- 4-** “Temps lexical ou temps politique ?”, *Lexicometrica* ( revue électronique), n°2, 2000.
- 5-** “"Nation" et "patrie" dans le discours de droite à la veille de la guerre : quels signifiants pour signifier quel patriotisme ?”, *Les Cahiers de Lexicologie*, n°76, 2000-1, pp. 133-150.
- 6-** “Exalter l'Empire pour désertier l'Europe : le repli impérial dans les discours munichois”, *Revue française d'histoire d'Outre-mer*, n°326-327, 1<sup>er</sup> semestre 2000, pp. 273-287.
- 7-** “La construction du sens en politique : "fascisme" et antifascisme dans le discours du Front populaire”, *Cahiers de la Méditerranée*, n°61, décembre 2000-janvier 2001, pp. 197-207.
- 8-** “History and Information Technology : The French are way behind”, *Lexicometrica* (revue électronique), n°3, 2001.
- 9-** “Discours politique, genres et individuation socio-linguistique”, in A. Morin & P. Sébilot (éds.), *JADT 2002*, Saint Malo, IRISA-INRIA, 2002, pp. 517-529.
- 10-** “1789 / 1917, deux héritages lexicaux concurrents dans le discours révolutionnaire de l'entre-deux-guerres”, *Mots*, n°69, juillet 2002, pp. 65-79.
- 11-** “L'Herméneutique numérique”, *L'Astrolabe. Recherche littéraire et Informatique*, (revue électronique), 2002.
- 12-** “Les corpus réflexifs : entre architextualité et hypertextualité”, *Corpus*, n°1, 2002, pp. 51-69.
- 13-** “Dire son identité politique. Etude du discours politique français au XX<sup>e</sup> siècle”, *Cahiers de la Méditerranée*, n°66, juin 2003, pp. 247-264.
- 14-** “Arbres et généalogie politique. Représentation arborée du discours de Jacques Chirac (1995-2002)”, (en collaboration avec X. Luong), *Histoire et Mesure*, volume XVIII – n°3/4, 2003, pp. 289-311.
- 15-** “L'ordinateur et le texte : pour une herméneutique contrôlée dans les sciences humaines”, *Cahiers du CERES, "Sciences Humaines et Sociales et Nouvelles Technologies"* (Actes de colloques–Tunis 30-31 mai/1<sup>er</sup> juin 2002), Hors Série n°1, Tunis, 2003, pp. 191-202.
- 16-** “Analyse logométrique de la cohabitation Chirac / Jospin (1997-2002). Explication de la défaite de Lionel Jospin à l'élection présidentielle de 2002”, in G. Purnelle, C. Fairon, A. Dister (éds), *JADT 2004. Le poids des mots*, Louvain, Presses universitaires de Louvain, 2004, volume II, pp. 785-792.
- 17-** “Formation(s) discursive(s) et discours politique : l'exemplarité des discours communistes versus bourgeois durant l'entre-deux-guerres”, *Texto !* (revue électronique), juin 2004.
- 18-** “De la lexicométrie à la logométrie”, *L'Astrolabe. Recherche littéraire et Informatique*, (revue électronique), mars 2005.
- 19-** “La rhétorique présidentielle (1958-2002)”, in P. Marillaud et R. Gauthier (éds), *Rhétoriques des discours politiques* (actes du XXV<sup>e</sup> colloque d'Albi - *Langages et signification*), Toulouse, CALS / CPST, 2005, pp. 237-248.
- 20-** “Les corpus politiques : objet, méthode et contenu. Introduction”, *Corpus*, n°4, 2005, pp. 5-19.
- 21-** “Analyse du discours politique et Logométrie : point de vue pratique et théorique”, *Langage et Société*, n°114, 2005, pp. 91-121.

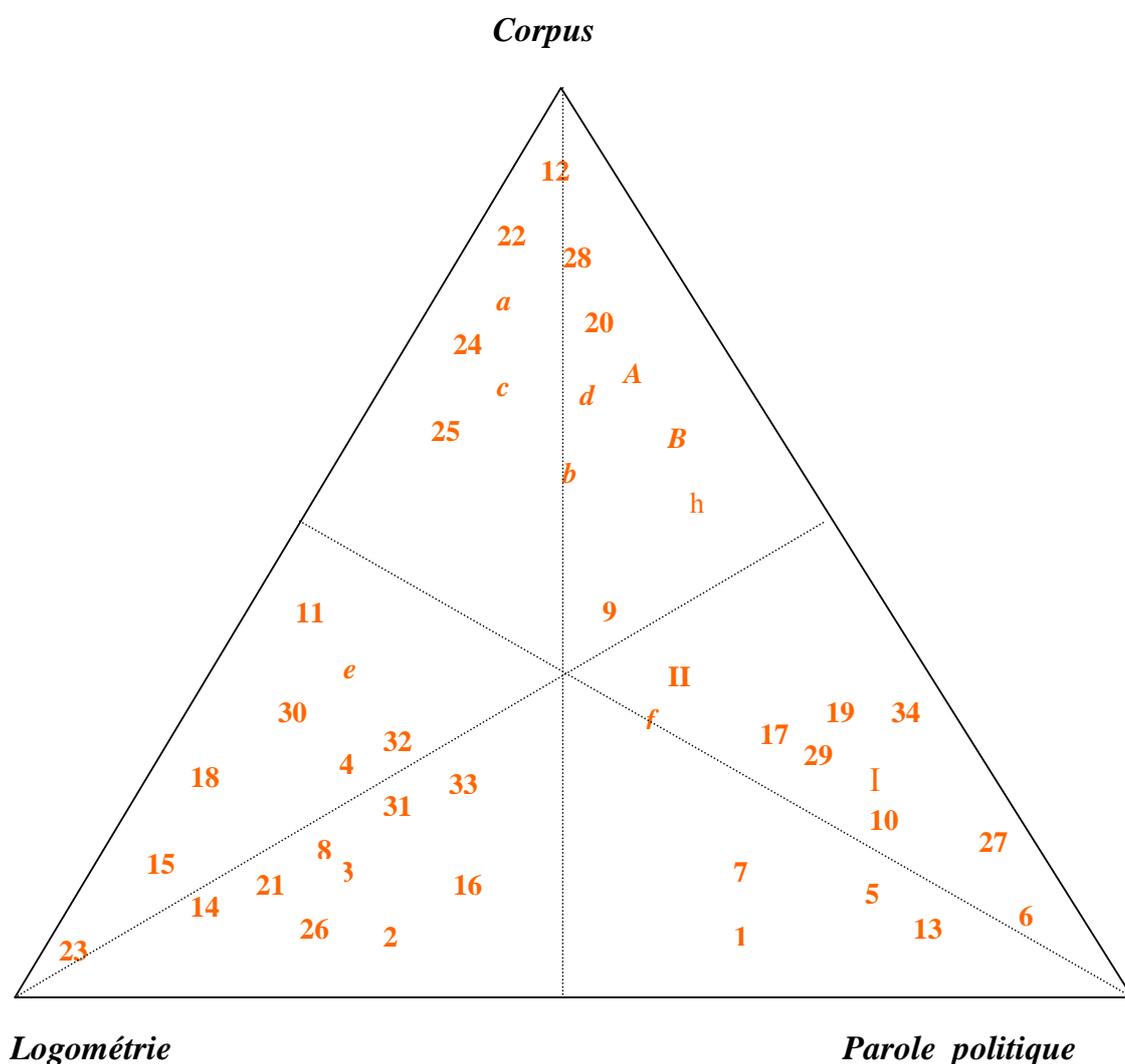
- 22- "Rôle et place du corpus en linguistique. Réflexions introductives", in Pascale Vergely (éd.), Actes du colloque *JETOU'2005*, Toulouse, Université de Toulouse-Le Mirail, 2005, pp. 5-17.
- 23- "Faut-il pondérer les spécificités lexicales par la composition grammaticale des textes ? Tests logométriques appliqués au discours présidentiel sous la V<sup>ème</sup> République", in Jean-Marie Viprey (éd.), *JADT 2006*, Besançon, Presses universitaires de Franche-Comté, 2006, pp. 677-685.
- 24- "Philologie et/ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques", in François Rastier et Michel Ballabriga (éds), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation*, Toulouse, Put, 2007, pp. 15-26.
- 25- "L'analyse de données textuelles aujourd'hui : du corpus comme une urne, au corpus comme un plan. Bilan sur les travaux actuels de topographie / topologie textuelle", *Lexicométrie*, 2007.
- 26- "Analyses logométriques et rhétoriques des discours", in Stéphane Olivési (dir.), *Introduction à la recherche en SIC*, Grenoble, Pug, 2007, pp. 153-180.
- 27- "Vocabulaire et discours électoral de Sarkozy : entre modernité et pétainisme", *La Pensée*, n°352, oct-nov-déc 2007, pp. 65-80.
- 28- "Effervescence autour des corpus", in Michel Ballard et Carmen Pineira (dir.), *Corpus en linguistique et en traductologie*, Arras, Artois Presses Université, 2007, pp. 61-71.
- 29- "Formações discursivas e discurso político : a exemplaridade dos discursos comunistas versus burgueses", in R. L. Baronas (organizador), *Analyse do discurso : apontamentos par uma história da noção-cenário de formação discursiva*, São Carlos (SP), Pedro et João editores, 2007, pp. 47-63.
- 30- "Quand « travail », « famille », « patrie » co-occurrent dans le discours de Nicolas Sarkozy. Etude de cas et réflexion théorique sur la co-occurrence", in S. Heiden et B. Pincemin (éds.), *JADT 2008, 9<sup>es</sup> journées internationales d'analyse statistique des données textuelles*, Lyon, Pul, vol. 2, pp. 811-822.
- 31- "L'entrelacement lexical des textes, cooccurrences et lexicométrie", *Texte et corpus* (revue électronique), n°3, 2008, Actes des Journées de la linguistique de Corpus 2007, pp. 91-102
- 32- "De l'occurrence à l'isotopie. Les co-occurrences en lexicométrie", *Sémantique & Syntaxe*, n°9, 2008 pp. 53-72.
- 33- "Histoire et Linguistique : le redémarrage. Considérations méthodologiques sur le traitement des textes en histoire : la logométrie", in J.-Ph. Genet (éd.), *Langue et Histoire*. 2009 (à paraître)
- 34- "2007. Débats pour l'Elysée", *Mots*, n° 89, mars 2009, pp. 5-9, (en col. avec D. Barbet).

### ÉDITION DE REVUE

- A-** *CORPUS*, n°4, "Corpus politiques : objet, méthode et contenu", 2005, 222 pages.
- B-** *MOTS*, n°89, "2007. Débats pour l'Elysée", mars 2009 (en col. avec D. Barbet).

### COMPTES RENDUS

- a-** François Rastier, *Arts et sciences du texte*, Paris, Puf, 2001 (*Corpus*, n°1, 2002, pp. 171-175).
- b-** Francine Mazière, *L'analyse du discours. Histoire et pratiques*, Paris, Puf, 2005 (*Corpus*, n°4, 2005, pp. 213-217).
- c-** Jean-Michel Adam et Ute Heidmann (éds), *Sciences du texte et analyse de discours*, Genève, Slatkine, 2005 (*Mots*, n°82, 2006, pp. 111-118).
- d-** Frédéric Darbellay, *Interdisciplinarité et transdisciplinarité en analyse des discours. Complexité des textes, intertextualité et transtextualité*, Genève, Slatkine, 2005 (*Corpus*, n°5, 2006, pp. 241-247).
- e-** Margarta Kastberg Sjöblom, *L'écriture de J.M.G. Le Clézio. Des mots aux thèmes*, Paris, Honoré Champion, 2006 (*Corpus*, n°5, 2006, pp. 232-236).
- f-** Jacques Guilhaumou, *Discours et événement. L'histoire langagière des concepts*, Besançon, Presse Universitaires de Franche-Comté, 2006 (*Corpus*, n°6, 2007, pp. 193-198).
- g-** Pascal Marchand, *Le grand oral. Le discours de politique générale de la V<sup>e</sup> République*. Bruxelles : Editions De Boeck Université, 2007 (*Mots*, n°88, 2008).
- h-** Alice Krieg-Planque, *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Besançon : Presses Univ. de Franche-Comté, 2009 (*Corpus*, n°8, 2009, pp. 250-253).



Dans ses détails, la représentation triangulaire rend donc justement compte du travail accompli dans sa triple préoccupation (aucune production n'a été unidimensionnelle ; toutes croisent à des degrés divers les différents points de vue, et essaient de les mettre en dialogue ; le plus précisément possible, les coordonnées des points sur la figure témoignent de l'importance relative des trois composantes théorique, méthodologique et pragmatique de chaque production). C'est pourtant une représentation moins juste mais plus simple qui nous servira de ligne conductrice dans cet essai. Quoique réductrice, elle renseigne globalement sur le mouvement général de la recherche.



Sur cet axe que l'on qualifiera *in fine* d'herméneutique car tendu vers l'interprétation des textes, les trois stations ont fait l'objet, de manière équilibrée, d'effort de réflexion continue. *Grosso modo*, année après année, un article au moins a essayé de faire avancer frontalement la réflexion sur la nature des corpus textuels ou sur la textualité, sur la logométrie ou la statistique textuelle, sur le langage ou la rhétorique politique.

S'il trahit l'imbrication complexe, les va-et-vient ou l'effort de synthèse du travail, ce schéma orienté, à une dimension, présente l'avantage de laisser au lecteur lui-même le soin de hiérarchiser l'importance qu'il convient d'accorder à chaque pôle.

Faut-il considérer le point de départ comme l'essentiel ? Au commencement est toujours le corpus<sup>2</sup>, et, de fait, tout procède de lui. Cette vérité nous apparaît toujours plus criante avec le temps, et au-delà de l'objet empirique que tel corpus constitue pour telle étude, au-delà de la pluralité des textes politiques traités (corpus entre-deux-guerres (1928-1939), corpus présidentiel V<sup>ème</sup> République (1958-2008), corpus cohabitation Chirac / Jospin (1997-2002), corpus campagne présidentielle 2007), nous avons essayé d'avancer des propositions théoriques pour mieux cerner l'objet « corpus textuel », en tant que lui-même, à l'heure du numérique.

Faut-il au contraire souligner l'importance du but ou de la finalité vers quoi tend l'analyse ? Jamais dans la démarche, même lorsqu'elle aura été technique ou au contraire conceptualisante, l'idée de faire concrètement parler le discours politique contemporain ne fut en effet absente. Le langage politique en général, le discours politique particulier d'un Blum ou d'un Thorez, d'un Chirac ou d'un Sarkozy, ne sont pas pour nous un alibi pour expérimenter une méthode ou théoriser une linguistique du texte, mais un champ d'étude d'autant plus passionnant qu'il se trouve aux confins de la science et de la société ; aux confins aussi, de plusieurs disciplines : la linguistique devenue désormais première dans notre réflexion, mais aussi l'histoire et les sciences politiques<sup>3</sup>.

Enfin, ne faudrait-il pas plutôt, *in medio*, remarquer la centralité – au sens propre et au sens figuré – de la méthode dans le parcours, sans laquelle l'objet ne peut être embrassé ni la finalité pressentie ? Non sans emphase, la logométrie a pu être qualifiée

<sup>2</sup> Cette plaisanterie biblique est empruntée à [Peschanski 1984 : 29]. Comme toutes les plaisanteries, elle comporte ses limites. Comme indiqué dans {I : 38}, l'hypothèse de travail précède (ou *préside à*) la constitution du corpus. Par ailleurs, notre propos sera de montrer ici que le corpus est un objet construit qui fournit à l'analyste des observables : une théorie des observables pertinents ou impertinents doit précéder donc, ici aussi, la mise en corpus.

<sup>3</sup> La question de la pluri- ou de l'inter- disciplinarité ne sera pas traitée de front dans cet essai. Un peu car le sujet est trop complexe, un peu car il peut être sensible pour un chercheur qui en est partie prenante. La pluri-/inter-/trans-disciplinarité peut hautement se théoriser (cf. l'ouvrage remarquable de [Darbellay, 2005] dont nous avons pu faire un compte-rendu {d}), mais elle peut aussi, plus modestement, se vivre au quotidien ; sans garantie de confort scientifique et institutionnel. Résumons donc notre trajectoire en deux phrases : notre thèse a été soutenue en Histoire ; cette HDR prétend l'être en Sciences du langage. Nos préoccupations restent identiques (comprendre la parole politique contemporaine), le centre de gravité disciplinaire de notre recherche a, lui, évolué.

d'epistèmè {21}. Posons en tout cas qu'il ne s'agit pas seulement, pour nous, d'une technè ou d'une empeiria mais d'un mode heuristique fondamental d'accès au sens, modifiant notre posture face au texte, nos habitudes et parcours de lecture, nos scénarios d'interprétation. La logométrie apparaît comme le moyen, aujourd'hui, de l'*herméneutique matérielle numérique* naissante : l'enjeu méthodologique ne saurait être minoré.

Précisément ; le seul intérêt de notre recherche se tient là : dans le refus d'apporter une réponse à ces questions sur le primat de l'un ou l'autre aspect des choses. Il se trouve dans l'acceptation qu'une pratique scientifique aboutie ne peut sacrifier ni la conceptualisation de l'objet, ni l'établissement de la méthode, ni la production de résultats pratiques ; et ceci, quitte à renoncer à l'excellence ou à l'hyperspécialisation dans la formalisation de ces trois domaines. A l'heure de résumer nos propositions et de plaider la cohérence de notre travail, c'est ce que l'on aimerait mettre en avant : cette volonté entêtée, au cœur même des productions, d'avancer sur trois pieds et ne jamais séparer le commencement, la fin et les moyens ; la chose, le procédé et les conclusions ; l'objet donc (le corpus textuel numérique), la méthode (la logométrie) et les résultats (l'étude de la parole politique contemporaine).

Ainsi, le plan que l'on se propose de suivre se trouve-t-il aux antipodes et de l'unité du travail accompli et de l'idée de synthèse que représente l'exercice HDR. Il s'impose néanmoins d'un point de vue didactique. En trois temps – successivement donc, là où tout est corrélé–, en trois parties – séparément donc, là où tout est imbriqué – les principales conclusions dans les trois domaines sus-évoqués seront rappelées pour en souligner le potentiel de recherches futures à mener nous-même ou à encadrer.

## 1. Un objet : les corpus textuels numériques

Les années 2000, durant lesquelles s'est déroulé l'essentiel de notre recherche, ont vu le triomphe du corpus en linguistique ; non pas que l'objet corpus n'ait existé de longue date auparavant, non pas que les linguistes l'aient ignoré jusqu'alors mais au sens où la linguistique *sans* ou *hors corpus* paraît aujourd'hui une spéculation intellectuelle marginale pratiquée seulement par une minorité. Sémanticiens, phonologues, lexicologues, dialectologues, etc. se revendiquent tous du corpus ; même la syntaxe semble désormais concernée (cf. *Corpus*, 9, 2010). Nous avons plusieurs fois {22, 28} fait le bilan de la fièvre qui a saisi sinon le monde en tout cas l'hexagone scientifique au début des années 2000. Depuis lors, le mouvement s'est encore conforté : on lira par exemple le bilan documenté dans [Laks 2008], on consultera les actes du colloque thématique du Cercle belge de Linguistique (22-24 mai 2008) [Mellet et Longrée 2009] ; on mentionnera encore le 23<sup>ème</sup> colloque international du Cercle Linguistique du Centre et de l'Ouest (Université de Poitiers - 5 et 6 juin 2009) intitulé « L'exemple et le corpus : quel statut ? ». A l'échelle niçoise, le succès de la revue *Corpus*, créé en 2001, et à laquelle nous avons essayé de contribuer, apparaît révélateur de l'engouement partagé. A Lorient, les Journées annuelles de Linguistique de Corpus (JLC), à côté des Journées internationales biannuelles d'Analyse de Données Textuelles (JADT), semblent s'être imposées dans le concert national et européen.

Cependant, loin de tirer un bénéfice direct du triomphe du corpus, la linguistique de corpus *stricto sensu* au sens par exemple de [Aijmer and Altenberg (ed.) 2002 ;

Biber, Conrad & Reppen 1998 ; Habert, Nazarenko et Salem 1997 ; Kennedy 1998, Partington 1998, Sinclair 1991, Tognini-Bonelli 2001 ; Rastier 2001, 2005-a, 2005-b, Williams 2005, etc.]<sup>4</sup> s'en trouve ébranlée. La banalisation de son objet a brouillé voire dissout son identité propre, comme le triomphe de la république en France à la fin du XIX<sup>ème</sup> siècle a vidé la gauche radicale-socialiste de sa raison d'être.

Dans ce cadre, notre travail a essayé de (re)spécifier l'objet « corpus textuel numérique » (dans sa dimension textuelle comme dans sa dimension numérique). C'est le sens indirect de la plupart de nos productions, et celui direct de {12, 20, 22, 24, 28}<sup>5</sup>.

### 1.1. Le corpus comme matrice

Travailler sur corpus, c'est vouloir travailler sur des données attestées ; sans quoi l'introspection suffirait. Comme les seules données langagières attestées – c'est-à-dire les seules performances linguistiques qu'un locuteur produit lorsqu'il s'exprime – sont les discours<sup>6</sup> puis, en tant que formes empiriques et stabilisées du discours, les textes, nous ne considérons, pour notre part, que les corpus textuels.

Peut-être est-ce la nature nécessairement textuelle des corpus traités qui modifie en substance les choses ?

Avec François Rastier, nous pensons qu'un texte n'a pas de signification, qu'une grammaire du texte est vaine, que les unités textuelles (quand bien même reposent-elles sur les formes matérielles graphiques repérables sur la chaîne) sont plurielles, mouvantes, complexes, parfois discontinues et opèrent à différents niveaux de granularité linguistique. Un texte n'a pas de signification mais un sens (ou plutôt des sens) qu'il ne s'agit pas de re-trouver mais de co-construire dans des parcours de lecture contrôlés. En linguistique du texte, il s'agit donc moins d'établir ou de restituer, que d'interpréter<sup>7</sup>.

Partant, un corpus de textes n'est pas une *base de données* à interroger : le sens ne se laisse pas enfermer dans une *base* ; le sens n'est jamais *donné*. Les corpus textuels

---

<sup>4</sup> Il va de soi que ces auteurs, entre eux, divergent sur certains points. Nous les rassemblons ici en tant que linguistes ayant pris à bras le corps l'objet corpus et l'idée d'une linguistique de corpus. Quelques grands ancêtres pourraient être ajoutés comme Firth ou Palmer.

<sup>5</sup> Dans la chronologie de la production, notons donc que la question du corpus n'apparaît centralement qu'en 2002 avec {12}. Dès 1998 néanmoins puis {I : 38-51} treize pages lui sont consacrées.

<sup>6</sup> « Discours » au sens large donc, c'est-à-dire aussi bien des monologues que des dialogues ou des séquences interactionnelles. Nos corpus recueillent majoritairement des allocutions solennelles, des discours de tribune, des tribunes de presse ; ils recueillent aussi parfois des interviews et des débats à plusieurs voix.

<sup>7</sup> La dimension herméneutique de la linguistique (textuelle) est, on le comprend dès à présent, au cœur de notre réflexion. Elle est au centre de l'œuvre de Rastier et particulièrement de [Rastier 2001] dont nous avons pu faire un compte rendu {a} : nous sommes marqué par le programme qui ouvre le chapitre IV *Herméneutique matérielle* : « Il reste à unir, au sein d'une sémantique des textes, les acquis de la philologie et de la linguistique comparée, pour restituer aux sciences du langage leur statut de disciplines herméneutiques » (p. 99). Nous retrouvons le même type de programme sous la plume de [Adam 2008 : 30] lorsque l'auteur parle « du tournant herméneutique et plus largement de l'ouverture de la linguistique à l'interprétation ». Nous verrons dans une note ultérieure que notre dette scientifique a été contractée, en amont de ces linguistes du texte, envers Jacques Guilhaumou. En France, c'est lui qui a engagé l'école française d'Analyse du discours dans ce *tournant herméneutique* décisif.

apparaissent ainsi très différents de certains corpus-ressources lexicographiques ou phonologiques qui consignent dans de vastes tableurs leurs données. Par la nature de ses composants – les textes –, le corpus textuel n'est donc pas une banque<sup>8</sup> ou une base de *data*. C'est un lieu construit où s'échafaude le sens, où se scénarise l'interprétation. En d'autres termes, nous avons essayé d'établir que le corpus est *moins le réceptacle du sens que sa matrice* {20 : 10}, moins un observatoire d'une langue qui serait déjà-là, qu'un observé vivant, mouvant, dynamique qui par sa constitution même et par son organisation, produit un sens toujours à inventer.

Cette première affirmation semble avoir des conséquences majeures. Nous en relèverons immédiatement deux au cœur de notre travail.

D'un point de vue méthodologique d'abord, la méthode pour traiter ce corpus-matrice – la logométrie – prendra une valeur heuristique plus que probatoire : interroger plutôt que prouver, interpréter autant qu'établir. Notre travail s'inscrit à l'intérieur d'une linguistique de corpus à vocation herméneutique – l'herméneutique matérielle numérique – et non dans le traitement automatique de la langue. Si la logométrie est, loin de l'impressionnisme, une méthode formelle ou formalisante s'appuyant fermement sur le matériel du texte, elle formalise moins des données que des parcours interprétatifs.

D'un point de vue épistémologique ensuite, le corpus-matrice redéfinit notre posture face au texte et le sens de la démarche. A moins d'aspirer à une démarche paradoxale et circulaire, qui consisterait à supposer (à « hypothéser ») un sens déjà-là qu'il suffirait de rechercher et d'établir, tout en prétendant que le corpus produit un sens qu'il nous resterait à inventer et à découvrir, notre mode heuristique demande à être renversé. A *minima*, nous parlerons, avec [Tognini-Bonelli 2001] et avec toute la littérature anglo-saxonne, d'études *corpus-driven* (*versus* d'études *corpus-based*) : un corpus qui n'est pas une ressource que l'on soumet à l'interrogatoire mais un objet qui nous dirige dans le questionnement. Plus hardiment, nous parlerons d'un retournement de la méthode hypothético-déductive qui domine en SHS : là où l'on interrogeait *top-down* le corpus, l'on se propose en effet de se laisser interroger *bottom up* par lui. Posons d'ores et déjà le principe naïvement avant de développer l'idée en détail plus loin (*infra*. 2. *Une méthode : la logométrie*).

## **1.2. Le corpus comme contexte**

Matrice du sens, le corpus l'est car, dans son entier, il informe les textes qui le composent. Principe d'architextualité ou détermination du local (dans ce cas, le texte) par le global (dans ce cas, le corpus) dirait [Rastier 2001 : 92] ; condition d'une comparaison différentielle non hiérarchisée des textes diraient [Adam et Heidmann 2005 : 102 et ss.].

Soulignons simplement ici que cette affirmation théorique et récente, que la plupart des auteurs de linguistique textuelle ou de linguistique de corpus revendiquent

---

<sup>8</sup> Ainsi, contrairement à la terminologie anglo-saxonne, nous distinguons clairement *banque* de textes (ressource matérielle collective dans laquelle le chercheur pourra puiser ses textes) et *corpus* de textes (objet construit de manière *ad hoc* par lequel le chercheur problématise sa recherche). Le *British National Corpus*, le *Brown Corpus* ou le *Lancaster-Oslo-Bergen (LOB) Corpus* relèvent pour nous des banques de textes au même titre que Frantext, la BMF ou le LASLA. (Ceci indépendamment du fait que corpus ou banques de textes doivent compter les textes intégraux et non pas des extraits de textes).

désormais, épouse le principe même, technique et originel, de la logométrie : l'idée d'un corpus-norme ou d'un corpus-référence, l'idée d'une statistique endogène ; comme si cette méthode était idéale pour ces linguistiques, comme si la méthode avait devancé, en pareil cas, la théorie. A la statistique endogène de Guiraud, Muller ou Brunet a en effet répondu une « stylistique endogène » [Viprey 1997], ou une « lexicologie textuelle » endogène au texte [Valette 2008] : finalement, la linguistique de corpus n'est rien d'autre qu'une *linguistique endogène*, et son outillage par la statistique (endogène par essence) apparaît naturel aussi bien chez [Biber 1988, 1995, 1998], [Habert, Nazarenko et Salem 1997] que [Malrieu et Rastier 2001].

Matrice du sens, le corpus entretient, dès lors, un dialogue direct avec la notion de *co(n)texte* puisque l'on admet que la *co(n)textualisation* est la condition de la maïeutique du sens. Le sens naît en/du *co(n)texte*, avons-nous plusieurs fois écrit {24, 30} en soulignant la vanité d'une linguistique ou simplement de pratiques méthodologiques décontextualisantes. Le sens naît en/du corpus pourrait-on renchéris liant corpus et contexte dans une relation étroite, quasi-synonymique.

Le corpus peut être en effet conçu comme une forme privilégiée du contexte. Plus précisément, nous définissons le corpus comme *la forme maximale du contexte* {30, 32}. De la lettre au mot, du mot à la phrase, de la phrase au paragraphe ou à la partie, de la partie au texte, du texte au corpus : le phénomène de *co(n)textualisation* (ou, autrement dit, l'extension de l'objet du linguiste) semble devoir s'arrêter là.

Précisons cependant : en définissant le corpus comme forme maximale du contexte, nous entendons forme maximale *formalisable* du contexte, car le contexte (le *co-texte* proche ou l'*intertexte* plus lointain) est insondable et à proprement parler *insaisissable*. Plus loin encore, le contexte lorsqu'il s'étend au-delà du *co-texte* ou de l'*intertexte* pour toucher à la situation socio-historique et aux conditions de production des discours est une chose qui échappe pour partie aux études strictement linguistiques<sup>9</sup>.

Pour un texte donné, donc, une infinité de corpus pourront être construits formant autant de *co(n)textes* maximaux, au sein desquels seront écrits autant de scénarios interprétatifs. Le corpus donne un corps – un corps linguistique – au contexte, et hors de la réification qu'il propose, le *co-texte* ou le contexte, l'*intertextualité* ou l'*interdiscursivité* sont un magma sans forme, sans aucun doute passionnant mais insaisissable. Dit plus simplement, le corpus représente pour nous la forme empirique ou matérielle du contexte ; celle qui accepte de se soumettre à l'observation linguistique. C'est le cadre matériel, immédiat, formalisé, d'une interprétation contrôlée.

### **1.3. Corpus réflexifs et herméneutique endogène**

Pour répondre à cette double définition – le corpus comme matrice du sens ; le corpus comme objectivation du contexte linguistique nécessaire à l'interprétation–, les corpus textuels doivent être bien formés.

Outre les critères désormais bien connus d'équilibre, de représentativité ou d'exhaustivité, d'homogénéité (notamment générique), de contrastivité et de clôture,

---

<sup>9</sup> Insistons encore : nous avons plusieurs fois écrit que le corpus comme *forme maximale du contexte* était un idéal qui nous paraissait intéressant et nécessaire pour réguler nos pratiques, mais dans l'absolu inatteignable ; c'est pour nous une tension ou un objectif plus qu'une réalisation définitive.

etc., nous avons proposé non seulement que les corpus soient gros pour offrir un cadre suffisant<sup>10</sup> mais structurés de manière adéquate.

La proposition qui a été la mieux reprise par la communauté scientifique est la nécessaire *réflexivité du corpus*. Sans revenir dans le détail sur une idée d'abord pressentie dans notre thèse {I} puis développée dans plusieurs articles {notamment 12 et 24}, posons dans la continuité de ce qui précède que l'enjeu du corpus réflexif est de constituer un ensemble sémantique auto-suffisant qui internaliserait les ressources contextuelles nécessaires à l'interprétation de chaque texte. « En miroir, les textes du corpus doivent s'éclairer mutuellement ; se *réfléchir* les uns les autres ; chacun d'entre eux constituant le co-texte immédiat de tous, et l'ensemble, l'intertexte de chacun. » {24 : 21}

Comme l'importance de la notion a déjà été soulignée précédemment, avançons ici l'idée que le corpus réflexif est la condition d'une d'herméneutique *endogène*. C'est au sein du corpus que les parcours interprétatifs sont proposés ; le corpus réflexif dans son ensemble formalisant l'intertextualité – une intertextualité parmi d'autres possibles – des textes constituants, soumis à l'analyse<sup>11</sup>.

Notons que ce co-texte ou cet intertexte que les corpus réflexifs entendent manufacturer peut apparaître objectif : il s'agira de traiter ensemble par exemple deux textes contemporains jugés comme apparentés historiquement – le premier constituant le co-texte immédiat du second, le second le co-texte immédiat du premier –, comme par exemple les discours de Jospin chef de gouvernement et les discours de Chirac président de la République, durant la période de cohabitation {II : chap. 9, 224-241, 14, 16}. Mais la réflexivité mise en scène peut être plus subjective et arbitraire et mettre volontairement en dialogue deux textes « étrangers » l'un à l'autre mais dont on suppose que la mise en rapport – le face à face réflexif – au sein du corpus produira des effets de sens et suscitera des parcours de lecture critiques et fertiles. Précisons par là avec force combien le corpus est un objet construit sur la base d'hypothèses de travail, et combien cette construction détermine l'analyse. Nous verrons *infra* pourquoi les hypothèses de travail ne doivent pas présider au traitement linguistique du corpus (cf. 2.3. *Administration de la preuve et contrôle de l'interprétation*) : il est suffisant pour nous qu'elles président au recueil – *moment philologique* [Adam 2005 : 83] – et à l'organisation – *projection herméneutique* – des textes le constituant.

---

<sup>10</sup> Nous n'avons ici fait qu'accompagner les avancées technologiques et une tendance naturelle allant des petits corpus historiques de St Cloud aux macro-corpus littéraires de Brunet. A vrai dire, nous n'avons jamais cherché à définir une taille critique. Nous savons simplement que le traitement statistique est d'autant moins contestable que les populations sont importantes. Et constatons que les ordinateurs et des logiciels repoussent chaque jour les limites de la veille. « More data, better data » pourrait être grossièrement notre devise s'il n'y avait la nécessité ensuite d'embrasser le texte qualitativement : trop gros, les corpus deviennent illisibles et ininterprétables. Puisqu'il faut donner un ordre de grandeur, posons, empiriquement, que nous traitons plutôt des corpus de 2.000.000 de mots que de 200.000 ou de 20.000.000.

<sup>11</sup> Internaliser les ressources interprétatives pour une herméneutique endogène : l'ambition est démesurée mais rappelons le modeste point de départ – un malaise épistémologique – de la réflexion. Une discrimination non acceptable sépare les textes-objets-du-corpus et les textes-ressources-interprétatives. Les premiers font l'objet d'un mode de sélection, d'une attention philologique et, par définition, d'un traitement linguistique minutieux. Les seconds sont convoqués à discrétion, cités à la hussarde et, hors du corpus, échappent au traitement proprement dit. Cette discrimination saute à l'œil car, sources ou ressources, il s'agit bien dans les deux cas de *textes*...

## 1.4. Le corpus comme texte

De la lettre au corpus, en passant par le mot ou la phrase, et en s'arrêtant sur le texte, la linguistique de corpus telle que nous l'entendons procède à/de l'extension de l'objet de la linguistique vers des réalités ou des globalités toujours plus vastes et toujours plus complexes.

Par facilité sans doute, à chaque palier de complexité franchi, l'analyste a eu tendance à se retourner vers le palier inférieur pour en faire remonter des schémas d'analyse qui lui étaient familiers. Ainsi, hier, par exemple, alors même qu'on passait du phrastique au transphrastique, s'est-on imaginé établir – en vain – une *grammaire du texte* comme il en existait une de la phrase.

Ainsi, aujourd'hui, en passant du texte au corpus peut-on envisager expliquer – avec fruit – la « corporalité » comme on explique la textualité ; et précisons que la tâche s'annonce passionnante mais d'autant plus compliquée que la notion de textualité est elle-même à peine stabilisée en linguistique textuelle et objet encore de riches discussions.

Peut-on considérer un corpus comme un texte ? Peut-on considérer un corpus textuel comme un macro-texte qu'il s'agirait alors de traiter avec des outils théoriques en partie balisés par Hjelmslev ou Bakhtine, Hasan, Halliday, Adam ou Rastier ?

Ce sont, au fond, les questions qu'avec la plupart des auteurs contemporains l'on est en droit de se poser<sup>12</sup>. Nos travaux se gardent bien d'apporter une réponse définitive à une interrogation qui engage sinon l'avenir de la linguistique de corpus en tout cas son intérêt actuel. Mais plusieurs indices, comme autant de pistes de réflexion, ont été pressentis. Deux méritent d'être rappelés ; nous verrons qu'ils sont nuancés ; et que dans ces nuances réside un programme de recherche majeur.

### ✓ *Cohérence-cohésion du texte / cohérence-cohésion du corpus*

Si l'on peut supposer que le corpus, à l'image du texte, est un ensemble cohérent et cohésif, il l'est nécessairement de manière différente. En s'aventurant dans le *distinguo heideggerien* sans doute peut-on prétendre que la cohésion-cohérence d'un texte lui est ontologique ; la cohésion-cohérence du corpus lui est ontique. Le texte est cohérent par nature, par essence, par définition<sup>13</sup> ; le corpus l'est par existence – en tant qu'*étant* –, par construction, par hypothèse.

Il ne s'agit pas ici de naturaliser (ontologiser) l'objet texte qui est lui-même un objet construit, artefactuel comme nous l'avons clairement souligné dans {24 ; voir aussi *infra*}, mais de rappeler que le texte existe – sous une forme ou une autre – dans la société sans l'analyse scientifique, là où le corpus existe uniquement en laboratoire par

---

<sup>12</sup> Par exemple, dans un article important, [Viprey 2005 : 62] s'interroge et conclut : « Soit la collection du *Monde Diplomatique 1980-2000*, éditée par le journal lui-même en CD-Rom. S'agit-il bien d'un texte ? Sans doute... ». Seulement, la conclusion positive de l'auteur tient dans le fait que cette collection a été réalisée par la rédaction *elle-même* du journal : qu'en est-il lorsque la collection est réalisée de l'extérieur par l'analyste ? D'autre part, on remarquera que Viprey a ici choisi une *série textuelle chronologique* dont on soulignera *infra* la singularité.

<sup>13</sup> Les « propos incohérents » qu'essayent de tenir certains auteurs n'en peuvent mais. Et c'est alors, précisément, cette incohérence du propos qui fait la cohérence du texte.

le fait du chercheur, et seulement le temps de la recherche<sup>14</sup>. Le texte est un construit, mais un construit social ou culturel « de première main » par le fait du couple auteur-lecteur. La construction du corpus textuel est, elle, de « seconde main » par le seul fait de l'analyste.

La textualité – ce qui fait qu'un texte est un texte – est définitoire du texte et peut être perçue par tout lecteur, *sans quoi il n'admettrait pas qu'il s'agit là d'un texte*. La corporalité, elle, est une pétition ou un parti pris, un espoir, un postulat, le fruit d'un travail singulier ou d'une projection particulière évidente pour le seul chercheur. Exprimé en langage mathématique, la cohésion-cohérence d'un texte est axiomatique ; la cohésion-cohérence du corpus est hypothétique. Bref, un texte qui ne serait pas cohérent-cohésif ne serait plus un texte. Un corpus qui n'est pas cohérent-cohésif est seulement un corpus manquant de pertinence et sans doute d'efficacité heuristique<sup>15</sup>.

La différence est donc importante. Pourtant, elle ne nous paraît pas définitive. Quoique d'une autre nature, la cohérence-cohésion du corpus est l'enjeu de la linguistique de corpus exactement comme la cohérence-cohésion du texte est celui de la linguistique textuelle : c'est cette tension commune vers une textualité / corporalité, conçue avec [Charolles 1995 : 10] comme « **principe général**<sup>16</sup> gouvernant l'interprétation » du texte / corpus, qui rapproche les deux disciplines. Simplement, si du point de vue de la cohérence-cohésion du texte, de [Halliday et Hasan 1976] à [Calas (dir.) 2006] ou [Adam 2008], l'essentiel est déjà réalisé, du point de vue du corpus, l'essentiel reste à faire, même si les travaux de [Viprey 1997, 2005, 2006] sur la micro-distribution des unités et la *texture* balisent une partie du terrain. Et à ce stade, pressentons seulement, d'un point de vue méthodologique, que le parallèle texte / corpus et textualité / corporalité demandera quelques ajustements : si le texte et la textualité peuvent encore être considérés comme des objets *micro* réclamant l'approche qualitative, le corpus et la corporalité, en tant qu'objets *macro*, semblent exiger une approche quantitative.

✓ *Sérialité du corpus / linéarité du texte*

Si texte et corpus (textuel) présentent certaines similarités au point que l'on peut envisager entre eux un simple rapport d'échelle, une différence profonde de structure semble les distinguer : le corpus est fondamentalement un objet *sériel* {12}, le texte est d'abord un objet *linéaire*.

Le corpus est une *collection* de textes réunis sur la base d'hypothèses de travail. Au-delà du stade critique d'une collection de textes qui en compterait un seul, les

---

<sup>14</sup> Ici il faut souligner les échecs répétés d'archives de corpus ou de banques de corpus. Nous avons tous rêvé à un moment donné de sauvegarder nos corpus (et non seulement nos textes) : force est de constater que ces tentatives ont été jusqu'ici vaines. Les corpus semblent destinés à disparaître après l'analyse et la validation de l'étude par des jurys. Sans doute la réutilisabilité des corpus sera un enjeu des années à venir. On notera cependant la pratique éditoriale des *recueils* ou *anthologies* que l'on peut considérer comme une variété laïque du corpus ; ici, ils existent dans la société et sont pérennes.

<sup>15</sup> Le lecteur aura remarqué le parti pris de mentionner ensemble, globalement, la *cohérence* et la *cohésion*. Dans le détail, et de manière hiérarchique, il serait facile de montrer que la *cohésion* du corpus pose plus de problème encore que sa *cohérence*.

<sup>16</sup> Surligné par Michel Charolles. Voir aussi son article moins abouti de 1983 : Coherence as a principle in the interpretation of discourse. *Text*, 3-1, 71-99.

corpus peuvent donc être considérés comme des séries. (Et faut-il souligner encore ici que les séries, en linguistique comme ailleurs, se prêtent bien au traitement statistique ?)

Certes, certaines de nos séries sont ordonnées linéairement. Nous pensons aux *séries textuelles chronologiques* dont André Salem a décrit les caractéristiques [Habert, Nazarenko, Salem 1997 : 207 et ss] et qui par leur *progression* peuvent à juste droit être traitées comme des textes : il s'agit-là d'un champ de recherche à part entière auquel nous nous sommes souvent affronté. Nos corpus ont été le plus souvent historiques, nos perspectives le plus souvent diachroniques, et sur les recommandations d'André Salem nous avons essayé de souligner l'importance du « temps lexical » qui opère, de manière structurante, en pareil cas {4, 25 ; cf. *infra* 3.2. *Temps discursif et temps politique*}

Mais hors des séries textuelles chronologiques, la plupart des corpus-séries n'ont pas de structure linéaire évidente ; de manière significative leurs parties (les textes qui les composent ou des regroupements de textes que l'on aura constitués en parties) peuvent être indifféremment ordonnées sans que le traitement en soit changé. Ainsi dans le corpus de la campagne électorale 2007, traité dans {27, 31, 34 et B}, les textes des candidats Laguiller, Buffet, Royal, Bayrou, Sarkozy et Le Pen peuvent contraster et se singulariser indépendamment de leur ordre de saisie<sup>17</sup>.

Si le corpus est avant tout sériel donc, et non nécessairement organisé linéairement, le texte lui est toujours linéaire. A l'exception de quelques productions surréalistes ou de quelques jeux d'auteurs marginaux en littérature<sup>18</sup>, un texte a toujours un commencement, un prolongement et une fin ; il peut être défini comme une *suite* [Maingueneau 1996 : 81 ou Détrie, Siblot, Vérine 2001 : 349]; et l'élément fondamental de sa lecture est sa progression, pour nous de gauche à droite, de haut en bas. Contrairement aux parties du corpus, les parties du texte (ses phrases, ses chapitres, ses séquences...) ne peuvent être inversées sans remettre en cause l'édifice. Certes, depuis l'abandon du rouleau pour le codex ou le *polyptychon*, rien n'interdit au lecteur de briser par sa lecture cette progression implacable et de papillonner aléatoirement d'une page à l'autre, d'arrière en avant, ou de chapitre en chapitre. Certes encore, aujourd'hui, le numérique permet des lectures hypertextuelles dont la caractéristique est justement de s'affranchir du linéaire : nous reviendrons en détail sur cette révolution majeure puisqu'elle est le cœur même de notre travail. Mais il n'en reste pas moins vrai que la linéarité apparaît irréductible à la textualité et constitue le socle de sa définition<sup>19</sup>.

Suite continue *versus* série discontinue, linéarité du texte *versus* sérialité du corpus : touchons-nous donc cette fois-ci à une différence définitive ? Non pourtant.

---

<sup>17</sup> Cette idée pourrait être nuancée (mais non contredite). On aura en effet noté qu'un *ordre* politique ici s'est imposé à nous. Et lors de l'analyse, certaines distributions semblent renvoyer à cet ordre ou cette *progression* du corpus. Ainsi par exemple constatera-t-on une progression de l'emploi de « patrie » {32 : 63 fig. 1} à mesure que le corpus se *déroule* (au sens politique *et* typographique) vers la droite.

<sup>18</sup> Précisément il s'agit là de jeux, dont la règle sous-jacente est bien la linéarité attendue... et transgressée.

<sup>19</sup> Citons ici la concession définitive du linguiste qui a remis le mieux en cause cette linéarité pour introduire dans le traitement la réticularité : « Nul ne saurait mettre en doute qu'un texte se manifeste dans l'ordre du temps et/ou de l'espace orientés, se caractérise par un début, un milieu, une fin, ordonnés et non interchangeables, et ce à quelque échelle que ce soit de l'organisation macro-séquentielle à la fine succession des périodes » [Viprey 2006 : 74].

Objet linéaire, *d'abord*, le texte est aussi traversé de sérialité et de réticularité : dans le cercle de nos collaborations, c'est l'apport essentiel des travaux de [Viprey 1997, 2005, 2006] que nous avons essayé de reprendre d'un point de vue théorique {24} et pratique {30, 31, 32}. Objet sériel, *en premier*, le corpus est – ne serait-ce que par ce qu'il est composé de textes linéaires – traversé par la linéarité et la séquentialité : au sein de notre UMR, c'est l'apport essentiel des travaux de [Longrée, Luong et Mellet 2004, 2006, 2008], [Mellet *et alii* 2009] et [Mellet et Longrée 2009] que nous aimerions développer après une première tentative en {25}.

Autrement dit, se dessine un double mouvement scientifique qui venant de deux pôles opposés converge en un programme de recherche commun : la linguistique de corpus, partant de la série, vise aujourd'hui à réintroduire la linéarité dans ses traitements. La linguistique textuelle, partant du linéaire, admet la sérialité comme élément complémentaire de son objet. A la suite de Jean-Michel Adam qui, venant du texte, a présenté l'unité de ce programme aux JADT 2006 [Adam 2006] avant d'en esquisser des pistes dans [Adam 2008 : 179-182], nous avons essayé, venant du corpus, d'en souligner quelques enjeux {25}.

Dit en des termes admis depuis longtemps en linguistique, il s'agit de rappeler que toute écriture / lecture (celle d'un texte ou celle d'un corpus ; *a fortiori* celle d'un *corpus textuel*) articule une compétence syntagmatique et une compétence paradigmatique. Longtemps essentiellement paradigmatique, l'approche statistique des données textuelles en général, notre travail en particulier, doit prendre en compte désormais aussi la dimension syntagmatique, séquentielle et plus généralement encore co(n)textuelle de son objet. Dans ce cadre, nos derniers travaux insistent classiquement sur le traitement des co-occurrences {30, 31, 32} car la co-occurrence articule, dans son essence même, un processus sélectif et un processus combinatoire<sup>20</sup>. Plus pertinemment nous semble-t-il, le calcul de la co-occurrence a pu y être conceptualisé comme le premier mouvement d'une statistique lexicale contextualisante, faisant passer nos pratiques, écrivions-nous, de la lexicographie à la lexicologie {30, 31, 32 : 53-54}. Précisément, dans la perspective herméneutique qui est la nôtre, nous avons pu définir la co-occurrence comme *la forme minimale du contexte* (forme minimale calculable) nécessaire (et parfois suffisante) à l'interprétation<sup>21</sup>. Ainsi pouvons-nous prétendre tenir les deux extrémités du contexte c'est-à-dire de la ressource interprétative : à un pôle, le corpus réflexif (forme maximale formalisable du contexte), à l'autre pôle la co-occurrence (forme minimale calculable du contexte)<sup>22</sup>.

---

<sup>20</sup> Concrètement, nous avons essayé de montrer que le calcul des co-occurrences mobilise / produit une liste paradigmatique (le mot pôle et les mots co-présents) et une fenêtre syntagmatique ou co(n)textuelle (sinon un ordonnancement) dans laquelle ces mots se combinent.

<sup>21</sup> En une phrase {32 : 54} : « Constaté que *a* et *b* sont co-occurents c'est contextualiser l'un par l'autre. »

<sup>22</sup> Nous avons choisi de ne pas développer, dans ce mémoire, nos travaux récents sur les co-occurrences {30, 31, 32}. Encore en chantier, ils seront seulement brièvement repris dans la conclusion générale avant d'être synthétisés ailleurs.

## 1.5. Corpus et numérique

A vrai dire, assimiler le corpus à un macro-texte à traiter est moins, aujourd'hui, une option théorique telle que nous venons de la discuter qu'une possibilité pratique offerte par le numérique. De manière générale, l'ensemble de notre réflexion sur les corpus retracée dans cette synthèse se trouve déterminée par le phénomène numérique. Particulièrement, la notion de *corpus réflexif* ne serait qu'une vaine spéculation si le numérique ne rendait pas possible la structuration architextuelle du corpus et une navigation / exploitation hypertextuelle généralisée {12} : ici c'est bien le numérique et lui seul qui organise le dialogue souhaité entre textes et objective l'intertextualité.

Après d'autres, nous avons donc écrit que le passage du papier au numérique ne représente pas un changement technique de support du vecteur principal de la culture humaine (le texte) mais une révolution culturelle et épistémologique (anthropologique aussi sans doute), sans guère de précédent dans l'histoire {24}.

Pour le simple propos qui nous intéresse ici – le statut des corpus textuels –, cette révolution a été résumée par une formule contradictoire qui retourne l'état de l'art traditionnel : le numérique dématérialise le texte et matérialise le corpus.

Là où l'on avait tendance à naturaliser le texte en l'assimilant, dans sa fixité matérielle, à son support physique traditionnel (la page et le livre), la philologie numérique rend évidente son artefactualité {24}. Pluralité des formats et des codages, choix multiples et individuels dans l'affichage, multiplication des niveaux d'étiquetage, d'annotations, d'enrichissement, circulation, sans limite et jusqu'à l'ubiquité, par fichier joint, etc. : tout se combine pour souligner aujourd'hui la volatilité ou la relativité du texte, sa dimension artefactualle, conventionnelle ou culturelle (*i.e* non naturelle). En l'arrachant du scriptorium et de la bibliothèque, en le « défixant » de l'ouvrage papier qui jusqu'ici le supportait, le numérique a définitivement dénaturalisé et dématérialisé le texte, retrouvant ainsi une pratique ancienne de l'Antiquité jusqu'au Moyen-Âge [cf. la réflexion engagée sur la *philologie numérique* par divers auteurs déjà cités : Rastier 2001, chap. III *Philologie numérique* 73-97 ; Viprey 2005]<sup>23</sup>.

Inversement, là où l'on considérait le corpus seulement comme une idée ou une virtualité, le numérique le matérialise, l'incarne, le réifie en le rendant, quelle que soit sa longueur, palpable et manipulable, exploitable et réexploitable, archivable et échangeable<sup>24</sup>. Si le terme de corpus avait tendu à disparaître dans les décennies précédentes, avant de s'imposer aujourd'hui, c'est qu'il était sans grande pertinence et sans contrainte ; flasque jusqu'ici, il est devenu désormais un concept dur. Le corpus était en effet un idéal (« potentiellement tous les textes susceptibles de m'intéresser ») : c'est aujourd'hui un matériau (« réellement, seuls les textes que j'ai *saisis* en machine et que je peux matériellement soumettre au traitement »). Hier encore horizon, il est devenu aujourd'hui, à la faveur du numérique, un continent, dont la clôture constitue une limite mais sur lequel il est désormais possible de circuler. En ce sens, rappelons

---

<sup>23</sup> Cf. aussi la réflexion de [Legallois 2006 : 7] : « Bien sûr, il faut mentionner l'hypertextualité liée à la mutation numérique du texte, qui oblige à une reconsidération de l'unité textuelle et du texte lui-même : en tant qu'ensemble de possibilités de parcours, le texte devient alors *une unité virtuelle* » (souligné par nous).

<sup>24</sup> Nous avons souligné ailleurs la difficulté d'archiver sur le long terme les corpus d'étude, mais le temps de la recherche le corpus numérique peut-être stocké, mobilisé et remobilisé, échangé, confronté à d'autres, etc.

que le développement de la linguistique de corpus (c'est-à-dire le dépassement du texte seul par le corpus, considéré alors comme le macro-objet de la linguistique) est un produit de la révolution numérique. Du *Brown corpus* au *British national corpus* en passant par le *Trésor de la langue française*, des corpus lemmatisés du Lasla aux corpus XMLisés de la *Base de Français Médiéval* en passant par le *Nouveau corpus d'Amsterdam* – sans rien dire des corpus particuliers –, tous les linguistes de corpus travaillent aujourd'hui sur corpus numériques, et tous réfléchissent à des méthodes numériques pour traiter leur objet numérique (cf. immédiatement 2. *Une méthode : la logométrie*).

## Conclusion

Dans un article récent, dans lequel l'auteur fait référence au sein d'une riche bibliographie internationale aux travaux de François Rastier, de Sylvie Mellet, de la revue *Corpus* et, plus modestement, aux nôtres, [Laks 2008] rappelle l'opposition fondamentale entre une linguistique du *datum* et une linguistique de l'*exemplum*. Magistralement, Bernard Laks démontre que les corpus existent depuis toujours en philologie et en linguistique, et qu'il ne saurait y avoir de modélisation du langage sans prise en compte des usages c'est-à-dire de données attestées recueillies dans de vastes compendiums.

Pour pertinente qu'elle soit dans l'histoire des sciences du langage, l'opposition retracée perd de son efficacité aujourd'hui, en opposant tout le monde à personne et en plaidant une cause entendue désormais par tous ; depuis la fin du XX<sup>ème</sup> siècle, il n'y a plus nécessité de dénoncer la grammaire générative et ses *exempla* controuvés, comme le temps des luttes politiques contre le rideau de fer apparaît révolu.

Comme indiqué dans {22 et 28}, ce n'est plus le corpus qui sépare puisqu'il est admis unanimement. C'est son statut heuristique qui fait toujours clivage.

Un corpus rassemblerait des données qu'il conviendrait ensuite d'extraire, de décrire, de modéliser<sup>25</sup> ? Mais la notion de « données », elle-même, nous semble dangereuse. Les données sont seulement *ce que l'on se donne* ironisent [Malrieu et Rastier 2001 : 554 note] ; elles ne peuvent que difficilement prétendre être le contenu objectif de la langue ou un échantillon représentatif de l'infini du langage. Dangereuse par son illusion objectivante, la notion de « données » semble surtout inopérante pour désigner un texte dans son épaisseur et sa complexité, dans son expression que la philologie toujours discute, dans son organisation et sa productivité sémantiques qui restent à découvrir.

Un corpus textuel – puisque tels sont nos corpus – ne donne jamais accès à aucun contenu objectif : il problématise la lecture et organise l'interprétation. Il ne restitue point un sens déjà-là : il le produit. Pas plus que nous pouvons envisager de travailler sans corpus, nous ne pouvons concevoir le corpus seulement comme une chambre froide de dissection ou comme un cercueil de données.

Le corpus est un ensemble dynamique et contrastif pour une sémantique différentielle. C'est un tout vivant, clos sur lui-même mais réflexif pour une

---

<sup>25</sup> Sur ce dernier verbe, nous sommes d'accord avec Bernard Laks pour refuser de laisser la linguistique de corpus caricaturée en béhaviorisme. La prise en compte des usages, l'étude de la Parole n'excluent pas une pensée modélisatrice. L'objectif de la description est la modélisation ; mais les modélisations sans état des lieux ni description nous paraissent hasardeuses.

herméneutique endogène. C'est une composition matérielle car saisie, contraignante par sa clôture, exigeante par sa réflexivité, mais que le numérique aujourd'hui fertilise et anime.

Il n'y a de sens que d'interprétation s'est-on permis d'écrire {24} : les corpus textuels numériques bien formés sont pour nous le lieu positif de cette interprétation et la condition nécessaire de son contrôle.

## 2. Une méthode : la logométrie

La méthode utilisée dans notre recherche depuis 12 ans souffre depuis ses origines d'un défaut d'état-civil. Encore aujourd'hui, l'incertitude planant sur son nom est le principal handicap à sa diffusion. Or, loin de clarifier la situation, sans doute avons-nous contribué à la confusion terminologique existante en proposant en lieu et place des termes déjà sur le marché scientifique (*linguistique quantitative, lexicologique quantitative, lexicométrie, textométrie, grammaticométrie, statistique lexicale ou statistique textuelle, analyse des données textuelles* ou *analyse statistique des données textuelles*, etc.) le terme « logométrie » {11 ; 16 ; 18 puis systématiquement ensuite<sup>26</sup>}. Nous reprendrons ici ce vocable, non par souci d'originalité, mais parce qu'il nous semble le plus ouvert par son premier formant (logos) et le plus précis par son deuxième formant (métrie). *Prendre la mesure du discours* (politique) par tous les moyens possibles, quantitatifs et qualitatifs, et dans toutes ses dimensions linguistiques, sans présager des dimensions utiles et celles impertinentes, tel est l'objectif non borné de notre travail<sup>27</sup>.

La méthode utilisée avec assiduité depuis 12 ans souffre surtout, dans l'usage vulgaire, d'une réduction voire d'une perversion de ses objectifs. Assignée seulement à administrer la preuve chiffrée grâce au décompte des mots, elle manque l'essentiel, se met sous une saine critique de scientisme et se condamne à n'être qu'un gadget convoqué par les littéraires, les historiens ou les politologues en mal de critères objectifs (*i.e.* quantifiés) pour sceller leur démonstration.

La logométrie que nous avons essayé de promouvoir est autre chose : c'est selon une expression martiale utilisée dans {32}, le bras armé de cette herméneutique numérique appelée modestement de nos vœux dès 2002 {11}<sup>28</sup>, qui voit aujourd'hui le

---

<sup>26</sup> Dans le détail en 2002 {11} a été essayé « logomatique ». En 2004, c'est l'adjectif « logométrique » qui apparaît dans le titre de {16} et « logométrie » qui est utilisé au cœur de {II}. En 2005 {18} nous titrons « De la lexicométrie à la logométrie ».

<sup>27</sup> En anticipant sur le développement à venir, rappelons la définition de la logométrie donnée dans {18} : « Ce que nous appelons Logométrie, c'est un ensemble de traitements documentaires et statistiques du texte qui ne s'interdit rien pour tout s'autoriser ; qui dépasse le traitement des formes graphiques sans les exclure ou les oublier ; qui analyse les lemmes ou les structures grammaticales sans délaisser le texte natif auquel nous sommes toujours renvoyés. C'est finalement un traitement automatique global du texte dans toutes ses dimensions : graphiques, lemmatisées, grammaticalisées. L'analyse ainsi portera sur toutes les unités linguistiques de la lettre aux isotopies, en passant par les n-grams, les mots, les lemmes, les codes grammaticaux, les bi-codes ou les enchaînements syntaxiques. ».

<sup>28</sup> L'expression *herméneutique numérique* a été utilisée en 2002 dans le titre de l'article {11}. Notons néanmoins qu'avec la naïveté qui convient aux travaux anté-doctoraux, nos deux premiers articles dès 1997 posent dans leur introduction {1 : 159} et {2 : 9} la question *herméneutique* ; et nous espérons dans la conclusion de notre thèse en 1998 avoir établi une *herméneutique rigoureuse* {I : 750}. Cette posture herméneutique qui se renforcera ensuite, nous l'avons dit, à la lecture de linguistes comme [Rastier 2001], [Viprey 2005] ou [Adam 2008], tient initialement beaucoup aux écrits de Jacques Guilhaumou qui en France a constitué « *l'AD en discipline interprétative à part entière* » [Guilhaumou 2006 : 12 et 25] ; sur le *tournant herméneutique* décisif (quoique remis en cause) de l'Analyse du discours à la française, lire particulièrement [Guilhaumou 1993].

jour sous des formes variées, et apparaît comme l'enjeu des décennies futures pour les sciences du texte. A ce titre, l'objectif premier de la logométrie n'est pas d'administrer la preuve mais de mettre en place un protocole de lecture pour baliser, au sein de corpus tels que définis plus haut, des parcours interprétatifs contrôlables. La logométrie est un mode de compréhension générale du corpus alliant descriptions macro susceptibles par exemple de produire, globalement, des typologies textuelles {**I** : 743 ; **II** : 45, 91, 115, 125, 202, 204, 231 ; **14** ; **16** ; etc.} et descriptions micro susceptibles de traiter, localement, de phénomènes linguistiques pointus du texte (par exemple l'usage des adverbes et de « naturellement » chez Chirac {**II** : 106-109, 160-170}). C'est une approche intégrale du texte alliant, comme indiqué *supra*, posture paradigmatique (création de dictionnaires, d'index, de tableaux de *sélection*) et posture syntagmatique (création de concordanciers, étude de motifs syntaxiques, prise en considération des *combinaisons*). C'est un effort systématique de déconstruction du texte qui permet de traiter statistiquement les régularités et les phénomènes linguistiques saillants, combiné à un effort de (re)contextualisation – jusqu'au retour au texte intégral – pour en construire le sens. Finalement, c'est une lecture révolutionnaire mais non destructrice qui cherche à adjoindre à la lecture *naturelle, linéaire, qualitative, traditionnelle* du texte – si ce n'est que cette lecture se passe sur écran et non sur papyrus, sur parchemin ou sur papier –, une lecture *hypertextuelle, quantitative, tabulaire, réticulaire* que seul autorise le numérique {**24**}. En un jeu de mots douteux proposé dans {**33**}, c'est une lecture totale car *alpha-numérique* du corpus usant dans un même mouvement de l'*alphabet* et du *nombre*, des chiffres et des mots.

## **2.1. Des lexies (lexi\*) au discours (logo\*) : retour sur les unités textuelles**

« Logométrie » s'est imposée à nous *versus* la lexicométrie traditionnelle en poussant l'étude des formes graphiques (lexies\* simples) ou même du lexique (lexico\*) vers des réalités linguistiques d'ordre grammatical, syntaxique, sémantique ou encore rhétorique qui composent, ensemble, le discours (ou logos\*).

Ce mouvement, en lui-même, n'est guère original<sup>29</sup> et ne mériterait pas d'être mentionné s'il ne constituait, au quotidien, la principale mutation de nos pratiques notamment de notre premier ouvrage issu du doctorat {**I**} à notre second sur le discours présidentiel {**II**}; dans cette mutation l'influence de Sylvie Mellet rompue aux corpus latins lemmatisés / étiquetés du Lasla a été déterminante [voir par exemple Mellet et Purnelle 2002], ainsi que celle d'Etienne Brunet dont le logiciel s'articule depuis 1998 aisément sur plusieurs lemmatiseurs (Winbrill, Tree Tagger, Cordial) ; et cette mutation symbolise grossièrement notre trajectoire interdisciplinaire d'abord centrée sur les sciences historiques et par là peu soucieuse de la qualité du matériau linguistique traité (le mot), puis, à partir de notre recrutement par la section 34 et notre intégration à l'UMR, *Bases, Corpus et Langage*, centrée sur les sciences du langage<sup>30</sup>.

<sup>29</sup> puisque [Muller 1963] renonçait déjà au texte brut au début des années 60.

<sup>30</sup> Rappelons en effet à cette occasion qu'après un cursus universitaire complet en Histoire, notre doctorat {**I**} a été encadré par un historien (R. Schor) et soutenu devant un jury exclusivement –à l'exception d'E. Brunet– composé d'historiens (S. Berstein, E. Brunet, J. Guilhaumou, G. Pervillé, D. Peschanski, R. Schor). Si notre thèse a su convaincre, ensuite, le jury de recrutement de la section 34 du CNRS, et la collection *Lettres numériques* de Champion, elle apparaît au détour de certaines phrases bien innocente linguistiquement. On lira par exemple avec le recul nécessaire la section III de l'*Introduction*

Sur ce mouvement vers des objets textuels plus complexes que les formes graphiques, il n'est donc pas nécessaire de s'étendre dans le détail tant la question a été traitée par d'autres<sup>31</sup>. Mais deux conclusions contradictoires peuvent être relevées sous forme de bilan : l'irréductibilité de la forme graphique dans laquelle s'enracine, et techniquement, et épistémologiquement, la discipline ; la pertinence de l'étiquetage (particulièrement morpho-syntaxique) par lequel la logométrie dépasse l'analyse dite de contenu pour s'ouvrir sur un traitement linguistique plus abouti.

### Matérialisme textuel et primauté de la forme graphique

Le traitement des corpus lemmatisés et étiquetés s'est aujourd'hui imposé. Au sein du projet ANR Textométrie (2007-2010) piloté par l'ENS et le laboratoire ICAR, et auquel l'équipe niçoise « Logométrie et corpus politique, médiatique et littéraire » participe avec d'autres, tout le monde envisage les futurs logiciels comme devant dépasser l'analyse des corpus des textes bruts et de ses mots. La logométrie telle que nous l'avons définie et pratiquée depuis {11} prend donc acte de ce dépassement.

Pourtant, dans le cadre de cette victoire, sans modestie, de Muller sur Saint Cloud, nous avons essayé de rappeler {18}, à contre courant de notre propre pratique, que la forme graphique restait non seulement la préférence technique du traitement, mais surtout sa raison d'être épistémologique.

Techniquement, les meilleurs tenants des corpus lemmatisés / étiquetés finissent par reconnaître, après de lourds débats, que l'unité opérationnelle d'un corpus sur laquelle repose le plus souvent le traitement est, par défaut sans doute, à regret peut-être, mais jusqu'à nouvel ordre, le mot graphique. Celui-ci est d'une essence *infra* linguistique très problématique pour l'analyse et il ne saurait être défini comme une unité *naturelle* du texte. Cependant, il peut être considéré comme une unité *consensuelle* – au sens où chacun comprend *grosso modo* de quoi il s'agit – ou une unité matérielle pratique sans laquelle la réalisation technique du traitement serait plus compliquée. Le modèle de conception des logiciels de l'avenir, qui s'est laissé définir par les experts du projet ANR Textométrie au cours de deux années de gestation 2007-2008, repose finalement sur la vision technique du texte comme d'une « trame », composée d'« items » minimaux (ou de « segments » ou de « tokens » ou « d'unités » selon les terminologies) munis de leurs coordonnées. Or ces items minimaux du texte se matérialisent, non exclusivement mais habituellement, non par postulat théorique mais par commodité technique, avant tout, dans les formes graphiques ou *concaténations de lettres comprises entre deux blancs* (ou autres caractères séparateurs comme la ponctuation). Ainsi l'adressage des unités textuelles complexes ultérieures (un lemme, une partie de discours, une catégorie grammaticale, un motif, etc.) s'appuie sur le substrat des formes graphiques. Quant au *retour au texte*, à la navigation hypertextuelle, et tous les phénomènes de recontextualisation en texte plein, conditions même de notre discipline interprétative, ils ne peuvent se faire, par définition, dans les logiciels que nous pratiquons, sans elles. Pour l'heure donc, le texte en formes graphiques ne donne

---

*générale* consacrée à la *Linguistique* {I : 26-38}. Nos lacunes en linguistique restent à ce jour évidentes, mais apparaissent chaque année un peu moins importantes jusqu'à prétendre aujourd'hui soutenir notre HDR en Sciences du langage.

<sup>31</sup> Nous renvoyons le lecteur à la bibliographie proposée dans {18} ; dans cet article nous survolons la question de la lemmatisation.

pas accès à un format de texte parmi d'autres (comme le texte lemmatisé, le texte grammaticalisé, ou un discours de Sarkozy ramené à un jeu d'étiquettes sont des formats de texte parmi d'autres) : le texte sous format graphique est, jusqu'à mise en œuvre de modèles conceptuels plus complexes [cf. par exemple des propositions dans Pincemin 2004 et 2008], la condition technique de tous les autres.

Epistémologiquement surtout, les objections fondamentales formulées par [Tournier 1985 et 1987] et le laboratoire de Saint-Cloud dans les années 1980 restent d'actualité et prennent aujourd'hui une acuité nouvelle. La logométrie (anciennement la lexicométrie) déconstruit le texte en unités les plus froides et les plus stables possibles pour établir des comparaisons objectives entre textes (ou bien à l'intérieur du texte entre les différentes parties, paragraphes, etc.) et amorcer l'interprétation. Dans un mouvement clairement orienté *bottum up*, il s'agit d'appuyer l'interprétation socio-linguistique sur les éléments les moins subjectifs ou contestables du texte : faute de mieux, sur sa surface matérielle brute ou graphique. Partant, toute forme d'interprétation préalable – de trahison ? – du texte matériel d'origine rend dérisoire ou artefactuelle la démarche faisant précéder les prémisses par la conclusion. La lemmatisation notamment est une forme insidieuse d'interprétation : elle consiste à ramener les unités du texte dont on postule pourtant la constitution *en contexte, en discours, en corpus* à des unités déjà constituées *en langue*. Elle suppose non seulement une lecture / interprétation du texte préalable à son traitement, mais un sens canonisé (le lexique consigné dans le dictionnaire) là où le lexicométricien (particulièrement celui du discours politique) pose que la lexicométrie est là, justement, pour amorcer la lecture et l'interprétation, et que le sens n'est jamais canonique mais se joue dans chaque prise de parole. Mieux : le discours (politique) lui-même que l'analyste doit décrypter, se définit avant tout comme un lieu problématique, une tension sociale et linguistique, un enjeu dans lesquels se jouent et se rejouent à chaque fois la signification des mots et le sens des choses. Et, répétons-le, la fonction du traitement lexicométrique des textes est, précisément, d'établir ce jeu et cet enjeu. En un terme qui nous est désormais connu, notre pratique ne peut être qu'*endogène* au discours et au corpus {32} : la convocation, en amont du traitement, du dictionnaire (lequel ? Et, surtout, avec quels pré-arbitrages sémantiques ou pré-interprétations de la part du chercheur ?) ne va pas sans poser un problème indépassable dans le procès de la démarche.

Au-delà de la lemmatisation, plusieurs décennies après Tournier donc, les logiciels d'étiquetage sémantique aux résultats douteux tel Cordial, qui corrompent le texte en amont du traitement, rendent aujourd'hui criante et définitive l'objection : ils prétendent dire le sens du texte que l'analyse logométrique se proposait de chercher ; ils concluent déjà, là où nous nous fixons l'objectif de chercher.

Même si la stabilisation graphique ou orthographique d'un texte est parfois problématique, toujours conventionnelle et demande elle aussi quelques arbitrages<sup>32</sup>, elle l'est infiniment moins que la lemmatisation que l'on sait dans le meilleur des cas

---

<sup>32</sup> Thorez, par exemple, dans les années 20 et au début des années 30 s'applique à orthographier « bolchevik ». Puis l'orthographe sera francisée (« bolchévique ») avec l'affirmation d'une voie française vers le socialisme. Au milieu des années 30, à un moment charnière donc, l'une ou l'autre orthographe se retrouve indifféremment selon le canal utilisé (*l'Humanité, le JO officiel, Les Cahiers du bolchévisme*). L'analyste se trouve obligé d'harmoniser... et cette harmonisation n'est pas neutre.

imparfaite<sup>33</sup>, ou que tout étiquetage sémantique plus ou moins folklorique. Bref, nous avons souligné *supra* le danger de la notion de « données » qui nous paraissent plus souvent des projections que des faits naturels, mais si les données textuelles (qui sont pour nous les observables ou les « entrées » nécessaires sur lesquelles s'appuient l'analyse et l'interprétation) devaient s'incarner le plus objectivement possible sans doute le seraient-elles dans les formes graphiques.

En d'autres termes, si la forme graphique n'est pas interprétable, notamment à cause des phénomènes bien connus de l'homographie, des agglutinations, des lexies complexes et composées, etc.<sup>34</sup>, et demande donc à être dépassée / complétée, il convient de bien mesurer le danger épistémologique que représente le traitement de données déjà traitées, la manipulation de matériaux déjà manipulés : l'interprétation d'unités déjà interprétées. La logométrie doit proposer des chemins interprétatifs et non des tunnels ; elle doit proposer un parcours herméneutique ordonné et non un cercle qui confond le commencement et la fin. Le corpus dans sa nudité c'est-à-dire dans la surface matérielle de ses textes est le point de départ stable du traitement sur lequel s'initie l'interprétation : c'est en ce sens où, faisant référence aux travaux de Szondi, nous avons parlé d'herméneutique matérielle {24}.

L'herméneutique *matérielle* numérique s'appuie sur la forme pour espérer atteindre le fond ; elle étudie l'expression brute du texte pour accéder à son contenu ; elle appréhende le mot en espérant atteindre le sens. Dans un procès bien compris que nous décrirons en détail *infra*, l'observation précède l'interprétation et la nécessaire *problématisation du sens* succède, conditionnellement, à l'appréhension du texte. Et ce qui se laisse observer et appréhender dans un texte est, d'abord, la chaîne graphique ou matérielle de ses mots.

Plus loin encore que le matérialisme textuel que nous revendiquons, notre démarche accepte donc la critique de « positivisme » textuel au sens où l'on envisage mal le « négativisme » textuel comme programme de recherche. Affirmer qu'un texte n'a pas de signification mais des sens à construire dans des parcours interprétatifs identifiés ne veut pas dire, pour nous, renoncer aux marques formelles du corpus comme autant d'entrées positives privilégiées ; comme premières entrées positives dans le cercle herméneutique (cf. *infra* : 2.3. *Administration de la preuve et contrôle de l'interprétation*).

---

<sup>33</sup> Il serait en effet trop aisé de moquer la lemmatisation. Restons-en au jeu de mots de [Brunet 2000] : « Qui lemmatise dilemme attise ». Particulièrement, la question du figement ou semi-figement à laquelle une lemmatisation poussée se frotte nécessairement est, selon l'aveu même de Saussure, insoluble. Faut-il lemmatiser « président-de-la-République » ? Oui bien sûr. Faut-il lemmatiser « président-de-la-France » (employé quasi synonymiquement) ? Sans doute non. Faut-il lemmatiser « droits-de-l'homme » ? Peut-être. Peut-on lemmatiser « déclaration-universelle-des-droits-de-l'homme-et-du-citoyen-de-1789 » ? Difficilement. « Bien-sûr » est lemmatisable. « C'est certain » ne l'est pas quand bien même il prendrait une valeur identique en discours. Le « programme-commun » de la gauche peut-être lemmatisé en 1972 (ou plutôt à partir de 1973, ou peut-être de 1974...) mais il ne l'est pas avant cette date et cesse progressivement de l'être après. Comme dans ce dernier exemple, la question de la lemmatisation devient particulièrement cruelle en corpus diachronique.

<sup>34</sup> Rappelons que le seul vrai problème de l'interprétation est la contextualisation. Aucune unité n'est interprétable sans être contextualisée ; et toutes le deviennent à partir du moment où elles le sont. Dans ce cadre, l'homographie n'est qu'un faux problème et seulement un cas particulier de l'ambiguïté sémantique.

En tout cas, dans le riche champ des méthodes d'analyse de texte, les analyses logométriques que nous entendons désormais encadrer ne seront pas confondues – et la frontière est parfois pourtant mince – avec ces appareillages statistiques qui traitent de données contrefaites et se fixent l'objectif de calculer le pourcentage de telle information que le chercheur aura lui-même projetée sur son texte<sup>35</sup>. Elles essayeront de se distinguer de ces outils d'extraction de balises importées tels que le permettent les tableurs comme Excel ou les moulinettes qui décryptent le XML. L'analyse logométrique est une méthode globale de lecture / compréhension / interprétation du texte. Elle a une dimension philologique et herméneutique {24} et quelques précautions sont nécessaires. L'intimité avec le texte brut ou matériel – sinon « naturel » – est la première d'entre elles. Mieux : le texte dans l'innocence certes relative de son expression graphique, dans l'évidence certes discutable de sa chair, se retrouve techniquement, dans nos procédures journalières, à l'origine de l'analyse (le mot, quelles que soient ses limites bien connues, est l'unité de segmentation minimale consensuelle ou utile du traitement) et à son terme (toute interprétation réclame la recontextualisation des unités en plein texte auquel le traitement finit toujours par nous renvoyer).

### L'avantage de l'étiquette morpho-syntaxique

Plaider en faveur des corpus lemmatisés ou étiquetés est cependant aujourd'hui devenu inutile<sup>36</sup>. Dénoncer la brutalité du texte brut n'est plus nécessaire. Compter et traiter ensemble, comme dans *Le poids des mots* {I}, les occurrences de « parti » dans les syntagmes « vive le parti ! » et « je suis parti » ne peut être plus longtemps justifié. Enfin, plus précisément dans notre travail, faire l'économie de l'étiquetage morpho-syntaxique et des informations qu'il donne sur un texte reviendrait à se priver de trésors linguistiques désormais accessibles à moindre frais grâce aux étiqueteurs sur le marché. Bref, les discours et leurs sens sont des affaires linguistiques sérieuses dont la lexicométrie a peu de chance de rendre compte si elle devait s'en tenir aux seuils des seuls mots et refusait de se muter en logométrie.

Nous ne plaiderons donc pas l'évidence pour seulement illustrer l'avantage de l'étiquetage par deux exemples concrets qui permettront d'anticiper l'exposé sur le langage politique contemporain (*infra* 3. Une fin : la parole politique). En effet, même dans le champ politique, le lexique – sans même plus parler de l'ambiguïté du mot graphique – n'est pas suffisant pour témoigner d'un discours. Souvent, l'idéologie transpire de phénomènes linguistiques autres que les choix lexicaux. Et si l'entrée par les mots ou les lemmes permet d'accéder rapidement à la thématique des textes, l'étude

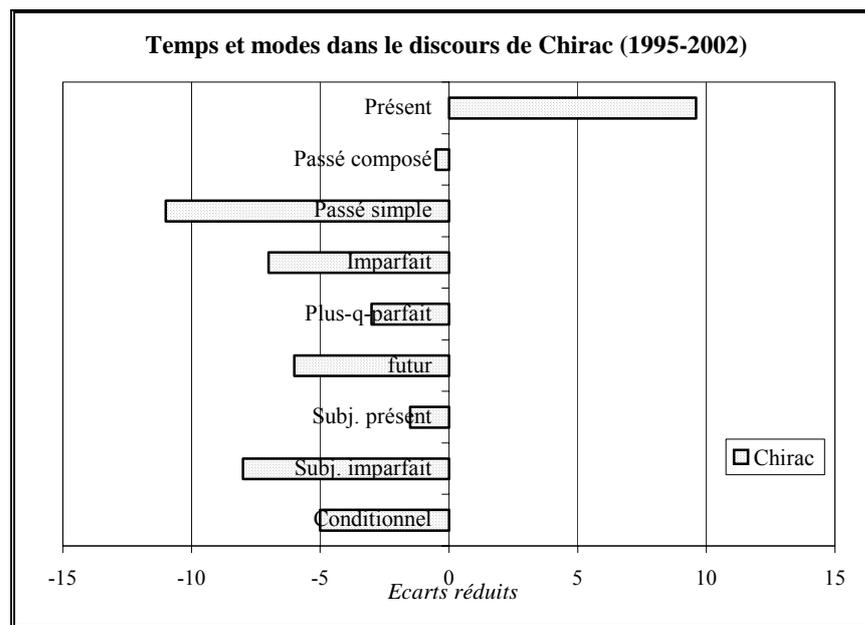
---

<sup>35</sup> « Données contrefaites » : l'expression est certes péjorative. Nous visons là en réalité deux types de balisage fort différents et qui ne peuvent être confondus. Evidemment et avant tout les balisages aveugles issus d'analyseurs automatiques dont il est inutile de souligner les limites. Mais nous pensons aussi à des balisages savants de chercheurs qui s'appliquent à qualifier toujours plus qualitativement leurs données et codent telle unité textuelle selon telle catégorie d'analyse personnelle, puis demandent à la statistique de mesurer leur interprétation (*i.e* leur balisage). Ce second usage du balisage et de la statistique est sans aucun doute pertinent scientifiquement mais il sort, pour nous, du champ logométrique *stricto sensu*.

<sup>36</sup> Le débat s'est déplacé aujourd'hui sur la qualité de l'étiquetage avec une exigence linguistique et philologique forte affirmée par de nombreux auteurs notamment au sein du consortium XML-TEI.

des parties de discours, des temps verbaux, des structures syntaxiques, etc. renseigne tout autant sur la performance socio-linguistique d'un locuteur.

- ✓ L'usage des temps verbaux chez Jacques Chirac par exemple est apparu édifiante : il est vrai que la signature lexicale chiraquienne est pâle et que d'autres indices étaient nécessaires pour caractériser son discours {II}. Après vérification de la fiabilité du décompte du présent de l'indicatif par le couple Cordial / Hyperbase<sup>37</sup>, le constat logométrique apparaît caricatural : Chirac utilise presque exclusivement le présent, et, en tout cas, le sur-utilise massivement par rapport aux autres présidents (illustration 1 ; voir aussi {II : 112-122})



**Illustration 1 : Temps et modes chez Chirac par rapport au corpus présidentiel (1958-2002)**

Avec d'autres éléments comme l'usage redondant des adverbes – et notamment des adverbes déictiques –, ce type de traitement a permis de proposer une interprétation du discours chiraquien tout entier tendu vers la *performativité* (nécessairement conjuguée au présent comme on sait)<sup>38</sup>. Le discours chiraquien est un discours-action plus qu'un discours-référence ou qu'un discours-programme ; et sa principale action ou performance illocutoire – dans le ici et maintenant du discours – consiste à (re)construire l'éthos d'un président (il)légitime {II : 11-13, 185-187, 255-259<sup>39</sup>}.

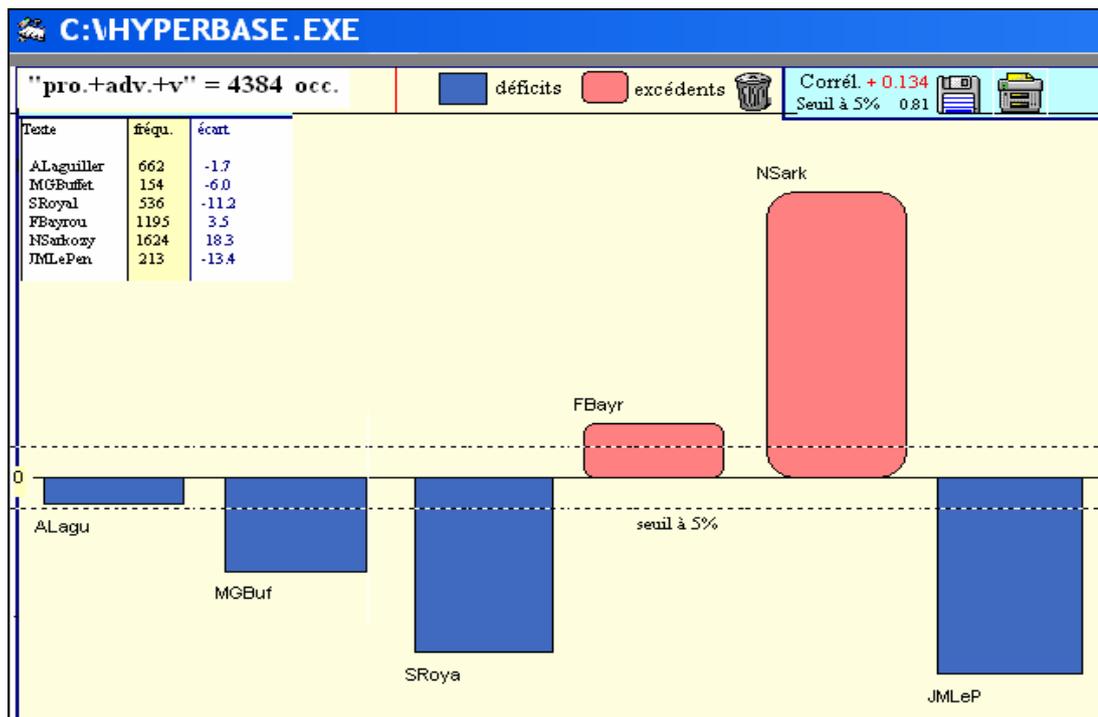
<sup>37</sup> Le passé simple et le présent de certains verbes – types « dire », « conclure », « fuir », etc. – sont en effet homographes aux trois premières personnes, entraînant des erreurs d'étiquetage. Plus grave, l'auxiliaire des verbes au passé composé peut être comptabilisé comme un présent de l'indicatif pour peu que le participe passé qui lui succède ne soit pas contigu.

<sup>38</sup> A rebours et en généralisant le propos, posons que le présent de l'indicatif, dans ce vaste acte de langage qu'est le discours politique, a comme principale fonction ou valeur d'être performatif (voir *infra* 3.3.2 *Temps verbaux*).

<sup>39</sup> On se souvient en effet que cette construction devait être d'autant plus travaillée que l'éthos présidentiel de Chirac – sa crédibilité – n'apparaissait pas évident. Les déchirements fratricides de la

Plus généralement encore, l'usage quasi exclusif du présent trahit la conception même de la politique chez Chirac : une politique comme gestion du présent ou de l'actualité immédiate, plus qu'une narration s'enracinant dans le passé ou qu'une projection tournée vers le futur {II : 112-117} ; en un mot, une politique qui n'est plus une politique ou qui refuse de s'afficher comme telle pour prétendre, au crépuscule d'un siècle fatigué, ne plus traiter que pragmatiquement – et non idéologiquement, et non politiquement – des affaires courantes {II : 156-159}. Sur les ruines de ce déni politique, le discours de Sarkozy étudié en {27, 30, 34, B} contrastera – une rupture ? – et apparaîtra reconstruire faisant explicitement référence à des mots-valeurs ou convoquant un héritage politique fort et syncrétique (Jaurès, Blum, de Gaulle) ; sachant utiliser aussi, pour le propos qui nous intéresse, d'autres temps que le seul présent de l'indicatif.

- ✓ Deuxième illustration de l'intérêt de traiter des unités textuelles plus complexes et mieux qualifiées linguistiquement que les mots : l'étude des structures syntaxiques. La structure [pronom + adverbe + verbe] par exemple est particulière en français car elle est quasi exclusivement celle de la négation. Or loin d'être équi-répartie dans le corpus de la campagne électorale de 2007, elle est une des saillances linguistiques les plus discriminantes du discours d'un des candidats (illustration 2).



**Illustration 2 : distribution de la structure « pronom + adverbe + verbe » dans le corpus de la campagne électorale 2007**

campagne 1995, les reniements électoraux sur la fracture sociale qui poussent aux grandes grèves et à la paralysie du pays dès l'automne 1995, la dissolution extraordinaire de 1997 et le désaveu électoral qui s'ensuit, la cote de popularité de son premier ministre concurrent qui lui fait ombre, les affaires judiciaires enfin qui s'accumulent : tout affaiblit l'image présidentielle durant la période.

Sarkozy n'affirme pas seulement avec force, il nie avec virulence. Discours du « *il ne faut pas* », discours du « *je ne veux pas* », discours du « *il n'est pas vrai que* », discours du « *vous n'avez pas le droit de* » : discours du père sévère à la nation ou du re-père ; discours castrateur {27} qui résume à lui seul l'image du candidat, la psychologie de l'électorat en 2007 et l'idéologie d'une droite décomplexée. « Il faut liquider l'héritage de mai 68 » affirme plusieurs fois Sarkozy<sup>40</sup>. « A bas l'autorité ! » raille-t-il pour résumer l'esprit délétère de cet héritage<sup>41</sup> : Sarkozy se pose en nouvelle autorité sachant dire « non »<sup>42</sup>, sachant dire « ne... pas », sachant réclamer l'ordre et l'obéissance là où mai 68 entendait interdire d'interdire et jouir sans entrave. A travers une simple structure syntaxique que la logométrie repère comme spécifique du candidat – et nulle part mieux qu'à travers elle – se profile le programme politique et sociétal de la nouvelle droite.

### Croiser les différents niveaux

L'historique débat sur la forme graphique, la lemmatisation ou l'étiquetage, et la question de la pluralité des unités textuelles désormais embrassables par le traitement, semblent pouvoir être résolus de deux façons, une *a minima*, l'autre de manière plus ambitieuse.

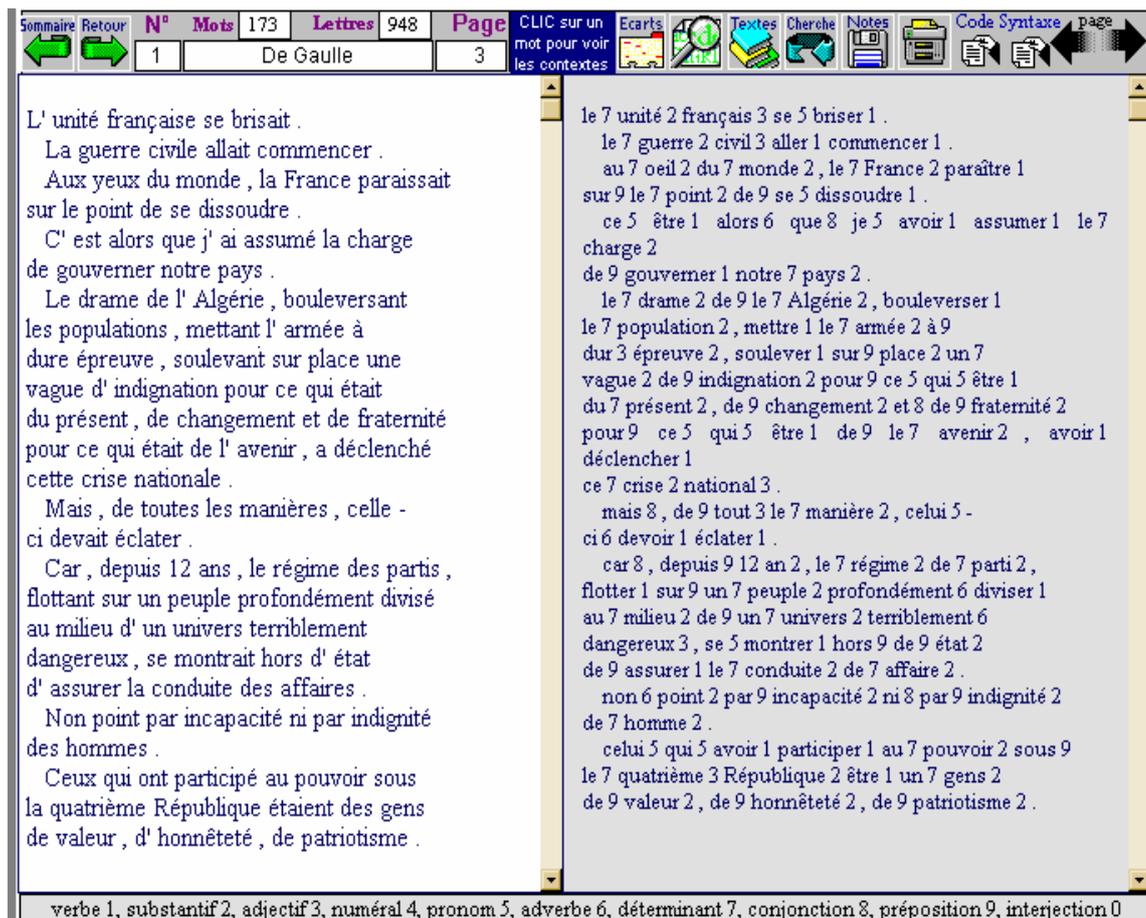
La première solution consiste à juxtaposer les traitements portant sur les différents paliers de description textuel (texte brut, texte lemmatisé, texte catégorisé, etc.). Partant de (et revenant à) la forme graphique, l'analyse s'étend aux lemmes puis aux catégories morpho-syntaxiques et aux enchaînements syntaxiques, puis encore éventuellement à d'autres étiquetages plus contestables. C'est le sens de l'œuvre d'Etienne Brunet [cf. outre ses ouvrages, le recueil de ses articles : Brunet 2009 et 2010], et la philosophie de son logiciel dans sa version lemmatisée qui s'applique dans l'ergonomie même à juxtaposer les différents états du texte laissant au lecteur le soin de choisir l'état qui lui convient le mieux et la possibilité de vérifier la correspondance entre le texte brut et le texte étiqueté (illustration 3).

---

<sup>40</sup> Discours du 23 février 2007 à Perpignan et discours du 3 mai 2007 à Montpellier.

<sup>41</sup> Discours du 23 février 2007 à Perpignan et discours du 17 avril à Metz.

<sup>42</sup> De fait, « non » est une spécificité positive de Sarkozy dans le corpus électoral de 2007 {27}.



**Illustration 3 :** Lecture parallèle du texte brut (partie gauche) et du texte lemmatisé / étiqueté (partie droite) dans Hyperbase (ici allocution de de Gaulle le 13 juin 1958)

A la suite de Brunet, c'est le sens par exemple de l'ouvrage de [Kastberg 2006] et de l'ensemble de nos travaux à partir de 2002 **{11}** : c'est, répétons-le, la définition minimale de ce que nous avons appelé « logométrie » en en faisant plusieurs fois l'état de l'art **{11 ; 15 ; 18 ; 21 ; 26, 33}**<sup>43</sup>. L'analyste, aujourd'hui, sait les fruits à récolter de ce genre de traitement : une performance linguistique accrue, un affinement successif des résultats, une multiplication vivifiante des points de vue lexical, grammatical, syntaxique sur le texte, d'utiles confirmations lorsqu'il s'agit d'établir des grandes typologies textuelles, etc.

La seconde solution est plus ambitieuse et constitue, à notre avis, un des programmes de recherche de la discipline pour les années futures : il s'agit non plus du juxtaposer les traitements mais de les croiser. Les analyses sur les formes, les lemmes

<sup>43</sup> Rappelons au passage la définition donnée dans **{18}** : « Ce que nous appelons *Logométrie*, c'est un ensemble de traitements documentaires et statistiques du texte qui ne s'interdit rien pour tout s'autoriser ; qui dépasse le traitement des formes graphiques sans les exclure ou les oublier ; qui analyse les lemmes ou les structures grammaticales sans délaisser le texte natif auquel nous sommes toujours renvoyés. C'est finalement un traitement automatique global du texte dans toutes ses dimensions : graphiques, lemmatisées, grammaticalisées. L'analyse ainsi portera sur toutes les unités linguistiques de la lettre aux isotopies, en passant par les n-grams, les mots, les lemmes, les codes grammaticaux, les bi-codes ou les enchaînements syntaxiques... »

ou les codes grammaticaux ne doivent plus être seulement additionnées mais factorisées.

Au côté de l'article programmatique de [Pincemin 2004] et des thèses récentes de [Poudat 2006] ou [Loiseau 2006] qui réfléchissent sur la complexité des observables du texte, à côté des travaux de [Luong, Longrée et Mellet 2004, 2006, 2008], [Mellet *et alii* 2009] et [Mellet et Longrée 2009] sur les *motifs* complexes croisant plusieurs dimensions, nous avons essayé d'avancer, d'un point de vue méthodologique, une proposition concrète {23}. Celle-ci n'a pas été reprise jusqu'ici – sinon oralement par Jean Véronis et techniquement par Etienne Brunet qui a bien voulu implémenter l'outil dans son logiciel. Elle nous semble néanmoins ouvrir une porte sur l'avenir et nous comptons proposer des sujets de recherche explorant les possibilités de généraliser le propos.

L'idée fondamentale est de prendre en considération, *dans le traitement statistique même*, les connaissances morpho-syntaxiques du corpus – dès lors qu'elles sont connues – pour en analyser les caractéristiques lexicales ; plus précisément, dans la proposition développée en {23}, les parties du discours doivent servir de base pour le calcul des spécificités lexicales.

Deux sentiments diffus comme prolégomènes à la recherche sont à l'origine de cette tentative qui mérite aujourd'hui prolongement.

D'abord, il y a le constat empirique, partagé par tous les logométriciens assidus, d'une redondance parfois suspecte de l'information que donnent les différents niveaux de traitement. Les grandes typologies textuelles, notamment, apparaissent en effet insensibles à l'unité textuelle traitée : une AFC par exemple discriminera de la même manière les discours de Mitterrand (*versus* les discours de de Gaulle ou de Pompidou) qu'elle fonde ses calculs sur les mots, les lemmes ou les codes grammaticaux. Les tenants du texte brut concluront là à la coûteuse inutilité de la lemmatisation et de l'étiquetage. D'autres affirmeront avec satisfaction tenir la preuve qu'un texte est *une boule* [Brunet 2007] et un tout linguistique indivisible dans ses dimensions lexicales, grammaticales ou syntaxiques. Au-delà des conclusions avancées, il en ressort pour chacun un malaise : nous voulions multiplier les points de vue linguistiques sur le texte et semblons, au fond, compter et constater toujours la même chose. Plus que jamais ici les traitements juxtaposés apparaissent comme de simples répliques là où nous aurions voulu qu'ils soient à chaque fois des originaux.

Il y a ensuite le constat, empirique lui aussi, de la prégnance grammaticale sur le choix lexical ; et ce constat empirique mériterait sans doute un débat linguistique théorique approfondi. Ce que nous avons appelé la structure rhétorico-grammaticale profonde des discours {II ; 19 ; 26} informe en effet – quantitativement – les choix lexicaux du locuteur jusqu'à les rendre – mathématiquement – prévisibles. Un locuteur politique – de Gaulle par exemple – qui tiendra un discours-récit ou discours-histoire au sens de Benveniste avec sa part de narration, de description, de didactisme utilisera plus de substantifs qu'un autre locuteur : et, par là, se donnera plus de chances statistiques – voire se condamnera – à sur-utiliser quelques grands substantifs incontournables du discours politique français (« France », « pays », « politique », « gouvernement », etc.) ; sans rien dire de la sur-utilisation automatique en pareil cas de « la », « de », « les » et de tous les déterminants. Inversement, un locuteur qui aura une rhétorique éristique – Mitterrand par exemple – et tiendra un discours-discours à l'énonciation tendue

prononcera un discours pronominal : et, par là, se donnera toutes les chances de sur-utiliser « je » ou « nous », « vous », « on » ou « tu », etc.

En d'autres termes, un locuteur dont on sait qu'il sur-utilise la catégorie nominale par exemple devrait sur-utiliser, dans les mêmes proportions, chacun des noms du corpus ; et s'il ne le fait pas, cette anomalie devrait être notée<sup>44</sup>. Bref : *la probabilité d'emploi d'un mot ne doit pas se mesurer à l'aune de tous les mots (c'est-à-dire du corpus) mais de la proportion des mots de la même partie du discours de son corpus* ; dit autrement : l'espérance mathématique de la fréquence d'un mot dans un discours est, nous semble-t-il, déterminée par la tonalité rhétorico-grammaticale ou la composition morfo-syntaxique dudit discours.

C'est ce que nous avons essayé de démontrer en développant un exemple précis. Giscard utilise moins de verbes que les autres présidents, et pourtant plus souvent le verbe « indiquer ». La statistique textuelle actuelle même dans son plus solide indice – les spécificités – ne dit rien du caractère extraordinaire de cette situation linguistique. La statistique textuelle telle qu'elle existe depuis 50 ans et telle qu'elle est implémentée aujourd'hui dans les grands logiciels dit : i) les verbes sont une forte spécificité négative de Giscard, ii) « indiquer » est une forte spécificité positive de Giscard. Mais elle ne sait pas mettre en dialogue ces deux constats qui sont pourtant parlants, ici par leur aspect contradictoire<sup>45</sup>.

La solution proposée est la plus simple et la plus radicale. C'est surtout celle qui à l'usage corrige le mieux, selon les tests effectués, les résultats du calcul traditionnel des spécificités lexicales sans les déformer. Partant donc de l'affirmation que *la probabilité d'emploi d'un mot ne doit pas se mesurer à l'aune du corpus mais de la partie du discours dans le corpus* il s'agit simplement de modifier non pas l'algorithme de calcul mais les paramètres de cet algorithme.

Là où nous avons, pour le modèle hypergéométrique présenté par [Lafon 1984], la formule :

$$\text{prob}(x=k) = \frac{f! (T-f)! t! (T-t)!}{k! (f-k)! (t-k)! (T-f-t+k)! T!}$$

et, selon la loi normale et le calcul de l'écart réduit rappelés par [Brunet 2001], la formule :

$$Z = \frac{k - fp}{\sqrt{fpq}}$$

avec

T = étendue du corpus,

t = étendue du texte (*i.e.* de la partie ou du sous corpus),

---

<sup>44</sup> On l'a compris donc, concrètement, si de Gaulle monopolise 20 % des noms du corpus, il devrait monopoliser 20 % des mots « gouvernement », « pays », « nation », etc.

<sup>45</sup> Le plus souvent les constats sont redondants : Giscard sur-utilise les noms ; puis, Giscard sur-utile « pays ». Le propos alors sera de minorer le performance lexicale (l'utilisation de « pays ») en considérant que la composition morfo-syntaxique à dominante nominale la détermine.

$$f = \text{fréquence du mot dans le corpus,}$$

$$k = \text{fréquence du mot dans le texte,}$$

$$p = t / T$$

$$q = 1 - p$$

nous gardons les mêmes formules avec

$T = \text{étendue du corpus}$ , étendue de la partie du discours dans le corpus (c'est-à-dire :  $T_v =$  étendue des verbes dans le corpus ;  $T_n =$  étendue des noms dans le corpus,  $T_{adj} =$  étendue des adjectifs dans le corpus, etc.)

$t = \text{étendue du texte}$ , étendue de la partie du discours dans le texte (c'est-à-dire :  $t_v =$  étendue des verbes dans le texte,  $t_n =$  étendue des noms dans le texte, etc.)

$f =$  fréquence du mot dans le corpus (la fréquence du mot est identique qu'on la compte dans le corpus ou dans la partie du discours du corpus).

$k =$  fréquence du mot dans le texte (la fréquence du mot est identique qu'on la compte dans le texte ou dans la partie du discours du texte).

$$p = t / T \text{ (c'est-à-dire, désormais } t_v / T_v, \text{ ou } t_n / T_n, \text{ etc.)}$$

$$q = 1 - p$$

La communauté scientifique n'a pas encore validé cette proposition. Mais l'usage que nous avons des nouvelles bases du calcul maintenant depuis 5 ans semble indiquer qu'elles sont plus performantes que les anciennes pour rendre compte d'un discours {23 : 679-680}.

Et si nous parlons d'un programme de recherche, c'est que ce que nous avons illustré ponctuellement par le calcul des spécificités lexicales doit se transformer en un travail systématique de dépoussiérage de l'appareil statistique traditionnel. Par exemple, selon la logique de la démonstration, le calcul de la distribution des temps verbaux dans un corpus doit se calculer non pas en fonction de l'étendue générale du corpus mais de l'étendue des verbes dans le corpus ; la proportion d'imparfait dans un discours n'a de sens que par rapport au nombre de verbes conjugués dans ce discours et non de tous les mots prononcés. Autre exemple : la proportions des mots suffixés en *-isme*, si intéressants dans le discours politique, doit se mesurer non par rapport à tout le vocabulaire mais par rapport aux seuls noms. Etc.

Les requêtes documentaires, aujourd'hui, savent croiser les différents états de texte : recherche d'une expression régulière ou d'un motif complexe de type [Déterminant + « peuple » (lemme) + verbe au passé]. De la même manière, la statistique doit désormais apprendre à le faire. Directement inspirée de la statistique lexicométrique, dont elle veut être l'héritière, la statistique logométrique reste en grande partie à inventer.

## **2.2. Quantitatif et qualitatif , macro et micro, global et local**

La logométrie puise sa force et son originalité dans le traitement quantitatif du corpus. Cela ne signifie pas qu'elle s'y réduise.

L'enjeu des analyses logométriques, simple à exprimer, difficile à mettre en pratique, apparaît dans le quotidien de notre recherche : articuler traitement quantitatif et traitement qualitatif, vision macroscopique du corpus et vision microscopique, dénombrement global – qui signifie décontextualisation et appréhension paradigmatique – et lecture locale – qui signifie recontextualisation et appréhension syntagmatique. C'est dans le va-et-vient critique entre un index de fréquences, une liste de spécificités, un tableau chiffré de distances intertextuelles d'un côté, et le texte de l'autre dans sa chaîne entière (texte intégral), comme dans ses extraits (concordances, fenêtres de contextualisation fixes ou coulissantes, etc.) que réside la force heuristique de la logométrie.

L'importance – pour ne pas dire le primat – du traitement quantitatif n'est pas une obsession scientifique visant à objectiver autant que possible le « signifiant » du corpus ou les « données » textuelles. Elle se trouve justifiée par la théorie rastiérienne de la détermination du local par le global [Rastier 2001]. En tant que globalités qui informent ses parties, les corpus textuels traités, par leur taille même, peuvent difficilement se passer d'un traitement macroscopique d'ordre quantitatif ; sauf à prétendre rendre compte des tendances profondes de masses de plusieurs centaines de textes et plusieurs millions de mots par de simples impressions. Plus simplement, nous l'avons dit, en tant que collections de textes ou que *séries*, les corpus ont tout à gagner d'une approche statistique à même de révéler les régularités / irrégularités linguistiques qui les animent à côté des accidents ou des épiphénomènes.

Pourtant, contrairement aux traitements TAL, les dénombrements proposés et, plus généralement, le traitement informatique mis en œuvre n'entendent pas se substituer à l'acte traditionnel de lecture : si le processus interprétatif est outillé par la statistique, il ne peut se réaliser sans le retour systématique au texte et une intimité avec l'œuvre. (Précisons seulement que ce retour au texte et cette intimité avec l'œuvre sont eux-mêmes favorisés par l'informatique et les outils de navigation, de lecture ou de recherche hypertextuelles).

Dès lors, la logométrie apparaît comme une *discipline* non pas au sens académique de champ de connaissance ou de branche d'activité, mais en tant qu'ensemble de règles à suivre garantissant la rigueur des résultats obtenus. Sa pratique, au-delà du va-et-vient fondamental mentionné entre la lettre et le chiffre ou la présence constatée et la distribution calculée, demande de la part du chercheur une posture adéquate face au corpus et de la part des outils mobilisés des fonctionnalités et une ergonomie *ad hoc*.

La posture, d'abord, allie deux principes complémentaires : la centralité du texte et le décentrement du lecteur ; ou, en termes généraux, la centralité de l'objet et le décentrement du sujet. Si le texte dans l'évidence de sa matérialité est central – au commencement et à la fin du traitement –, sa lecture ne saurait être im-médiate. La lecture naturelle du texte se trouve encadrée ou médiatisée par la statistique et l'informatique afin de retarder l'entrée dans la subjectivité interprétative {idée exprimée dès 2, puis régulièrement ensuite}. L'affirmation est triviale : dans les sciences, l'objet (ici le corpus textuel) ne peut être appréhendé par le sujet (ici le lecteur-expert) sans une mise à distance, c'est-à-dire, concrètement, sans la médiation d'un outil (ici les logiciels de logométrie, ailleurs le microscope ou la lunette). En d'autres termes, la proximité vantée entre le chercheur et l'œuvre ne saurait être confondue avec une promiscuité ; pas plus que l'outillage qui sert à mettre à distance l'objet n'est pour nous une finalité

qui nous déracinerait du corpus. Et précisons avec force que ce décentrement du sujet par rapport à son objet, nécessaire partout, nous semble particulièrement indispensable en matière textuelle, tant le texte donne l'impression trompeuse d'être naturel, et le sens l'illusion d'être immédiat.

L'outil, ensuite, en tant que froid prolongement de la main et du cerveau ou en tant que médiateur, prend, dans ces conditions, un rôle non négligeable. Aussi avons-nous toujours illustré nos démonstrations par les fonctionnalités des logiciels utilisés et plus particulièrement par l'ergonomie d'Hyperbase que nous espérons avoir contribué à améliorer. Toute l'ergonomie d'Hyperbase, de l'historique page d'accueil aux fonctions les plus récentes de l'hiver 2008-2009, est tendue vers l'organisation de la double lecture logométrique : lecture quantitative (« fonctions statistiques ») et lecture qualitative (« fonctions documentaires ») (illustration 4)<sup>46</sup>.



Illustration 4 : Page d'accueil d'HYPERBASE (base « Republic »)

Dans le détail du logiciel, chaque outil statistique permet par simple clic de retourner aux textes pour engager une lecture traditionnelle, de même que la lecture traditionnelle du texte plein est outillée par des liens nous renvoyant au dictionnaire de fréquences, à la liste des spécificités, etc. Théoriciens et praticiens du domaine s'accordent pour penser que c'est dans cette mise en scène logicielle, d'apparence technique, que se joue l'essentiel. Des outils logométriques pointus statistiquement mais qui gêneraient le retour au texte amèneraient le parcours interprétatif dans des impasses ; pire encore, favoriserait les sur-interprétations (d'une AFC par exemple)

<sup>46</sup> A côté de ce premier choix ergonomique fondamental (la séparation des fonctions documentaires et des fonctions statistiques), rappelons le second choix ergonomique important : la juxtaposition/comparaison du texte brut et du texte lemmatisé/étiqueté.

voire les mésinterprétations (d'une arborée par exemple). Inversement, ces mêmes logiciels, s'ils n'offrent que des possibilités de recherches documentaires et de lecture, sans instrumentation statistique, seraient aussi inopérants qu'une loupe pour observer l'univers. Sur de gros corpus, les recherches documentaires focalisées en effet sont parfois trompeuses et contreproductives : attestation (locale) n'est pas raison (globale) ; et présence ne veut pas dire significativité. Bref, si l'on veut bien considérer avec [Bachelard 1934 : *Première partie, introduction*] que l'outil est toujours et seulement une *théorie matérialisée*, c'est sans doute par ses outils que la logométrie traduit le mieux sa différence théorique avec la linguistique textuelle traditionnelle d'un côté et le TAL de l'autre. Et pour cette raison, sans avoir de compétences informatiques propres, nous ne sommes pas et ne deviendrons pas insensible aux développements informatiques qui sont menés dans l'UMR 6039 (refonte et réécriture d'Hyperbase) et au sein de projets tel l'ANR Textométrie 2007-2010 (création d'une plateforme textométrique accessible en ligne).

### **2.3. Administration de la preuve et contrôle de l'interprétation**

Lorsqu'elles ne sont pas supercherries, les pratiques statistiques qui prétendent être inférentielles ou probatoires sont des pratiques dangereuses en matière textuelle. Plus généralement, au-delà de notre domaine, de la statistique lexicale et des inférences aveugles et automatiques (attribution d'auteur, arbitrage judiciaire de lettres anonymes, etc.) qu'elle peut parfois entraîner, les méthodes les plus rigoureuses doivent renoncer, dans les Sciences humaines et sociales, à vouloir administrer la preuve là où elles nourrissent seulement la démonstration.

Pour cette raison, nous avons plusieurs fois répété {dès **1**, **2** et **3**, puis **11**, **15**, **18**, **21**, **26**, **33**} que la logométrie a une valeur heuristique plus que probatoire et que son objectif est de contrôler l'interprétation plus que d'objectiver ou d'inférer « le » sens.

Si nous avons parfois insisté, au début de notre recherche, sur la précision descriptive susceptible de trancher des débats historiques sur le discours politique contemporain, la dimension heuristique et la finalité herméneutique se sont imposées chaque année un peu plus au cœur de notre démarche<sup>47</sup>.

Renoncer à prouver paraîtra suffisamment modeste ; prétendre mieux interpréter reste un programme ambitieux.

La pratique logométrique amène en effet à réfléchir à l'herméneutique – au sens linguistique réduit d'interprétation des textes et non dans celui philosophique large d'interprétation du monde–. En lisant Schleiermacher, Szondi ou Bollack, nous nous sommes affronté à son mur c'est-à-dire à son cercle : l'acte interprétatif met le tout et ses parties dans un rapport sans commencement ni fin, le premier expliquant les secondes, la somme des secondes expliquant le premier. De ce cercle aporétique, aucun auteur ne prétend sortir, et notre *position pratique* décrite dans {**24**} a été inspirée par Heidegger :

---

<sup>47</sup> En conclusion de {**I** : 753-754} nous avons en effet narré l'accident scientifique à l'origine de notre conversion à la logométrie ; et cette narration insiste beaucoup sur la dimension probatoire du traitement. Nous rappèlerons encore *infra* le pouvoir descriptif de la logométrie, mais préférons insister désormais sur sa vertu heuristique.

« l'essentiel [...] n'est pas de sortir du cercle, mais d'y entrer de la bonne manière » [Heidegger cité sous des formes différentes par Szondi 1989 : 10 et 105]

Pour notre part, nous entrons dans le cercle herméneutique par le bas c'est-à-dire par le corpus ; par sa matière attestée plus que par son idée supposée, par son expression observable plus que par son contenu (déjà) interprété, par son appareil formel plus que par son dispositif informel : la logométrie est un matérialisme textuel ou une herméneutique matérielle ; c'est une démarche à dominante inductive qui s'appuie sur les éléments positifs du corpus ; et nos pratiques sont essentiellement émergentistes.

La modestie de ces affirmations, moins glorieuses qu'une profession de foi théorique ou que ces démarches hypothético-déductives qui font la part belle à l'intuition du chercheur, nous est imposée de toute part, c'est-à-dire aux trois niveaux de réflexion qui sont synthétisés dans cet essai : réflexion sur l'objet (les corpus textuels), réflexion sur la méthode (la logométrie), réflexion sur la fin (le langage politique).

En amont d'abord, nous l'avons dit, le corpus est décrit comme la matrice active du sens et non comme son réceptacle passif. C'est pour cette raison qu'il convient moins de l'interroger *top down* que de le laisser nous interroger *bottom up*. La dynamique du corpus – pour qu'elle reste une dynamique – ne doit pas être corsetée par des hypothèses de travail trop contraignantes : celles-ci, répétons-le, président à la sélection des textes mais ne doivent pas gouverner l'analyse linguistique. Supposer de manière transcendantale un sens, c'est le plus souvent déjà l'importer là où nous disions qu'il réside de manière immanente et toujours renouvelée dans le corpus, sa composition, son organisation, sa réflexivité. Si le sens naît du bouillonnement contrastif ou réflexif du corpus, il convient de le laisser émerger du tréfonds plutôt que de l'hypothésier de l'extérieur ou du dessus. Notre linguistique de corpus est *corpus-driven* et non *corpus-based* [Tognini-Bonelli 2001]. Affirmer que le sens n'est pas une donnée objective des textes mais le produit de parcours interprétatifs au sein du corpus nous paraît suffisamment subversif pour la sémantique traditionnelle pour que ces parcours interprétatifs eux-mêmes ne se chevillent pas sur le matériel textuel et se trouvent investis, informés, surplombés à tout moment par l'hypothèse ou la thèse du chercheur. Les analyses de discours, notamment, qui selon [Sarfati 2003 : 432] risquent de *manquer le texte en tant que tel* pour favoriser d'un côté les conditions de production des textes et de l'autre, pour ce qui nous intéresse ici, « le point de vue du chercheur sur les textes » ne nous apparaissent pas sans danger.

En aval ensuite, nous le verrons, le discours politique (très) contemporain est un matériau brûlant que l'on a tout intérêt à refroidir. Dans l'analyse, le citoyen perce sous le chercheur, et les sacro-saintes hypothèses scientifiques de travail sont le plus souvent pré-jugés partisans, *a priori* politiques et reflet de l'idéologie dominante du moment : à tout prendre, nous les jugeons non pas comme l'expression de la liberté créatrice du chercheur mais comme l'expression de son aliénation. En tout cas, la projection des hypothèses de travail dans l'objet de recherche est un risque classique en SHS que nul ne peut ignorer ; c'est le danger majeur en analyse du discours politique que nous devons sérieusement prendre en compte. Dans le détail, il est possible de montrer {33} que ce danger se décline de deux manières : trouver toujours ce que l'on cherche (établissement de conclusions artefactuelles ; la thèse est induite par l'hypothèse) ; ne

pas trouver ce que l'on ne cherche pas (ignorance d'éléments importants du corpus, non reconnus faute d'avoir été pressentis comme dignes d'intérêt)<sup>48</sup>.

Au milieu donc, enfin, la méthode logométrique est heuristique et a comme principal objectif de favoriser, par l'organisation des parcours de lecture, l'émergence du sens. Mieux : par le contrôle des parcours interprétatifs – notamment en initiant le mouvement *par le bas*, et en chevillant l'interprétation *sur le* texte – la logométrie entend favoriser non seulement l'émergence du sens mais l'émergence des hypothèses de travail qui nous mènent au sens ; il s'agit donc moins de contrôler l'interprétation en tant que telle que de maîtriser le *processus* interprétatif, toutefois dans son premier et si fondamental mouvement d'entrée dans le cercle.

C'est en effet le plus souvent du traitement systématique et exhaustif logométrique – là où la lecture humaine est aléatoire, partielle, partiale; déjà orientée – que surgissent les hypothèses de travail. En bon ordre – de manière hiérarchique – le traitement logométrique fait remonter du corpus des caractéristiques majeures qui deviennent autant d'interrogations. Les milliers de sorties machine qui ont été consultées depuis 12 ans se sont transformées en effet en autant d'hypothèses de travail, de problématiques, d'interrogations. Entre mille, donc, illustrons le procès de la démarche par deux exemples, le plus lointain avec des conclusions arrêtées (c'est-à-dire un processus interprétatif poussé à son terme), et le plus récent avec des conclusions encore en devenir.

(i) Après comptage exhaustif et systématique de tous les mots d'un corpus contrastif de l'entre-deux-guerres {I : 229 et ss.}, le traitement des spécificités a fait ressortir qu'une des premières caractéristiques lexicales ou graphiques du discours de droite *versus* le discours de gauche dans les années 30 est le mot « a ». Pourquoi ?

C'est de ce constat que l'on peut difficilement imaginer plus épuré, c'est de cette interrogation que l'on n'envisage pas plus élémentaire que vont naître les hypothèses de travail non seulement sur la fonction du verbe-auxiliaire *avoir* dans la langue française mais sur son usage en discours ; usage bizarrement inégal selon l'appartenance gauche / droite des locuteurs dans le discours politique des années 30. Le processus heuristique est ainsi – et pas autrement – engagé, et de questionnement en questionnement, de vérification en vérification, de traitement en retraitement, nous avons pu montrer que la droite des années 30 pensait ou parlait au passé composé (lorsque la gauche parle notamment au futur). Penser et parler au passé composé (et aussi au plus-que-parfait et à l'imparfait) est une constatation non triviale pour la France de l'avant-guerre si l'on veut bien s'essayer à une interprétation socio-linguistique puis historique. La droite de l'avant-guerre tient sans toujours s'en rendre compte un discours passéiste ou un discours de l'accompli, un discours du bilan ou un discours du constat<sup>49</sup> au détriment de toute projection ; et ce discours du constat est d'autant plus douloureux que la droite des années 30 pose ses yeux nostalgiques sur un monde qui s'écroule. De fait, le malaise républicain de la droite, certes originel, prend une acuité nouvelle avec le 6 février

---

<sup>48</sup> Inversement, [Rastier 2009] constate : « Comme souvent, quand l'instrumentation permet un nouveau rapport à l'empirique, on ne trouve plus ce que l'on cherche, et l'on trouve ce qu'on ne cherchait pas ».

<sup>49</sup> Sur-utilisation massive du passé composé donc, sur-utilisation aussi du plus-que-parfait et de l'imparfait, sur-utilisation encore du gallicisme « il y a », ou « il y avait », sur-utilisation enfin des dates : telles sont les sorties-machine sur lesquelles s'initie l'interprétation {I : 229 et ss.}.

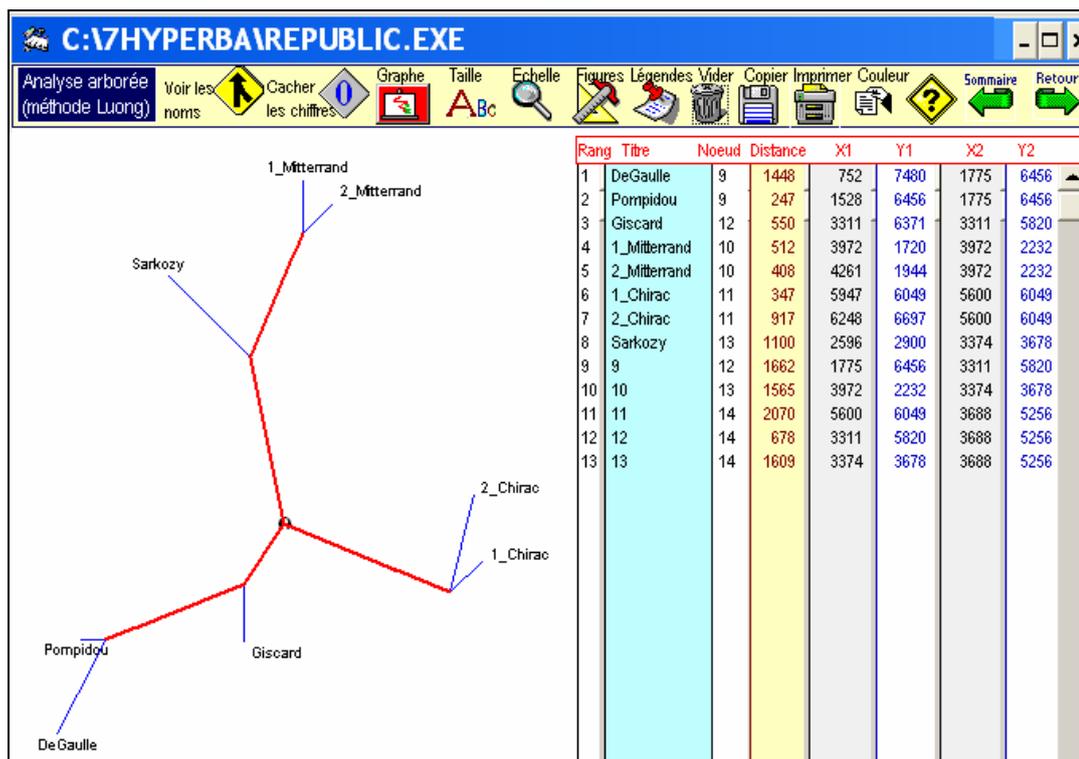
1934, le danger communiste qui s'incarne dans le Front populaire ou la confrontation / répulsion / attirance avec les fascismes : sans horizon politique la droite se retourne alors vers de vieux modèles bonapartistes hérités du siècle précédent. Surtout, la crise économique mondiale qui touche la France à partir de 1930, ébranle les certitudes économiques d'une droite française fraîchement convaincue par le capitalisme industriel : sans horizon économique, la droite se retourne ici encore vers un modèle économique agraire et paternaliste hérité du XIX<sup>ème</sup> siècle. Bref, c'est une droite rétrograde, qui apparaît épuisée idéologiquement, qui voit arriver la guerre, et qui s'abandonnera *quasi* unanimement dans la réaction et le vichysme. Plus loin donc, Vichy, son idéologie, son discours n'apparaissent plus comme des accidents de l'histoire issus de la défaite militaire mais comme l'aboutissement logique d'une classe politique incapable d'imaginer – de dire, de *conjuguer* – l'avenir, et qui se tourne, dès les années 30, vers un *passé* césariste et ruraliste re-composé<sup>50</sup>.

(ii) Sur le corpus contrastif des présidents de la République de de Gaulle à Sarkozy<sup>51</sup>, le global et puissant calcul de la distance intertextuelle [cf. *Corpus 2*, 2003 ], montrera, dans des publications encore à paraître, que le discours de Nicolas Sarkozy est relativement proche du discours de François Mitterrand (illustration 5). Pourquoi ? (Et on comprend que l'interrogation, surgie du corpus et non imposé à lui, est ici d'autant plus intéressante que le constat est contre-intuitif au regard de nos *a priori* sur la différence de culture des deux hommes, différence d'appartenance politique, différence de génération, etc.).

---

<sup>50</sup> Toujours plus loin dans le processus interprétatif, résumons encore une autre conclusion forte tirée, de proche en proche, du constat de la sur-utilisation du passé composé dans le discours de la droite. Le pétainisme comme aboutissement logique du discours passéiste d'un Flaminio Piccoli ou d'un Tardieu ne peut plus être situé dans la généalogie tripartite de René Rémond. Pour [Rémond 1954-1982], le vichysme est l'héritier de l'extrême-droite *légitimiste* : il apparaît pourtant ici comme l'héritier d'une droite *orléaniste* qui refuse le présent et l'avenir. Dit autrement, la tripartition de René Rémond, qui s'applique à dresser trois généalogies à droite, en les supposant étanches, ne tient pas pour la période ; et sans doute pas plus pour les autres. Cette conclusion a été discutée, puis validée, par les historiens de notre jury de thèse.

<sup>51</sup> Le corpus Sarkozy est en cours de constitution et de traitement, au moment où ces lignes sont écrites seuls les discours des 18 premiers mois de sa présidence ont pu être saisis et traités. Les résultats présentés ci-dessous demanderont confirmation.



**Illustration 5 : Distance intertextuelle calculée sur les lemmes. Représentation arborée selon la méthode Luong**

Faut-il envisager une explication d'ordre générique? Sarkozy reprendrait à Mitterrand un mode de communication politique spécifique tel que les interviews à bâton rompu et un genre de discours particulier. L'hypothèse de travail ici est forte car elle poserait à nouveau la question de la prégnance du genre sur une production discursive ; prégnance définitivement démontrée par plusieurs auteurs dont Brunet dans le domaine littéraire, mais que nous nous sommes permis de nuancer – voire de contredire – dans le discours politique en montrant en français {17} et en portugais {29} que le facteur idéologique était plus important que le facteur générique, et que la *formation discursive* façonnait davantage le discours politique que la norme du genre (voir *infra* 3.1.1. *Thorez versus Blum ou la question du genre*).

Faut-il envisager plutôt une explication d'ordre thématique ? Le rapprochement des deux discours viendrait du fait que Sarkozy n'hésite pas à utiliser des thèmes et un vocabulaire de la gauche mitterrandienne. Et si tel était le cas, l'analyse devra être poussée pour mesurer jusqu'à quel point l'emprunt lexicopolitique est important. L'idée de rupture discursive sarkozienne que nous avons nous-même reprise {27, 34, B} devra être alors revisitée : las de rupture, il s'agirait alors d'une continuité insoupçonnée avec un président inattendu.

Faut-il avoir recours à des explications rhétoriques et à une catégorie d'analyse récemment (re)théorisée comme le *registre* [Gaudin et Salvan 2008] ? Pour convaincre, Sarkozy et Mitterrand n'hésiteraient pas à utiliser, l'un comme l'autre, le ton polémique – discours énonciativement tendu avec prise en compte et prise à partie de la pensée de l'adversaire – là où les autres présidents favorisent la tonalité didactique – discours universalisant, détendu, avec effacement de l'interlocution.

Plus simplement et avant tout, quels sont les mots qui expliquent ce rapprochement global imprévu remonté des profondeurs du corpus ? Sont-ce quelques grands substantifs (« pays », « politique », « gouvernement » etc.) ? Sont-ce plutôt des verbes, notamment les énonciatifs, les modaux et les (semi-)performatifs tels « penser », « dire », « vouloir », « falloir » ? Sont-ce encore les pronoms et d'abord le premier d'entre eux « je », lorsqu'un de Gaulle ou un Pompidou utilisent plutôt un « nous-la-France » ? Etc.

Nous laissons ici volontairement ces questions ouvertes pour témoigner du processus heuristique engagé et parce qu'à ce stade nous ignorons la réponse, mais les hypothèses de travail imposées par le traitement logométrique du corpus et non projetés par le pré-jugé de l'analyste ne manquent pas. Et si les conclusions que nous serons amené à tirer resteront à coup sûr discutables, les interrogations qui ont émergé là où elles sont habituellement importées, auront quelque légitimité (et quelque obligation) à avoir été posées.

En de simples mots, et à l'heure de décrire sans maquillage notre pensée, nous avons la faiblesse de croire qu'il est fondamentalement différent, d'un côté, d'hypothéser par exemple une proximité entre les discours de Sarkozy et ceux de de Gaulle et de chercher à confirmer ou infirmer – confirmer le plus souvent ! – notre géniale intuition, et de l'autre côté, de constater (ici grâce au calcul de la distance intertextuelle) un certain rapprochement entre deux locuteurs (ici Sarkozy et Mitterrand) et de réfléchir, sur la base de ce constat empirique impérieux, sur les raisons linguistiques et socio-linguistiques de ce rapprochement<sup>52</sup>.

Le cercle herméneutique est toujours un cercle – c'est ce qui fait sa fertilité –, mais nous y entrons différemment. Dans le premier cas, il s'agira d'objectiver (de démontrer) la subjectivité (l'hypothèse de travail). Dans le second cas, il s'agira de subjectiver (d'interpréter) l'objectivité (un constat linguistique empirique dûment attesté). Impasse symétrique, pensera-t-on, dont les écueils symétriques seraient soit la conclusion artificielle induite par l'hypothèse, soit le bond interprétatif qui nous projette à un moment hors du simple constat<sup>53</sup>. Impasse dissymétrique, affirmons-nous, car seul le second cas aura fait avancer une description reproductible de l'objet.

S'il est probable que les interprétations de tel phénomène linguistique en effet varieront – les nôtres ont toujours été très engagées –, posons *a minima* que la communauté scientifique pourra désormais réfléchir sur des bases descriptives communes : sur-utilisation matérielle du passé composé dans le discours de la droite de l'entre-deux-guerres ; proximité discursive attestée entre Mitterrand et Sarkozy ; sous-utilisation massive du vocabulaire marxiste chez Blum dans l'entre-deux-guerres ; affaiblissement du vocabulaire bolchevik et apparition du vocabulaire jacobin dès 1931

---

<sup>52</sup> Fondamentalement différent d'un côté d'hypothéser – non sans *a priori* historique – une rupture politico-discursive du PCF en 1935 avec le Front populaire, et de voir de l'autre côté émerger un tournant discursif dès 1931, 1932 ou 1933 et de proposer une interprétation de cette mutation précoce (cf. une des principales conclusions de notre thèse {I : 407-418, particulièrement l'AFC p. 477}, voir aussi *infra* 3.2.1). Fondamentalement, différent de pré-supposer – non sans *a priori* citoyens – un pétainisme inavoué chez Sarkozy et de voir émerger, par un traitement systématique des co-occurrences, l'utilisation simultanée de « travail » « famille » et « patrie » dans le discours, et de proposer une interprétation de cette co-occurrence {30,31, 32}. Etc.

<sup>53</sup> Bond interprétatif ? Dans ce deuxième cas, [Peschanski 1989 : 8] parle de la nécessité de « franchir le rubicon de l'interprétation »

chez Thorez ; nominalisation systématique du discours chez Giscard et pronominalisation du discours chez Mitterrand ; sur-emploi de « naturellement » et des adverbes chez Chirac ; co-occurrence statistique de « travail », « famille », « patrie » dans le discours de campagne du candidat Sarkozy en 2007 ; spécificité positive spectaculaire de « je » chez Blum puis chez Mitterrand ; inclassabilité lexicale et discursive du discours de Jospin en 2002 (*versus* son discours antérieur entre 1997 et 2001) ; montée en puissance des segments répétés de longueur importante dans le discours de Chirac au fil de sa présidence (1995-2007) et sur-emploi constaté du présent de l'indicatif par rapport aux autres présidents ; longueur objective de la phrase chez Royal (*versus* la phrase chez Sarkozy) , etc<sup>54</sup>.

Résumons donc pour conclure : la statistique ne prouve rien : une *probabilité* d'emploi n'est pas, par définition, une *preuve*. Elle a vocation non à inférer mais à interroger de manière inductive. Du reste, l'inférence est toujours chose délicate dans une démarche inductive : d'un constat empirique, même bien établi, il est risqué de tirer une loi ou une relation de cause à conséquence définitives ; en matière textuelle, il est souvent imprudent de conclure du général le systématique<sup>55</sup>. En revanche, l'induction s'arme volontiers de la statistique. La remontée d'hypothèses de travail ou de constats, sur lesquels l'on commencera à réfléchir, ne peut en effet s'envisager que si les moyens de description de l'existant (les régularités et irrégularités linguistiques du corpus) sont sûrs et solides ; et sur nos grands corpus de textes, souvent inaccessibles à l'œil et la mémoire humaine, nous envisageons mal de moyens d'exploration plus réglés que l'informatique (dépouillement systématique et exhaustif du matériau brut ou qualifié du texte) et plus solides que la statistique (dénombrements, indices, étude de la macro et de la micro distributions des termes, repérage des récurrences ou des absences, des co-occurrences, etc.). Reste enfin que si le traitement quantitatif a la lourde tâche, par le compte rendu qu'il donne du corpus, de contrôler la description et d'initier le processus interprétatif, c'est par le retour au texte que ce processus pourra aboutir. *Le nombre fait sens* (déchiffrement quantitatif des saillances linguistiques du corpus) ; *le sens naît en/du (con)texte* (lecture contrôlée de la chaîne du texte) : entre ces deux affirmations, que l'ergonomie d'un logiciel comme Hyperbase met minutieusement en forme, se situe notre pratique quotidienne.

---

<sup>54</sup> On l'a compris, nous énumérons ici en vrac quelques constats linguistiques qui nous ont servi d'*entrées* dans le processus interprétatif, et ont abouti, au terme de ce processus, à des interprétations tranchées. Pour épuiser le sujet, répétons donc après {I : 754} et {II : 177}, la règle herméneutique intangible fixée dans notre recherche : *ne jamais rien dire qui ne soit initié par un constat logométrique issu d'un traitement systématique du corpus*. Dans le détail de nos ouvrages, équivalents à 1100 pages tapuscrites, nous ne voyons que 5 pages {II : 177-182} qui contreviennent à cette déontologie ; et nous nous excusons auprès du lecteur de cette contravention.

<sup>55</sup> Précisons par là, qu'en matière textuelle, nous nous méfions des corpus dits échantillonnés. (Un échantillon de texte(s) ?). Lorsqu'un corpus – exhaustif si possible – arrive à témoigner de lui-même, ne lui demandons pas de témoigner, en plus, d'autres réalités. Autrement dit, « le texte est unique en son genre » selon l'expression de Riffaterre rapportée par [Adam 2008 : 13] : il ne peut que difficilement *être représenté* par autre chose que lui-même ; ni *prétendre représenter* autre chose que lui-même.

## Conclusion

L'engagement méthodologique au cœur ou à la périphérie de chacune de nos productions scientifiques depuis 12 ans et que nous entendons poursuivre dans l'encadrement de la recherche paraîtra naïf.

Les textes sont des objets si complexes qu'aucun formalisme ne pourra jamais prétendre les épuiser. Surtout, les textes ont une dimension esthétique : leur science est pour beaucoup aussi un art qui en appelle autant au sentiment qu'à la raison.

Dans ces conditions, s'enfermer dans une pratique logométrique rigoureuse convoquant la machine et les mathématiques, l'informatique et les statistiques peut apparaître réducteur voire contre-productif. De fait, il faut admettre avec le philosophe ou le poète, avec certains historiens travaillant sur textes rares comme certains littéraires sur pièces précieuses, l'inutilité de passer certains corpus textuels en machine pour leur meilleure compréhension.

Pourtant, on ne saurait ignorer que les sciences humaines souffrent dans leur ensemble d'un déficit et de méthode et de crédibilité dans les résultats obtenus. Cela affecte particulièrement les sciences ayant affaire aux textes et à leur interprétation.

La raison de ce déficit méthodologique, auquel nous sommes prêt à répondre par un excès de méthode, est connue. La croyance dans la transparence du texte et l'immédiateté du sens demeure une croyance bien établie. En Histoire, par exemple, l'apparente évidence du contenu dénoncée déjà par [Robin 1973] est toujours aveuglante pour un locuteur-analyste natif, et les méthodes de traitement trop contraignantes peuvent apparaître superfétatoires : c'est le constat que nous avons pu tirer en français {3} et en anglais {8} juste après notre thèse {I}. Les échanges que l'on a pu avoir récemment grâce à Jean-Philippe Genet {33} autour du projet ANR ATHIS (Atelier international Histoire et Informatique – 2006-2009) ne démentent pas aujourd'hui le *French way behind* que nous regrettons<sup>56</sup>. Plus fondamentalement, la subjectivité peut être érigée, parfois à juste droit, en système dans les humanités ; ici en système indépassable d'interprétation des textes. Tout formalisme peut alors être perçu comme de l'objectivisme, et l'objectivisme caricaturé en scientisme : la logométrie, nous le savons, n'échappe pas à cette critique. Plus simplement, enfin, le texte ou le discours peut être considéré seulement comme un média, comme un support, comme un moyen, et non comme un acteur social en lui-même, un *événement* [Guilhaumou 2006], un objet scientifique, une fin ; et au titre de simple média qui transmettrait le monde là où, pour nous, il le construit, convenons qu'il ne nécessite pas de méthode d'analyse plus approfondie qu'une lecture immédiate et intuitive.

Dans ce cadre, et après avoir plaidé de manière complexe les vertus de la logométrie, revenons pour conclure au plus simple. Deux raisons prosaïques mais fondamentales président à notre engagement méthodologique passé et à venir, pour faire de lui une évidence : la taille des corpus traités ou à traiter (cf. *supra* 1. *Un objet* :

---

<sup>56</sup> Concrètement, c'est ce retard qui nous a poussé à postuler en section 34 *Langue, Langage, Discours* du CNRS et aujourd'hui définitivement à soutenir notre HDR en Sciences du langage. Notons au passage que la section 34 a fait alors preuve d'ouverture pluri-disciplinaire ; après le recrutement de Jacques Guilhaumou en 1986, c'était la deuxième fois de son histoire – et dernière fois jusqu'à nouvel ordre – qu'elle recrutait en son sein un chercheur dont la thèse avait été soutenue en Histoire.

*les corpus textuels numériques*) et leur nature politique (cf. *infra* 3. *Une fin : le langage politique* ).

Les petits recueils de petites phrases ou les opuscules de *formules* sont l'objet de journalistes ou d'érudits [sur les *formules* voir l'ouvrage remarquable de Krieg 2009, compte rendu dans {h}]. Notre objet et nos préoccupations sont autres. Nous avons en effet, du mémoire de Maîtrise en 1995 à la thèse que nous co-dirigeons aujourd'hui, cherché à appréhender de gros corpus, souvent exhaustifs, avec l'espoir d'en rendre compte non seulement ponctuellement et localement mais systématiquement et globalement ; *a minima*, avec l'espoir d'y naviguer sans s'y noyer. (Et devant l'inflation textuelle que produit chaque jour notre société numérique, la taille de nos corpus ne devrait pas diminuer). Dès lors, l'assistance de la statistique et de l'informatique – la logométrie – ne se vit pas comme une option mais comme un impératif.

Surtout, le discours politique n'est pas un discours comme les autres (lire immédiatement 3. *Une fin : le langage politique*). Sa fonction sociale l'oblige. Le discours politique a moins une dimension esthétique qu'une dimension pragmatique et plus précisément propagandique. La pertinence d'un discours politique se mesure à son efficacité pratique, jusqu'à sa force perlocutoire. *Et la redondance est la condition de cette efficacité*. La construction de l'*ethos* du locuteur politique par exemple passe par une performance linguistique à chaque fois rééditée de tribune en tribune, de discours en discours, de paragraphe en paragraphe. Les thématiques – en période électorale notamment – pour s'imposer en *agenda setting* doivent être développées puis reprises avec des répétitions qui ressemblent parfois à des bégayements. Les mots, tout simplement, qui apprivoisent la réalité et l'enferment, ne s'imposent comme référents sociaux ou comme *praxèmes* qu'à l'occasion d'une récurrence d'emploi (*récurrence combinée* dans le cadre des co-occurrences) proche du martèlement. A côté de la taille des corpus embrassés, c'est donc la redondance intrinsèque du discours politique – redondance supérieure à celle du discours littéraire par exemple ; redondance que l'on préférera contrôler plutôt que ressentir par ailleurs – qui fait de la logométrie, pour nous, non pas un luxe mais une nécessité.

### 3. Une fin : le langage politique contemporain

Si le discours est selon la définition minimale et historique de [Benveniste 1966, t.1 : 242] « toute énonciation supposant un locuteur et un auditeur, et chez le premier l'intention d'influencer l'autre en quelque manière », alors le discours politique est sans doute la forme archétypale du discours : son *adressage* et son *intentionnalité*, au cœur du propos du linguiste, sont explicites et non à supposer.

Si, dans la perspective pragmatique ouverte par Morris puis Austin, le discours met toujours en jeu la dimension actancielle du langage ou sa dimension performative, alors, à nouveau, le discours politique est une forme privilégiée de discours : locution, illocution, perlocution s'y entremêlent à parts égales, comme la dimension esthétique est constitutive du discours poétique.

En termes plus élémentaires, loin de tout modèle théorique, le point de départ de nos études tient dans le sentiment qu'il n'existe aucune innocence et aucune vanité, aucun hasard<sup>57</sup> et aucune gratuité, dans les choix linguistiques des échanges verbaux politiques : une analyse logométrique minutieuse, au sein de corpus bien formés, s'en trouve d'autant plus justifiée.

Le texte (plongé dans un corpus textuel) est pour nous l'objet d'étude car il constitue dans l'évidence de sa matérialité linguistique la forme empirique et observable du discours. Reste que la préoccupation ultime de notre travail semble épouser le programme général de l'Analyse du discours telle qu'elle s'est constituée dans les années 1960 : comprendre, par le discours, l'articulation entre la langue et le monde, le locuteur et la société, la performance linguistique et la circonstance sociale ou politique<sup>58</sup>.

Posons avec Jacques Guilhaumou que le langage informe l'événement politique : l'événement n'advient ou ne se constitue que dans et par sa « donation linguistique » [Guilhaumou 2006 : 90]. Sans nier l'existence d'un factuel historique ou politique (la prise de la Bastille, l'exécution de Louis XVI..., l'élection de Nicolas Sarkozy), il s'agit d'admettre comme essentiel le truchement linguistique dans le fait – ou la *factualisation* – politique, et ceci aussi bien pour le contemporain-acteur qui vit,

---

<sup>57</sup> Hasard. Le mot n'est pas à prendre seulement dans son sens commun. Il doit être relié à la méthode exposée *supra*. L'approche statistique est là pour chiffrer le moment où la distribution de certains mots, leur récurrence ou leur absence, dans telle ou telle partie du discours, ne peuvent plus être imputables au seul hasard.

<sup>58</sup> Notre position vis-à-vis de l'AD, et particulièrement vis-à-vis de l'évolution de l'œuvre de Dominique Maingueneau a été précisée dans *Langage et Société* {21}. Résumons : nous avons pris nos distances vis-à-vis de l'AD qui se pratique au début du XXI<sup>ème</sup> siècle pour deux raisons. D'abord car celle-ci semble avoir « oublié le texte » selon l'expression de [Sarfati 2003 : 432] ; la linguistique semble de moins en moins présente au profit notamment des sciences de la communication. Ensuite, et plus encore peut-être, parce qu'elle semble avoir oublié l'exigence méthodologique pour traiter ce texte ; l'exigence méthodologique – lexicométrique notamment – avait été pourtant posée originellement par Harris ou Pécheux, Robin ou Guilhaumou, les premiers Maingueneau ou les premiers Bonnafous, Habert, Labbé, Saint-Cloud, les 10 premiers numéros de la revue *Mots*, etc.

fait ou observe l'actualité politique que pour l'historien-rédacteur qui sera amené, par son récit, à la recomposer.

En retour, posons avec Maurice Tournier que le politique informe le langage : le politique « prend la langue » [Tournier 1997 : 292], la contraint, se l'approprie. Sans nier l'importance du système linguistique cette fois-ci, il s'agit d'admettre par exemple que les mots sont des « produits sociaux » [*ibid.* : 287], traversés par la politique ou l'idéologie, et qu'ils prennent sens dans et par l'usage social. « Liberté » ou « égalité », « peuple » ou « bourgeoisie », « Etat » ou « nation », « travail » ou « famille », « je » ou « parti », « nous » ou « insécurité », « naturellement » ou « réformer », etc. : ces mots, dans nos corpus, n'ont pas de sens littéral, mais un sens politique ; « l'unité de langue est un leurre [...], la fonction sociale d'un terme prime sa signification. » [Tournier 1992 : 12-13].

A l'intersection de ces deux programmes complémentaires (« l'histoire langagière des concepts » [Guilhaumou, *Discours et Événement. L'histoire langagière des concepts*, 2006] et « l'histoire des usages linguistiques, vus du côté des historiens » [Tournier, *Des mots en politique. Propos d'étymologie sociale* 1997 : 298]), posons donc dans notre travail la consubstantialité de l'activité politique et de l'activité langagière, et admettons que le discours politique est un objet interdisciplinaire qui réclame compétence linguistique (particulièrement nécessaire au niveau de sa description) et compétence historique (indispensable au niveau de l'interprétation)<sup>59</sup>.

Concrètement, dans cette partie, nous voudrions rappeler quelques grands constats linguistiques et quelques interprétations socio-linguistiques abouties que l'on a été amené à faire sur le discours politique français contemporain, et en prolonger l'analyse. Le terrain d'étude a été le discours de la III<sup>ème</sup> République {**I, 5, 6, 7, 9, 10, 13, 17**}, et le discours de la V<sup>ème</sup> République sous sa forme présidentielle ou ministérielle {**II, 14, 16, 19, 26**} et sous une forme électorale {**27, 30, 31, 32, 34, B**}<sup>60</sup>. Pour répondre à l'ambition chronologique présentée lors du concours de recrutement du

---

<sup>59</sup> Pour [Darbellay 2005], la complexité des approches interdisciplinaires tient dans la nécessaire *intrication* des disciplines pour le traitement de l'objet – sans quoi l'on parlerait d'approches pluridisciplinaires. Un objet interdisciplinaire est défini comme tel lorsque plusieurs points de vue *entrecroisés* – et non juxtaposés – sont nécessaires pour en rendre compte. Cependant, pour régler – réguler – la relation jamais ancillaire entre Linguistique et Histoire, entre Histoire et Linguistique, nous posons schématiquement dans nos pratiques que la linguistique prime dans la description du matériau textuel et que l'histoire encadre nécessairement l'interprétation ; sur cette base clarifiée, l'intrication peut commencer à être tricotée. En ce sens, nous partageons ici le positionnement de la *Critical Discourse Analysis* et du programme de travail que [Fairclough 2001] présente lorsqu'il distingue un temps pour la description des propriétés formelles du texte et un temps pour l'interprétation. En ce sens nous nous mettons sous la critique de Rastier dans l'analyse qu'il fait de l'Analyse du discours [Rastier 2001, *Enonciation et Analyse du discours* : 242-246].

<sup>60</sup> Consubstantialité du langage et du politique : l'affirmation nous semble valide en général. Remarquons néanmoins que nous avons toujours choisi, pour plus de précautions, des corpus particuliers dans lesquels la consubstantialité du dire (linguistique) et du faire (politique) était évidente : les années 30 décrites par les historiens comme un moment de guerre civile verbale ou d'affrontement simulé (là où ailleurs en Europe, l'affrontement était martial et non seulement verbal) ; le septennat de Jacques Chirac – le président de la parole – durant lequel le président, privé dans une cohabitation longue de tout pouvoir d'agir, fut condamné à seulement discourir ; les périodes électorales enfin qui apparaissent comme des moments caricaturaux dans nos logocraties, durant lesquels l'action politique se réduit entièrement au discours.

CNRS en 2000 – l'ensemble du XX<sup>ème</sup> siècle était alors évoqué – , indiquons que le discours de la IV<sup>ème</sup> République manque encore : nous ne l'étudions que par procuration dans le co-encadrement de la thèse de Julien Bonneau portant sur le corpus de Pierre Mendès France<sup>61</sup>, avant d'en faire notre propre terrain d'analyse dans les années à venir.

### **3.1. Individuation politique et linguistique ?**

Peut-être s'agit-il là seulement d'un artefact de la méthode ? Mais à analyser des corpus qui contrastent – volontairement – plusieurs locuteurs, nous sommes toujours arrivé à identifier des personnalités linguistiques. Partant, comme nos locuteurs représentent ou appartiennent à un courant de pensée politique identifié ou à préciser, nous avons essayé de mettre en correspondance l'identité linguistique constatée avec l'identité politique supposée ou à découvrir.

Individuation linguistique et politique : derrière la banalité du constat voire son artefactualité, la question posée est d'importance. Existe-t-il un rapport inférentiel entre l'identité discursive d'un locuteur et son positionnement politique ou idéologique ?<sup>62</sup> Plus précisément, des différents facteurs qui contraignent le discours d'un orateur, et que nos corpus en général croisent (facteur politique : chaque locuteur est situé politiquement ; facteur générique : chaque discours est prononcé dans un genre déterminé ; facteur chronologique : chaque discours est historiquement daté et traite d'une actualité donnée), le facteur politique – l'engagement personnel politique du locuteur – est-il essentiel pour comprendre l'identité discursive qui se dégage ?

En termes simples : s'exprime-t-on différemment selon qu'on est de gauche ou de droite, communiste ou réformiste, républicain ou révolutionnaire, orléaniste ou gaulliste ? Et inversement, cesse-t-on d'être réformiste ou amende-t-on le réformisme selon les propos que l'on va tenir lorsqu'on incarne ce courant de pensée ?

En termes althusériens ou foucaaldiens en partie décalés dans le paysage marxiste actuel et dans lesquels le rapport d'inférence est plus clairement orienté<sup>63</sup> : les formations discursives déterminent-elles le discours des hommes ? Telle position de classe ou partisane entraîne-t-elle telle pratique langagière particulière ? Quel reflet donne le discours du positionnement socio-idéologico-politique des locuteurs ?

#### **3.1.1. Thorez versus Blum ou la question des genres de discours**

Sans revenir en détails sur notre thèse {I}, rappelons qu'elle a permis d'établir dans sa principale conclusion l'identité linguistique de quatre dirigeants politiques

---

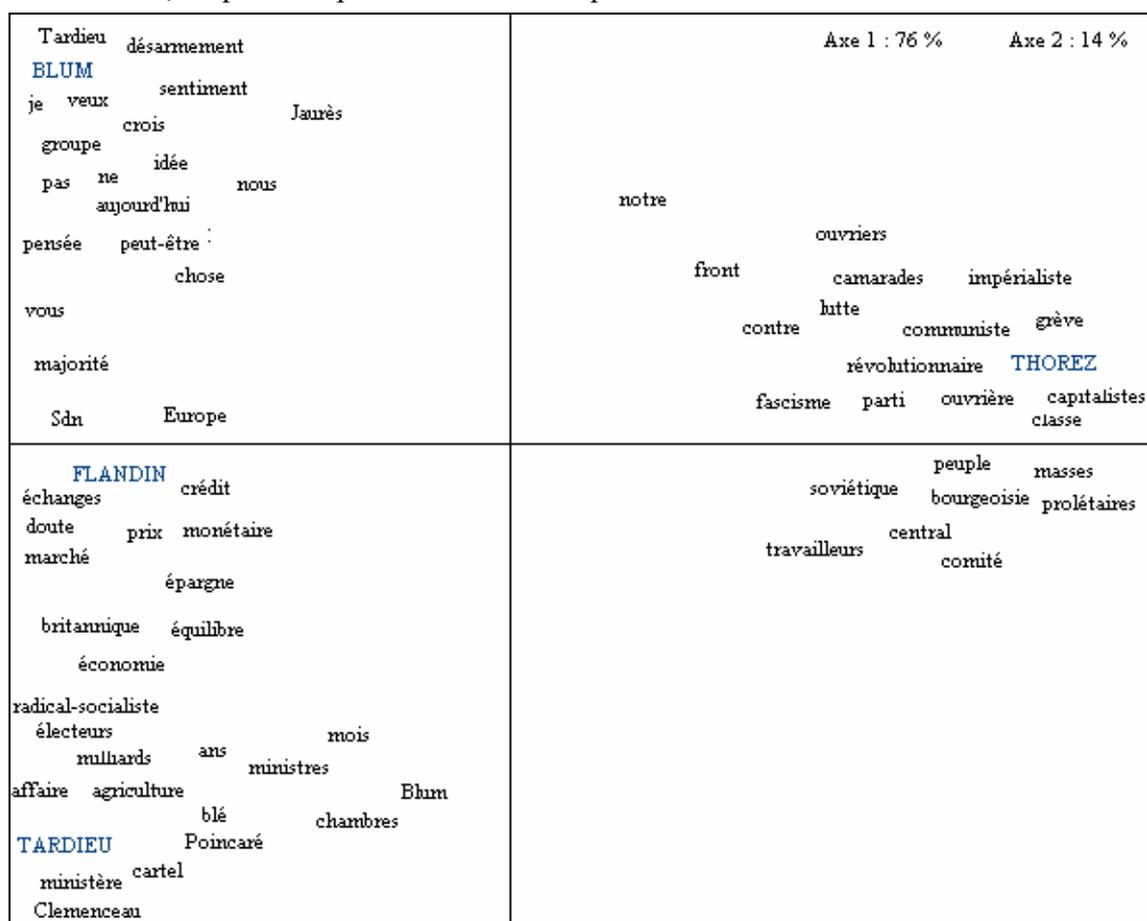
<sup>61</sup> Julien Bonneau, *Approche mathématique de la textualité. Application au corpus Mendès France (1924-1960)*, thèse en cours sous la co-direction de D. Mayaffre et S. Mellet.

<sup>62</sup> Suggérons que s'il devait y avoir un rapport d'inférence, il s'agirait d'une inférence non orientée. En vertu du rapport consubstantiel évoqué entre langage et politique, un positionnement politique peut inférer une pratique discursive, comme l'énoncé d'un discours peut inférer (modifier) un positionnement politique. Ici nous sommes bien dans le cadre d'une corrélation ou d'une co-variance non orientée (cf. note suivante)

<sup>63</sup> Ici nous quittons donc l'idée de corrélation ou de d'inférence non orientée pour entrer dans une forme de détermination. C'est la position politique qui détermine le discours (cf. note précédente).

majeurs de l'entre-deux-guerres<sup>64</sup>. En étudiant les discours de Thorez (PCF), Blum (SFIO), Flandin (droite orléaniste), Tardieu (droite bonapartiste), nous avons conclu à une individuation linguistico-politique forte, relativement stable sur une décennie (1928-1939). Cette identité est particulièrement marquée du côté du locuteur communiste, et, par contraste, l'identité linguistico-politique de Blum apparaît apparentée à celle des dirigeants de droite, loin de la dichotomie gauche/droite attendue.

Nous pouvons illustrer le propos par une AFC de 60 mots majeurs, sélectionnés ici à dessein, et que nous prendrons donc simplement à titre illustratif.



**Illustration 6 : AFC de 60 mots dans le corpus entre-deux-guerres**

Sur le premier axe, Thorez, seul, se marginalise en s'opposant aux trois autres dirigeants. Au réformisme politique de Blum bien connu<sup>65</sup> correspond un réformisme

<sup>64</sup> Pour des raisons méthodologiques l'identité linguistique en question était ramenée à une simple identité lexicale : dans les années 1990, les lemmatiseurs n'étaient pas encore fiables ; au regard de la taille du corpus étudié une lemmatisation manuelle était exclue ; Hyperbase comme Lexico ne traitaient pas encore des corpus étiquetés.

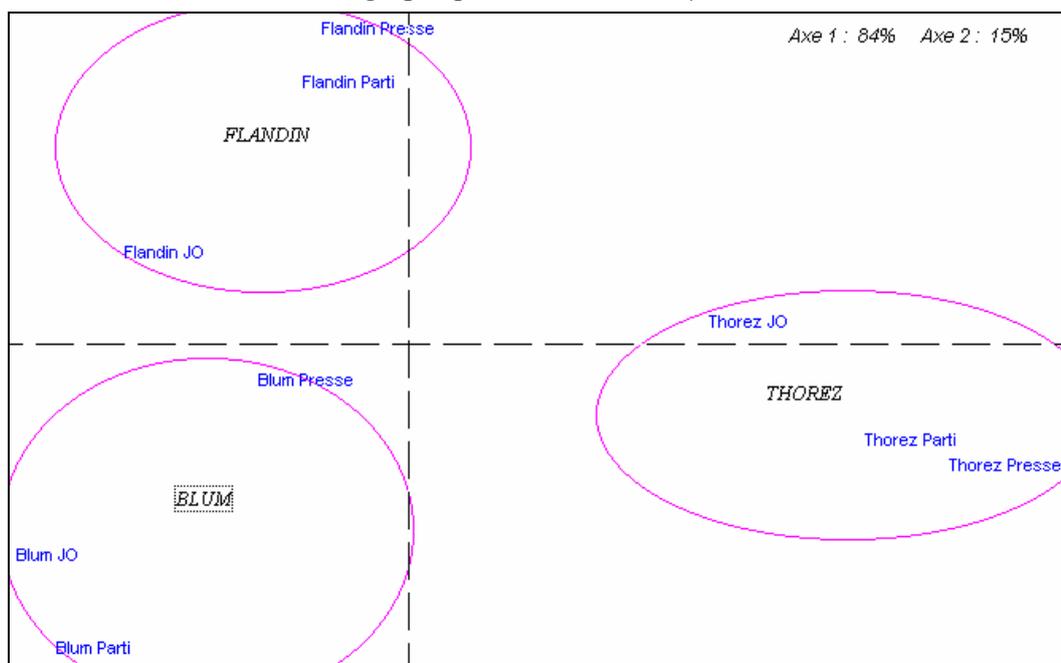
<sup>65</sup> On sait que Blum a théorisé dans les années 20 la distinction entre conquête (révolutionnaire) du pouvoir et exercice (républicain) du pouvoir, inventant ainsi le réformisme ; de fait, ensuite, il fut le premier chef de gouvernement socialiste de la République française en 1936. Cf. son célèbre discours de la Bellevilloise, le 10 janvier 1926 :

« Bien qu'en ce qui concerne la conquête du pouvoir, je ne sois pas légaliste, je le suis en ce qui concerne l'exercice du pouvoir. J'estime que si le déroulement des pratiques parlementaires nous appelle à exercer le pouvoir dans le cadre des institutions actuelles, nous devons le faire légalement, loyalement,

lexical qui rapproche le dirigeant socialiste (sans le confondre certes) des deux dirigeants de la droite républicaine. Le vocabulaire marxiste notamment (« prolétariat », « bourgeoisie », « classe », etc.), voire seulement ouvrieriste (« ouvriers », « ouvrières », « lutte », « grève ») appartient en propre au dirigeant communiste à l'exclusion de celui de la SFIO {pour plus de détails, voir **I** : notamment 390-399}.

Dix ans après notre thèse, cette conclusion nous paraît toujours valable. Cependant, nous avons cherché à la contredire dans un premier article {9} et dans un second paru d'abord en français {17} puis en portugais brésilien dans un ouvrage collectif {29}. Comme ces articles retravaillaient une notion fondatrice de l'école française d'Analyse du discours (les « formations discursives »), le propos a pu être repris par Francine Mazière dans son précis sur l'AD en France de 1969 à nos jours paru aux Presses Universitaires de France [Mazière 2005].

L'objectif était de *mettre à l'épreuve des genres* les identités linguistico-politiques brièvement décrites ci-dessus. En séparant / contrastant les discours d'origine partisane, les discours d'origine journalistique et les discours d'origine parlementaire de chaque locuteur (Thorez/Blum/Flandin)<sup>66</sup>, nous avons pu constater que les identités linguistico-politiques décelées résistaient aux variations génériques {9, et. ci-dessous les deux illustrations sur les mots graphiques et les lemmes}.



**Illustration 7 : Analyse factorielle du vocabulaire (l'ensemble des mots graphiques)**

---

sans commettre cette espèce d'escroquerie qui consisterait à profiter de notre présence au gouvernement pour transformer l'exercice du pouvoir en conquête du pouvoir. »

<sup>66</sup> Tardieu a dû être exclu de l'analyse car il a toujours refusé d'adhérer à un parti : son corpus de compte pas de discours partisans.

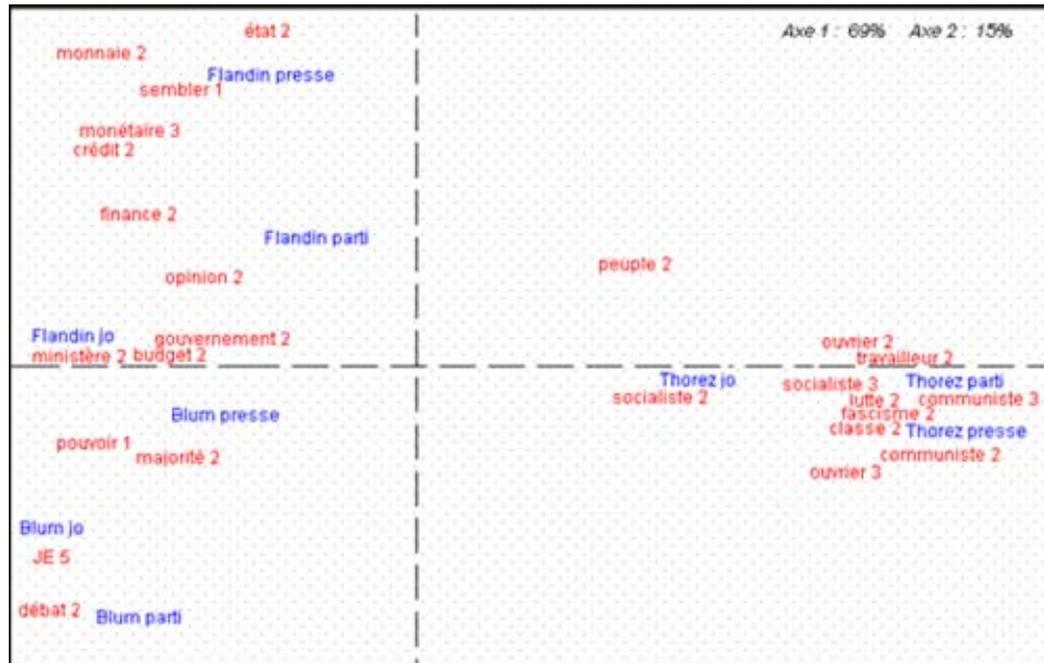
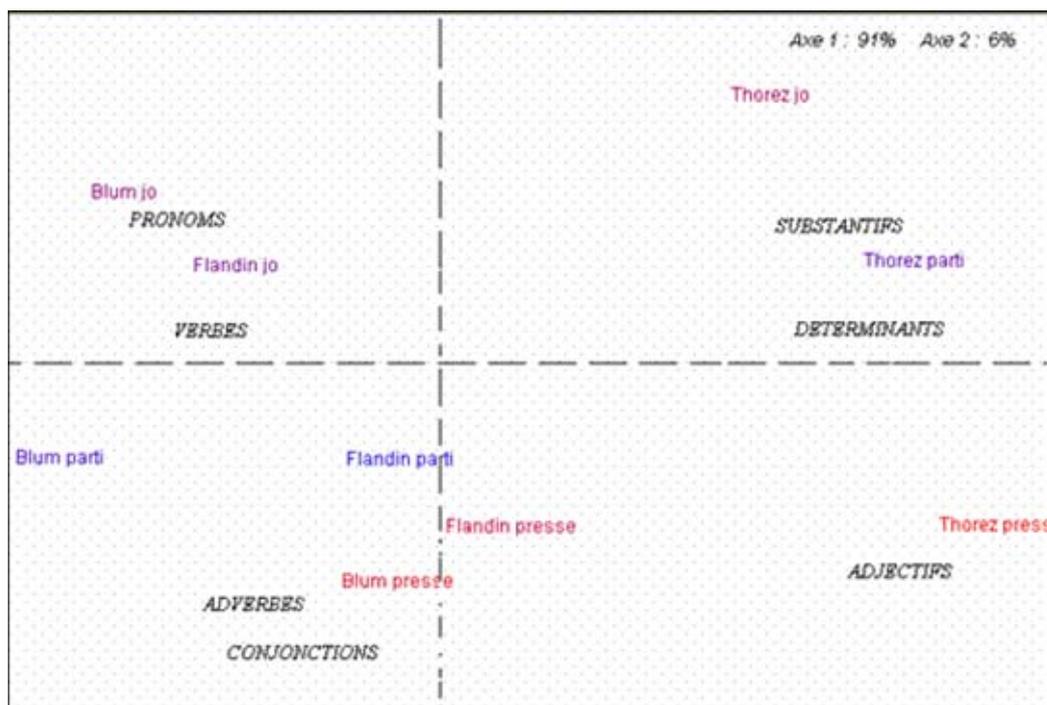


Illustration 8 : Analyse factorielle de 25 lemmes

De manière sans doute trop définitive, nous avons ainsi conclu que rien ne ressemblait plus à un discours communiste qu'un autre discours communiste ; qu'il soit prononcé à l'oral ou rédigé à l'écrit, produit à huis clos devant le Comité central du parti ou bien en public devant le Parlement, publié dans l'*Humanité* sous forme d'éditorial ou au *Journal officiel* sous forme de compte-rendu de débat parlementaire.

La faiblesse de l'analyse tient dans la modicité du corpus traité qui malgré son ampleur en taille ne croise ici que quelques locuteurs durant une seule décennie : en dépit de nombreuses études antérieures qui paraissent convergentes [Courdresses 1971 ; Labbé 1977 et 1983 ; Gerstlé 1979], nos conclusions sur un parler communiste *versus* un parler républicain-bourgeois gagneraient à être vérifiées sur d'autres corpus, avec d'autres locuteurs, durant des périodes différentes. La force de l'analyse, en revanche, réside dans le fait que les identités déterminées sont désormais enregistrées aussi bien d'un point de vue lexical (cf. les deux graphiques précédents) que d'un point de vue grammatical ou morphosyntaxique {cf. graphique dans 9 : 526 ou illustration suivante} : si on comprend aisément que le discours communiste puisse embrasser des thématiques ou un vocabulaire stables et particuliers (le « parti », la « classe ouvrière », la « révolution », etc.), nous restons plus étonné que le discours communiste ait un régime énonciatif ou une tonalité rhétorico-grammaticale qui lui soit propre quel que soit le genre (discours non verbal avec des nominalisations systématiques, discours didactique au sens de [Courdresses 1971], énonciation toujours collective sous la forme d'un « nous » subsumant, masquage systématique du locuteur, usage des noms propres, etc.).

En effet, de manière certes moins marquée que pour le vocabulaire (on pourrait ici insister sur l'axe 2), l'analyse factorielle des parties du discours distribue les 9 sous-corpus ainsi et isole à droite de l'axe 1 le langage communiste :



**Illustration 9 : Analyse factorielle des parties du discours**

Nous renvoyons le lecteur au détail de l'analyse pour en discuter ses principales conclusions. Prenons le risque seulement, dans le cadre de l'HDR, d'explicitier le positionnement épistémologique implicitement défendu alors ; en France, dans le contexte scientifique actuel, ce positionnement apparaît non consensuel.

Le marxisme ou, dans le cadre de l'AD, le marxisme althusserien ont trop longtemps insisté, jusqu'à la fin des années 1980, sur le déterminisme social pesant sur l'activité discursive. Le discours – sinon la langue elle-même – comme toute autre chose se voyait déterminé par quelque infrastructure. *En dernière instance*, selon l'expression d'Althusser, c'était la position de classe du locuteur, sa place dans l'appareil de production, qui expliquaient une performance discursive. *Grosso modo*, dans un jeu de déterminismes en cascade dont il n'est pas le lieu ici de rediscuter le détail, les intérêts de classe produisaient une (fausse) conscience idéologique, qui elle-même se traduisait par un positionnement politique sur l'échiquier, qui lui-même s'exprimait par un discours particulier ; de fait, c'est explicitement que le parti communiste entend servir dans et par son discours les intérêts matériels de la classe ouvrière.

D'un excès à l'autre, dans un mouvement de balancier, le danger existe aujourd'hui pour que nous sous-évaluions ces déterminismes jusqu'ici sur-évalués. On les remplacerait alors de manière *ad hoc* par d'autres contraintes, étant entendu que personne ne conçoit le discours comme une expression affranchie de toute contrainte comme les associations libres prétendaient l'être dans le surréalisme. Le genre semble devenu, au détour des années 1990-2000, ce déterminisme principal pesant sur les discours, ce facteur explicatif premier de l'expression d'un locuteur. Ainsi, avec des positions certes différentes et des définitions très variées du genre, on lira par ordre alphabétique : Achard 1995, Adam 1999, Baroni et Macé (eds.) (2007), Branca-Rosoff 1999, Bronckart 1997, Maingueneau 2002, Rastier 2001, etc. (Nous limitons ici volontairement ces quelques références aux seules années 1990-2000 sans ignorer que

l'Antiquité avait déjà cherché à distinguer les genres de discours, ni négliger les écrits fondateurs de Bakhtine puis de Todorov ; pour un état de l'art sur le genre cf. [Poudat 2010]).

Il est impossible de nier l'importance du genre dans une production discursive. Comme l'on parle le mandarin ou le français – à un autre niveau certes ! –, nous nous exprimons dans des genres différents qui tendent à normer nos discours. Mais dans le domaine politique, le genre ne semble pas le *principe premier*, comme il l'est dans le domaine littéraire, pour expliquer, identifier, caractériser un discours : le facteur individuel c'est-à-dire, pour nous, le facteur politique – issu directement ou non d'une position de classe du locuteur – nous est apparu toujours plus important.

Des propos définitoires de Jacques Bres, significatifs car récents, et s'appuyant précisément sur les genres du discours politique, nous ne retiendrons donc que l'essentiel, à savoir le programme de recherche, sans pouvoir souscrire à l'ensemble :

« La notion de genre apparaît aujourd'hui comme fondamentale en analyse du discours : puisque tout discours se réalise dans (au moins) un genre et que *le genre détermine les thèmes, le style et la structure du discours*, il apparaît heuristique de mettre en relation discours et genre pour étudier leur interaction. ». [J. Bres, in Détrie, Siblot, Verine, 2001 : 131. Le passage critique est souligné par nous].

Lorsque Jacques Chirac décide de communiquer sur l'insécurité en 2001-2002, il le fait aussi bien, longuement, en dialogue avec des journalistes lors de l'interview télévisée du 14 juillet 2001 que plus rapidement dans son monologue solennel du 31 décembre, avant de revenir constamment sur la question dans ses meetings (oraux) électoraux, dans ses tribunes (écrites) de campagne ou dans ses tracts et profession de foi au début de l'année 2002 : *sa thématique ne paraît pas déterminée par le genre du discours*, mais par un engagement politique à droite de l'échiquier, un désir de contraster politiquement avec Jospin, une volonté de battre la gauche plurielle portée, elle, par ses résultats économiques et la thématique sociale.

Quand Léon Blum, dans l'entre-deux-guerres, abandonne l'idéal révolutionnaire et la conscience prolétarienne, il le fait aussi bien dans ses éditoriaux du *Populaire* que dans les travées du Parlement : ici aussi son vocabulaire fondamental, hérité plus de Ferry que de Guesde, ne semble affecté que secondairement par le genre varié de ses discours ; il est plus probablement déterminé politiquement par la base sociologique de l'électorat SFIO, composée, dans les années 30, de la petite bourgeoisie rurale à laquelle l'ancien Maître des requêtes au Conseil d'Etat s'adresse.

Quand François Mitterrand utilise l'égotisme comme principe de gouvernement et la polémique en lieu et place de l'argumentation, son *style éristique* (énonciation tendue fondée sur une omniprésence du « je », sur-présence de l'opposition « je/vous », sur-utilisation de l'impératif, sur-utilisation des verbes performatifs, des modalités énonciatives, etc.) ne se laisse lui non plus pas enfermer par les genres discursifs et reste prégnant aussi bien à l'oral qu'à l'écrit, à la télévision que dans la presse ou par exemple dans sa *Lettre à tous les Français* du printemps 1988. Et nous avons posé qu'il s'agissait là, sans doute, de la rançon politique (et non générique) que réclamait la social-démocratie à ses dirigeants : au final, dans le réformisme, seul le « je » du leader, dans d'innombrables polémiques, peut arbitrer l'impossible compromis entre le programme révolutionnaire du parti et la pratique légaliste gouvernementale. En substance, Mitterrand répète durant 14 ans aux adhérents du parti socialiste, aux électeurs de

gauche et finalement au peuple entier : « JE sais que l'on ne fait pas le programme marxiste du parti établi à Epinay, mais JE VOUS promets – et JE suis le seul à avoir une vue d'ensemble de la difficulté de gouverner – que l'on ne peut pas faire autrement dans la situation actuelle. Et JE VOUS demande, MOI qui ne VOUS ai jamais trahis, de ME faire confiance dans l'établissement de ce qui est possible présentement »<sup>67</sup>.

Aujourd'hui enfin, Nicolas Sarkozy entend incarner le volontarisme politique et agir au-dessus des partis, en dialogue direct avec ces Français qu'il oppose aux élites, aux corporations, à l'intellectualisme ou à la pensée unique. Ici encore, les genres de son discours, que l'on qualifiera de populiste, n'affectent qu'indirectement le fond et la forme de son propos : nous retrouvons aussi bien dans ses discours de tribune, dans ses interviews à bâton rompu que dans ses allocutions solennelles la sur-modalisation en vouloir (« je veux... ») ou falloir (« il faut... »), la récurrence d'un « on » indéfini et menaçant (« on veut vous faire croire que... mais »), les phrases courtes, directes, hyposyntaxiques, un vocabulaire simple et antinomique, des anaphores rhétoriques répétitives {27 et B ; cf. *infra* 3.1.3. *Sarkozy versus Royal*}. La complexité de la pensée y perd, mais l'efficacité d'une communication dirigée vers le plus grand nombre – notamment l'électorat qui s'était réfugié dans l'abstention ou votait Front national – s'en trouve renforcée<sup>68</sup>.

Si le genre est un espace normatif et qu'il régle le discours selon des canons d'autant mieux identifiés qu'ils sont transgressés – selon l'aveu même de Todorov –, il ne constitue donc pas pour nous un « déterminisme causal » ni des thèmes du discours (repérables, en logométrie, par les récurrences / absences lexicales) ni même de sa rhétorique fondamentale (repérable, en logométrie, *via* la récurrence / absence de traits grammaticaux, morphosyntaxiques ou syntaxiques). Sans jamais juger le genre comme inessentiel, nous le tenons, à rebours des tendances actuelles, comme moins contraignant sur le discours d'un dirigeant politique que ces déterminismes sociaux et idéologiques que l'école française d'Analyse du discours a mis, un temps, en avant<sup>69</sup>. Qu'un locuteur utilise par exemple plus d'adverbes à l'oral qu'à l'écrit, plus la première personne dans un entretien télévisé que dans une tribune est pour nous une chose

---

<sup>67</sup> Cette phrase est donc inventée sur la base de mots statistiquement mitterrandiens. Mais l'on en trouvera d'authentiques très voisines reproduites dans {II : 82-83} notamment lorsqu'il s'agit pour Mitterrand de justifier le tournant droitier de 1983-1984 et l'abandon du Programme commun. Surtout, nous avons montré que ce type de rhétorique était trait pour trait celui de Blum au nomment de justifier l'abandon du programme du Front populaire et la non-intervention en Espagne en 1936. Comme si les deux figures historiques du réformisme français avaient été condamnées à utiliser les mêmes procédés linguistiques pour justifier leur ligne politique {pour une comparaison précise de ce *discours polémique*, cf. II : 77-87 et I : 211-219}.

<sup>68</sup> Après la *thématique* et le *style*, reste, dans la définition de J. Bres, la *structure du discours* qui, elle, semble directement soumise aux contraintes génériques.

<sup>69</sup> Aujourd'hui, notre positionnement apparaît presque comme marginal en France. Cependant, le paradigme conceptuel de l'école française d'Analyse du discours reste très présent dans le continent sud-américain aussi bien de langue espagnole que portugaise, et très présent dans l'œuvre de Fairclough et de la *Critical Discourse Analysis*. Il est amusant de voir que certains concepts français sont aujourd'hui « réintroduits » en France par l'école latino-américaine ou anglo-saxonne. Cf. par exemple, pour la seule année 2009, *Semen* n°27 (avril 2009) qui propose la traduction française de quelques articles fondateurs de la CDA – elle-même ayant traduit en anglais Pêcheux, Foucault ou Althusser –, et la tonalité des débats du colloque international *Les discours politiques en Amérique latine* (12-14 février 2009, Université Paris-Est ; actes à paraître pour partie dans *Mots* en 2010).

évidente, mais pas plus évidente que d'autres caractères lexicaux, énonciatifs, rhétoriques en lien avec son identité politique et son positionnement social.

### 3.1.2. Chirac, entre de Gaulle et Mitterrand

L'individuation linguistico-politique est aussi au cœur de notre ouvrage sur le discours présidentiel sous la V<sup>ème</sup> République {II}.

De manière attendue, aux cinq présidents étudiés de la V<sup>ème</sup> République ont correspondu cinq identités discursives différentes – nous ne comptons pas encore les discours de Sarkozy.

De manière rassurante, au sein du corpus présidentiel, les deux septennats Mitterrand – nous ne comptons pas alors les deux mandats de Chirac – se sont toujours confondus dans nos études. La chronologie, nous le verrons, affecte le discours mais pas jusqu'à dissoudre l'identité discursive de celui qui le porte ; si l'identité discursive mitterrandienne, par exemple, évolue entre 1981 et 1995, elle n'est abolie ni par les périodes de cohabitation (1986-1988 ; 1993-1995) ni par la trajectoire politique enregistrée en 14 ans (cf. *infra* 3.2. *Temps discursif et temps politique*).

De manière cohérente, encore, le discours de Pompidou, que l'on qualifiera dès lors de discours gaullien, s'est toujours rapproché des discours de de Gaulle. Lexique commun, parties du discours partagées dans les mêmes proportions avec par exemple un goût prononcé pour les adjectifs, temps verbaux proches, etc. : la filiation que les sciences politiques ont soulignée entre les deux hommes apparaît évidente d'un point de vue discursif.

C'est donc à la suite de ces conclusions qu'il a fallu constater que le discours de Chirac s'individualise moins facilement que les autres {II, section II – *Chirac : le président de tous les Français* : 95-188}. Ce fut la difficulté de notre travail et son intérêt : l'identité linguistico-politique de Jacques Chirac a des traits pastels.

D'un point de vue lexical par exemple, la distance intertextuelle, calculée sur l'ensemble du vocabulaire, donne le graphique suivant sur lequel Chirac apparaît au centre, directement rattaché à la racine de l'arbre :

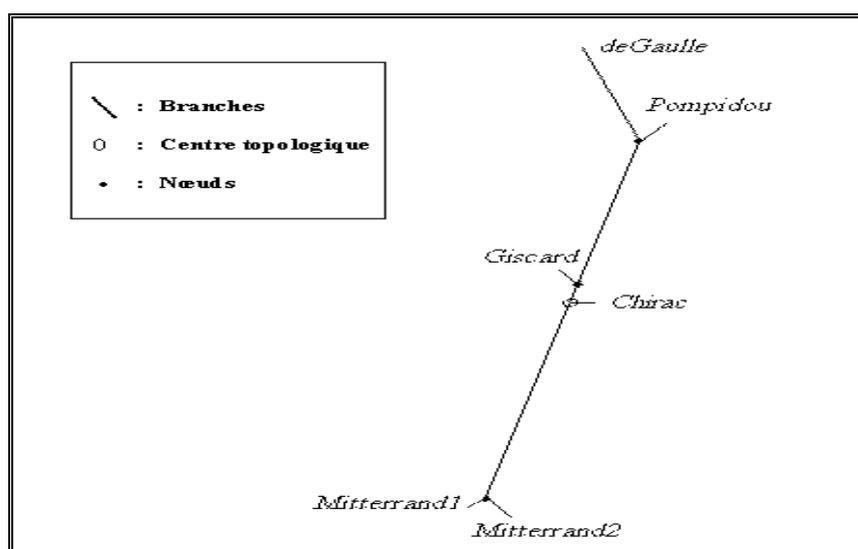
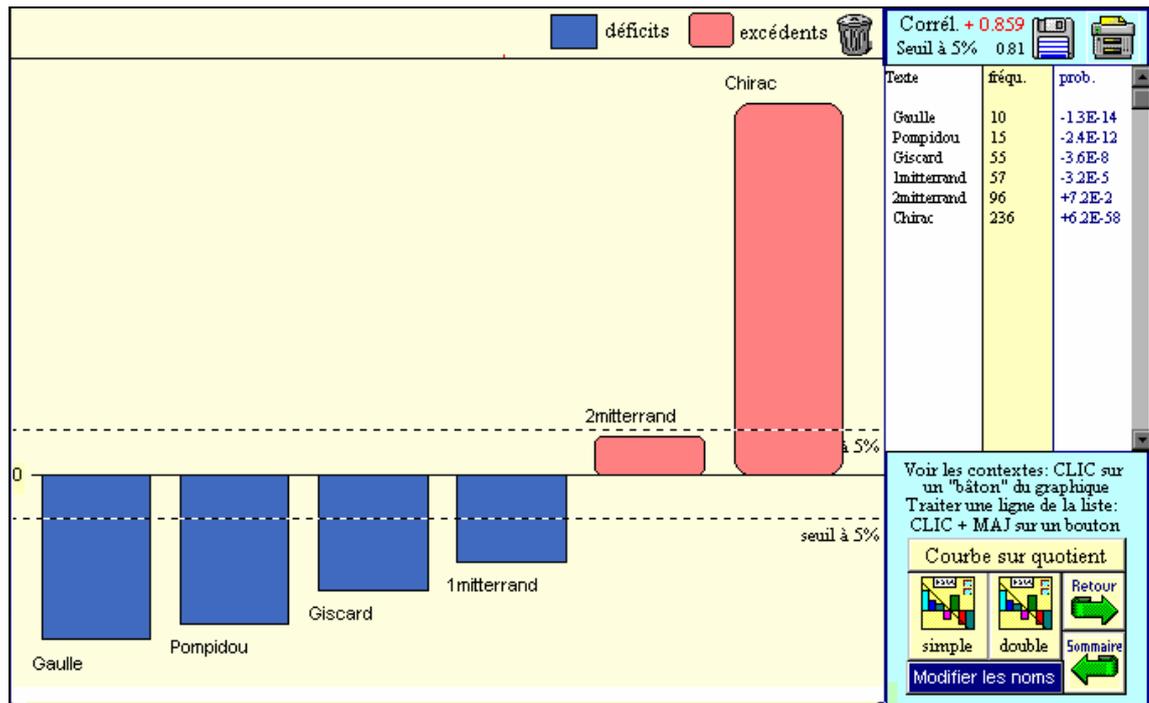


Illustration 10 : Analyse arborée de la distance intertextuelle calculée sur le vocabulaire

Ni classique ni moderne, ni de gauche ni de droite, Chirac tient un discours du juste milieu entre celui de de Gaulle et celui de Mitterrand, un discours patchwork ou un discours de synthèse qui emprunte quantitativement à celui de ses prédécesseurs des traits lexicaux, grammaticaux et rhétoriques.

Certes, quelques récurrences lexicales typiques dans le discours chiraquien ont pu être relevées. « Démocratie » par exemple – un contre-exemple donc – connaît une distribution chronologique remarquable dans le corpus, là où nous aurions pu penser qu’il s’agissait d’un mot incontournable de la langue politique républicaine, partagé par tous, et à la distribution étale.



**Illustration 11 : Distribution de « démocratie » dans le corpus présidentiel (1958-2002)**

Au-delà de la progression générale et continue du mot durant 50 ans, sur laquelle on ne peut revenir ici, constatons donc, à rebours de notre démonstration, que Chirac atteint un seuil de sur-utilisation extraordinaire. Dans un discours remarquable surtout par sa banalité quantitative, et plus généralement par sa vacuité politique, cette saillance d’un mot fort a mérité une explication ; depuis lors, la lecture [Agamben, Badiou *et al.* 2009] nous a confirmé l’intérêt qu’il convient de porter à ce *mot caoutchouc* [...] *qui condamne au malentendu ceux qui l’emploient* (p. 9). Grâce à un travail sur les co-occurrences et un retour systématique au contexte d’utilisation, deux usages ou sémantisations distincts – voire opposés – ont alors été relevés. Un usage particulier ou particularisant : la « démocratie » est souvent chez Chirac au détour des années 1990 une « démocratie de proximité » ; il s’agit là d’un concept nouveau que la science politique doit étudier {II : 149-155}. Et un usage universalisant et rhétorique : la « démocratie » est chez Chirac une valeur abstraite et universelle, incantatoire, susceptible de justifier par exemple le bombardement de Sarajevo en 1999 comme la « liberté » (*free, freedom*) permettra de justifier dans le discours de Georges Bush le bombardement de Bagdad en 2002-2003 {II : 147-149}.

La banalité ou « centralité » du discours de Chirac, dont nous venons de souligner ici un contre-exemple lexical, est plus remarquable encore peut-être d'un point de vue grammatical {cf. graphique dans II : 101, et immédiatement illustration suivante élargie à la période la plus contemporaine<sup>70</sup>}.

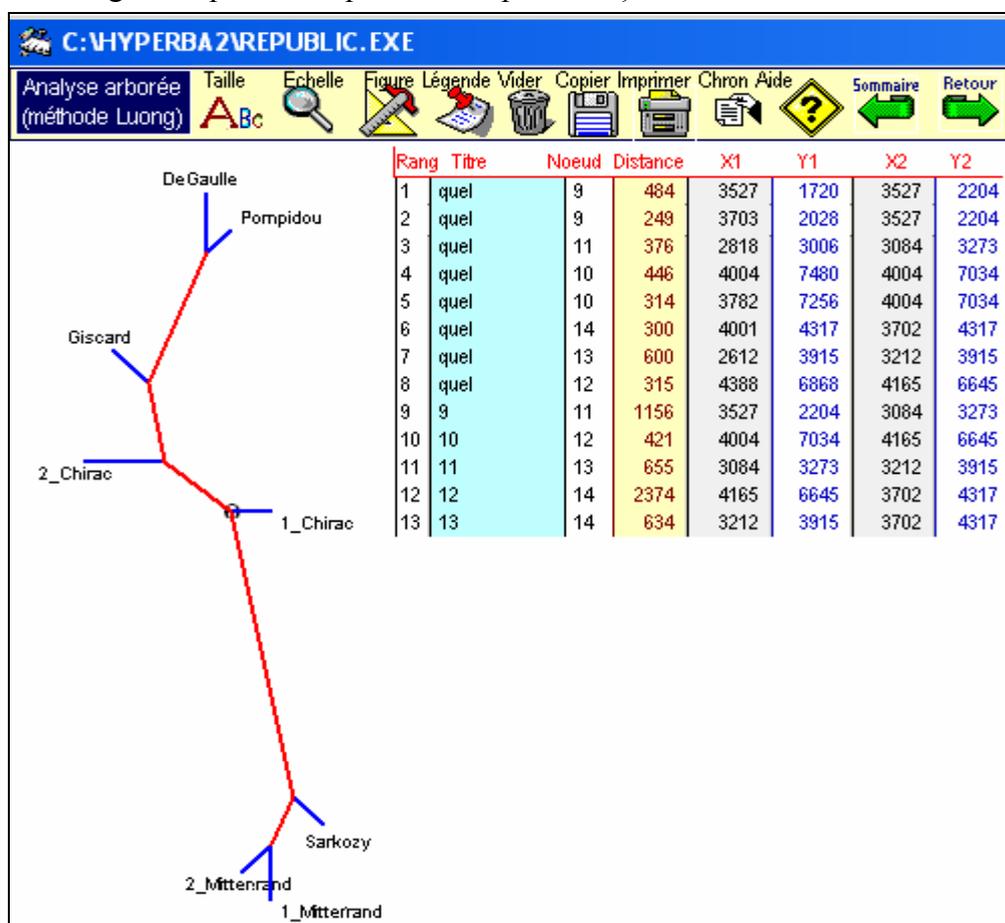


Illustration 12 : Analyse arborée de la distance intertextuelle calculée sur les codes grammaticaux

Dans la grande discrimination chronologique relevée entre un discours nominal au début de la période (1958-1980) puis un discours verbal (1980-2008) (cf. *infra* 3.2.3. *De de Gaulle à Chirac : dichotomie discursive*), Chirac se tient presque à un point d'équilibre. Si Chirac a certes tendance à utiliser plus les pronoms et les verbes que la moyenne, il le fait moins que ne le faisait Mitterrand et que ne le fera Sarkozy. Nous le verrons, à tout prendre, la caractéristique principale de Chirac est la sur-utilisation des adverbes que nous avons exemplifiée avec l'usage remarquable de « naturellement » (cf. II : 106-112, 160-170, et *infra* 3.3.1. *L'adverbialisation du discours contemporain*). Au-delà, la tonalité rhétorico-grammaticale se fait *mezza voce* comme si Chirac par l'équilibre de son discours ne voulait désobliger aucune oreille.

<sup>70</sup> Par rapport à {II}, l'ajout dans le corpus du deuxième mandat de Chirac et des premières années de Sarkozy ne modifie que très peu les choses ; preuve d'une robustesse du traitement et de la permanence de l'identité discursive de chacun dans une trame chronologique assez longue. Dans le graphique suivant, on remarquera néanmoins que le deuxième Chirac quitte légèrement le point médian parfait (centre topologique de l'arbre) pour essayer de renouer avec un parler gaullien.

Au final, ces rappels sur la banalité du discours chiraquien visent seulement à souligner un élément trop souvent oublié dans le traitement quantitatif – ou dans l’idée que l’on se fait des traitements quantitatifs. La logométrie met en valeur, le plus souvent, les récurrences importantes c’est-à-dire les éléments saillants d’un discours. Parfois, de manière symétrique, ce sont les absences remarquables qui sont exhibées (les spécificités négatives par exemple) afin de proposer une caractérisation en creux du discours. Pourtant parfois, c’est la banalité quantitative qui fait sens et offre à l’analyste des chevilles interprétatives ; comme essayera de le montrer Julien Bonneau dans sa thèse, à la suite de certains travaux saint-clousiens, le vocabulaire dit *banal* est un axe de recherche important qu’il conviendra de retravailler<sup>71</sup>. Pour plaire au plus grand nombre, Jacques Chirac semble s’être appliqué à utiliser dans une juste mesure un vocabulaire commun et des traits rhétoriques partagés. Là où d’autres – et nous pensons particulièrement au discours sarkozien (cf. *infra* immédiatement) – tiennent un discours typé, qui fait clivage ou dissensus, Chirac a choisi le discours moyen du consensus.

### 3.1.3. Sarkozy versus Royal

On pointera une difficulté importante dans l’étude précédente. Si le corpus présidentiel fait contraster des locuteurs, il fait contraster du même chef des périodes historiques différentes. A strictement parler, il est impossible d’imputer à l’identité politique des hommes ce qui relève peut-être de la chronologie d’une époque. Deux variables (le temps et les présidents) se superposent strictement sans que l’on puisse isoler l’une de l’autre.

Notre étude doctorale, elle, présentait l’avantage de comparer des locuteurs dans une même période donnée (1928-1939). Néanmoins, un autre paramètre parasitait l’analyse : la période étudiée est si lourde d’événements (crise de 29, arrivée d’Hitler au pouvoir, 6 février 1934, Front populaire, guerre d’Espagne, Munich, début de la 2<sup>ème</sup> guerre mondiale) qu’il devenait difficile de la traiter en synchronie. Si l’on veut bien mesurer, par exemple, l’évolution du discours communiste sous le coup des événements, on conclura à l’existence non pas d’un discours mais de deux discours successifs, le premier d’essence bolchevick, le second d’essence jacobine {**I**: *Les mutations du discours de Thorez* : 407-446 ; **10** ; voir aussi *infra* 3.2.1. *Thorez (1930-1939) : continuum lexical et hiatus chronologique*}. De la même manière, il n’existe pas un discours de Tardieu mais trois successivement entre 1928 et 1939, qui épousent la trajectoire politique de l’homme : un discours orléaniste (1928-1932), un discours bonapartiste (1932-1934) et un discours légitimiste (1934-1939){**I**: *Tardieu ou la tentation autoritaire* : 607-740}.

Bref, les études logométriques exigent des corpus contrastifs dont le tout constitue la norme par rapport à laquelle les parties comparées se caractérisent. Seulement, il est difficile d’isoler les variables principales de contraste des variables secondaires qui affectent le traitement. Ainsi veut-on contraster Chirac et Jospin pour comprendre ce qui différencie encore le discours de gauche et le discours de droite à la fin du XX<sup>ème</sup> siècle ? Las, nous contrastons (aussi, du même coup) génériquement le discours d’un président de la République d’un côté et le discours d’un Premier ministre

---

<sup>71</sup> Après Charles Muller, Pierre Lafon par exemple notait dès le premier numéro de *Mots*, l’importance des *formes de bases* dans le discours politique [P. Lafon, *Mots*, 1, 1980 « Sur la variabilité de la fréquence des formes dans un corpus, pp. 127-165].

de l'autre {14}. Ainsi veut-on comparer de Gaulle à Mitterrand pour comprendre deux personnalités historiques majeures ? Las, nous comparons (aussi, du même coup) les années 1960 aux années 1980, et l'ORTF qui diffuse le discours gaullien à la télévision berlusconienne<sup>72</sup> qui médiatise pour partie le discours miterrandien.

Quoique restée superficielle car réalisée à chaud, l'étude du discours de la campagne électorale 2007 {27, puis 30, 31, 32, B} a le mérite de neutraliser au mieux les différentes variables qui informent un discours, pour espérer isoler, seule, l'identité discursive des locuteurs. La chronologie en effet est stricte et identique pour tous les orateurs étudiés : 5 mois entre janvier et mai 2007. Le genre est stable et bien identifié : il s'agit toujours de discours publics tenus lors de meetings électoraux. Le statut des locuteurs quoique différent dans le détail est identique pour l'essentiel : ils s'agit officiellement de candidats à la magistrature suprême.

Au-delà de l'image générale que l'on peut dresser de l'échiquier politique français de Buffet à Le Pen (cf. ci-dessous, par exemple, l'illustration 13 sur l'identité lexicale de chacun à partir de 30 mots significatifs), cette étude a permis de décrire pour la première fois le discours sarkozien. Les premiers éléments que nous avons pu alors mettre à jour ont été confirmés depuis aussi bien dans l'ouvrage de [Calvet et Véronis 2008] que par les études récentes de [Marchand <http://pascal-marchand.fr/spip.php?rubrique4>], le numéro spécial de [Mots 2009 {B}] ou l'ouvrage collectif de [Perrineau (éd.) 2008]. Surtout, comme mentionné plus haut, nos conclusions sur le discours du candidat-Sarkozy semblent pouvoir être reprises pour l'essentiel pour les discours du président-Sarkozy que l'on se propose d'étudier dans les prochaines années.

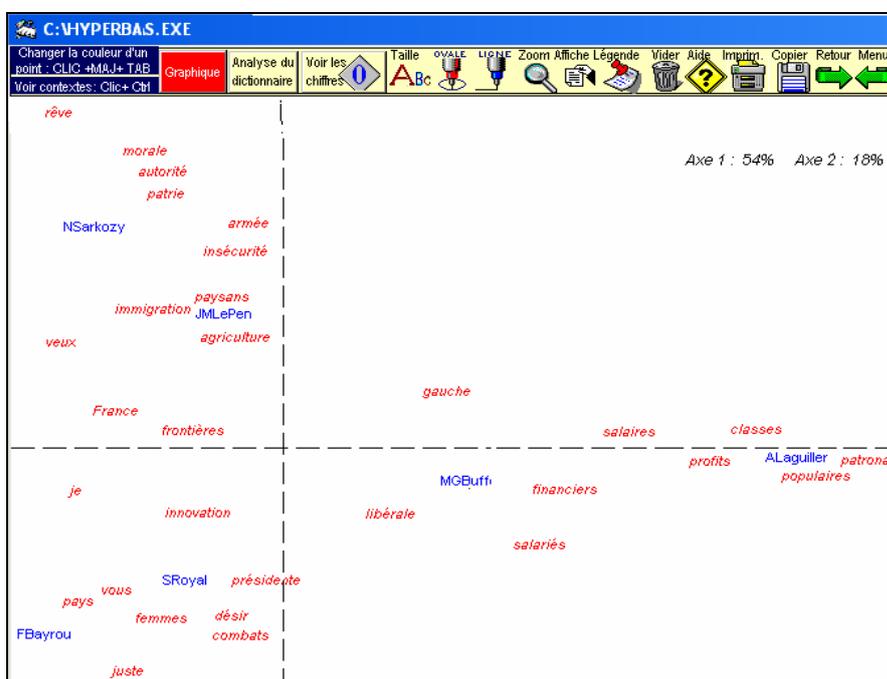


Illustration 13 : AFC de 30 mots durant la campagne 2007 (graphe non commenté ici)

<sup>72</sup> Nous avons noté ailleurs {II : 254} que la libéralisation ou privatisation des chaînes de télévision a été marquée en France par le rachat inaugural, en 1986, de la 5<sup>ème</sup> chaîne par Silvio Berlusconi ; suivra, en 1987, le rachat de TF1 par le milliardaire Bouygues.

A l'inverse de celle de Chirac, l'identité discursive de Sarkozy est forte et volontairement marquée. Sarkozy signe ses discours quantitativement et qualitativement, quand une forme d'anonymat rhétorique avait fini par s'imposer au plus haut sommet de l'Etat. Trois traits linguistiques méritent d'être rappelés, car ils accèdent à l'idée de ruptures discursives – sinon de ruptures politiques –, et érigent ainsi le parler sarkozien en objet d'étude privilégié pour l'avenir.

(i) A l'analyse, le discours de Sarkozy est d'abord celui de l'extrême lexical ou de l'hyperbole lexicale {27} là où le discours républicain traditionnel est habituellement tempéré. Explicitement, Sarkozy s'en prend aux *euphémismes* qui seraient le propre de la pensée unique et finalement responsables de la mort de la politique en tant qu'action déterminée sur le monde. Plusieurs fois en meeting, Sarkozy répète ce paragraphe, dans lequel, notons-le au passage, une habile confusion est entretenue entre le dire et le faire :

Un jour j'ai utilisé le mot « racaille » en réponse à l'interpellation d'une habitante d'Argenteuil [...]. On me l'a reproché. C'est mépriser la jeunesse que de lui parler par euphémismes sous prétexte qu'elle ne serait pas capable de regarder la réalité en face. Quels éducateurs serons-nous si nous nous laissons aller à ces petites lâchetés ? Si les multirécidivistes n'ont rien à craindre ? Si les mineurs peuvent se livrer aux pires excès sans être punis ? Si nous apprenons à nos enfants que l'âge excuse tout ? Si les voyous ne peuvent même pas être appelés des voyous ? (30 mars 2007, Nice ; 19 avril, Marseille, etc.)

Dès lors, la logométrie ne manque pas de repérer dans les discours la récurrence de mots forts que Sarkozy s'applique à répéter et qui discriminent son discours par rapport aux autres candidats (ou aux autres présidents) plus timides lexicalement : « haine », « détestation », « rêve » (verbe et nom), « capitalisme », « polygamie », « voyous », « assistanat », « excision », « morale » (nom et adjectif), « viol », « barbarie »<sup>73</sup>. Lexique de fer donc contre la langue de bois ou de coton supposée de ses adversaires : ce fut la première conclusion de notre étude.

(ii) Le deuxième trait frappant du discours de Sarkozy a déjà été noté : il cherche le *dissensus* lorsque le discours habituel recherche le *consensus* ; il marque les clivages lorsqu'un Bayrou durant la campagne ou un Giscard durant sa présidence par exemple aspirent à l'unanimité ; il incrimine lorsqu'une Royal ou un Pompidou concilient. En ce sens, il s'approche plus du discours de l'extrême droite que celui de la droite traditionnelle. Ce dissensus recherché peut être repéré par le jeu de vocables antinomiques souvent axiologiques (« bon » / « mauvais » par exemple) ou par un vocabulaire ou des thématiques traditionnellement polémiques dans le débat politique français (la « sécurité » ou « l'insécurité », « l'immigration » ou « l'identité », l'« autorité », la « morale », la « religion »). Mais il est surtout important de noter que ces clivages souvent manichéens s'appliquent à traverser – à abolir – les classes sociales pour se situer dans le seul *domaine moral* : il y a là un *distinguo* important avec le discours d'extrême gauche (notamment de l'entre-deux-guerres), lui aussi dissensuel mais d'une toute autre manière. Chez Sarkozy, en effet, la ligne de fracture ne passe pas entre l'ouvrier et le patron, mais au sein même des classes populaires entre « celui qui se lève tôt » et l'« assisté » ; elle ne passe pas entre le patron et les salariés, mais au

---

<sup>73</sup> Nous ne reproduisons pas ici les indices statistiques. Le lecteur aura compris que l'analyse repose sur les mots spécifiques de Sarkozy par rapport à notre corpus de références (cf. pour le détail {27}).

sein même du patronat entre le bon patron et le « patron voyou » ; elle ne passe pas entre les Français et les immigrés, mais au sein même des immigrés entre l'immigré intégré et « l'immigré clandestin », « l'immigré sans papier » ou celui qui « égorge son mouton dans son appartement »<sup>74</sup>. Cette série de mises à l'index, transclassiques et bornées au domaine moral, donne au discours de Sarkozy un aspect violent mais reste électoralement efficace : on remarquera en effet que l'incrimination pointe toujours une minorité voire une marginalité, voire une déviance que la quasi totalité du corps électoral réproouve : le voyou, le polygame, l'exciseur, l'assisté sont des boucs émissaires idéaux. Enfin, si le discours de Sarkozy est un discours dissensuel, alors la négation en est sa clef de voûte. La logométrie classe en effet le « ne... pas » (et le « non ») comme le premier facteur quantitatif discriminant du parler sarkozien et en donne ainsi sa grille de lecture rhétorique (« il ne faut pas... », « vous ne pouvez pas... », « je ne veux pas... », « il n'est pas vrai que... »). Le discours et la pensée de Sarkozy se construisent en s'opposant ; contredisent avant de dire ; nient avant d'affirmer ; interdisent avant d'autoriser ; se nourrissent du conflictuel et des antagonismes dans un premier temps, pour mieux paraître rassembleurs dans un second.

(iii) Faute d'épuiser entièrement un discours complexe, un troisième et dernier trait majeur mérite d'être relevé dans notre étude naissante: Sarkozy s'applique à tenir un discours *oralisé*. C'est au fond sa première signature qui contraste aussi bien avec Royal lors du second tour de l'élection présidentielle qu'avec ses prédécesseurs à l'Elysée lors de leurs allocutions<sup>75</sup>. Même en campagne électorale où l'on sait que le texte est préparé ligne à ligne, et parfois récité à la virgule près d'un meeting à l'autre, Sarkozy feint l'improvisation, la spontanéité, le dialogue direct. C'est peut-être à cette aune simple que l'on peut mesurer le mieux l'importance de la plupart des phénomènes discursifs et décrypter le plus simplement le parler sarkozien. L'hyperbole lexicale mentionnée, par exemple, a pour fonction de donner au discours un aspect entier et spontané, loin d'une langue administrative ou technocratique contrôlée, qui qualifierait peut-être la « racaille » de « contrevenant à l'ordre public ». L'omniprésence du « je », si caractéristique, montre un orateur *in praesentia*, debout face à son auditoire, et qui assume – incarne – directement son propos, sans se réfugier derrière un texte impersonnel ou un collectif politique sans forme. La récurrence des interrogations – le

---

<sup>74</sup> Lors de sa principale émission électorale à la télévision, Sarkozy a ainsi pu déclarer le 5 février 2007 sur TF1 à 21 heures : « Je suis le premier homme politique de droite à dire qu'il faut une immigration choisie, mais je vous dis aussi une chose avec la plus grande force, personne n'est obligé d'habiter en France et quand on habite en France, on aime la France et on la respecte... [Interruption] Non ce qui rend raciste monsieur, c'est cette espèce... [Interruption] Non, si Le Pen dit le soleil est jaune, je ne vais pas être obligé d'arriver en prétendant qu'il est bleu. Personne n'est obligé, je répète d'habiter en France, mais quand on habite en France, on respecte ses règles, c'est-à-dire qu'on n'est pas polygame... [Interruption]. J'y viens, on n'est pas polygame... [Interruption]. Je vais y venir. On ne pratique pas l'excision sur ses filles, on n'égorge pas le mouton dans son appartement et on respecte les règles républicaines. »

<sup>75</sup> Notons ici au passage que la question de l'oralité des discours est une limite importante de notre travail que l'on aimerait essayer de lever dans les années prochaines. Jusqu'à maintenant nous travaillons uniquement sur des textes issus, pour beaucoup, d'une retranscription officielle, disponible à ce titre sur le site de l'Elysée, de Matignon ou les sites officiels de campagne (ou encore dans le JO pour le discours ministériel ou parlementaire). Nous écoutons en général ces discours à la télévision, à la radio ou en vidéo mais sans affronter la question de la fidélité de la transcription ; sans étudier jamais, par exemple, la prosodie des discours ; ou encore sans analyser la gestuelle de l'orateur en meeting ou à la télévision.

point d'interrogation est hautement spécifique du discours de Sarkozy – a encore pour vocation de simuler l'interaction orale avec l'auditoire et un dialogue supposé : à moindre frais, par l'auto-interrogation, Sarkozy simule le débat participatif. Les redondances et les répétitions, dont la forme la plus aboutie sont les anaphores rhétoriques lourdes à force d'être nombreuses dans le discours<sup>76</sup>, font ici aussi penser à un discours qui se construit sur le champ, dans une forme d'improvisation bégayante. Le « on » encore – deuxième caractéristique quantitative après le « ne... pas » – qui méritera un développement particulier dans la lignée des travaux d'Irène Tamba, pourrait être ramené, encore, à cette dimension, si l'on veut bien le considérer comme une forme relâchée et orale du « nous » académique, etc.

Hyperbole lexicale, conflictualité ou dissensus, oralité et forme relâchée : bien sûr, il ne s'agit là que d'un résumé des traits quantitatifs d'un discours qui reste, à ce stade, plus un sujet de recherche pour les années à venir qu'un sujet épuisé. Mais le discours de Sarkozy nous paraît d'ores et déjà un objet d'étude privilégié en matière d'identité discursive tant celle-ci paraît marquée et en rupture avec l'identité discursive policée des présidents successifs ou des derniers présidentiables. Et ici concluons que cette identité discursive épouse et construit une identité politique tout aussi marquée. En termes rhétoriques, dans la lignée des travaux de [Amossy 1999 et 2006], de [Maingueneau 2004] ou récemment du [n°3, 2009, *Argumentation & Analyse du discours* : « Ethos discursif et image d'auteur »], on notera que l'efficacité du propos sarkozien tient dans l'adéquation remarquable entre l'éthos discursif que construit le discours et l'éthos pré-discursif (disons, l'image politique préalable) que Sarkozy s'est forgé depuis plusieurs années<sup>77</sup> : durant la campagne, les « je veux » du candidat par exemple, qui avec le « ne... pas », et le « on » restent la signature linguistique principale, renvoient et alimentent l'image du ministre de l'Intérieur qui a pris en charge fermement le problème de l'insécurité entre 2002 et 2007. Les « ne...pas » renvoient au père sévère ou au premier flic de France sachant poser des interdictions dans une société qualifiée depuis mai 68 de laxiste et déliquescence. Plus généralement, ce phrasé direct, rapide<sup>78</sup> et cette langue relâchée, ce lexique extrême et cette pensée décomplexée, ces thématiques aux confins de l'extrême droite comme l'immigration ou l'identité nationale, ces valeurs assumées comme « l'ordre », « l'autorité », la « morale »<sup>79</sup> renvoient au corpus idéologique de cette nouvelle droite dynamique, populaire ou populiste, bonapartiste ou autoritaire qui en Europe (Berlusconi, Aznar) ou aux Etats-Unis (Bush) a su séduire, au-delà des personnes âgées et des retraités, des classes aisées et de l'élite conservatrice, la plèbe au début du XXI<sup>ème</sup> siècle<sup>80</sup>.

<sup>76</sup> Ici aussi l'approche quantitative est parlante : durant la campagne électorale près d'un 1/3 des phrases prononcées sont des reprises anaphoriques.

<sup>77</sup> [Amossy 1999 : 147] parlerait aussi d'adéquation entre *éthos discursif* et *éthos institutionnel*.

<sup>78</sup> Avec le point d'interrogation (?), le point (.) est hautement spécifique de Sarkozy *versus* les autres candidats à la présidentielle de 2007. La conclusion est facile à tirer : Sarkozy fait des phrases plus courtes que les autres, en moyenne de 22 mots *versus* 33 mots pour Royal. Ce même rapport d'un tiers de moins a été aussi constaté entre la phrase sarkozienne et la phrase gaullienne, pompidolienne ou mitterrandienne.

<sup>79</sup> Sur le lexique et la thématique sarkozienne voir {27, mais aussi 30, 31 et 32, B}

<sup>80</sup> La sociologie électorale a bien identifié le vote sarkozien. Il s'agit d'abord des personnes âgées : 44 % des 65-74 ans ont ainsi voté Sarkozy dès le premier tour (contre 31 % de la population globale = +13). Il s'agit ensuite des classes très aisées : 56 % des personnes aux revenus supérieurs à

### 3.2. Temps discursif et temps politique

Le temps contraint le discours comme il contraint la langue ; dans des diachronies plus serrées évidemment et de manière plus fine. L'individualité discursive d'un locuteur particulier ou collectif, dont il a été question jusqu'ici, est partie prenante de l'histoire ou plus modestement de l'actualité politique que le discours tout à la fois traite et constitue. Si l'identité discursive d'un homme politique se doit de composer avec les contraintes génériques, elle se trouve obligée plus encore de composer avec le temps qui passe. Et en historien du discours ou linguiste de l'histoire discursive, il était naturel d'observer des corpus chronologiques – *les séries textuelles chronologiques* [Salem 1991] – et d'essayer de comprendre le *temps lexical* [ibid.] ou temps discursif qui opère en pareil cas.

Jamais l'approche interdisciplinaire n'est apparue plus justifiée. C'est la posture fondamentale de l'historien que nous avons pu apprendre dans notre cursus universitaire jusqu'à la thèse, et telle que la rappelle par exemple Jacques Le Goff, qui se trouve appliquée à l'objet linguistique que constitue un corpus textuel : embrasser, d'un point de vue discursif, « la dialectique de la continuité et du changement » [Le Goff 1996 : 14], observer les continuums lexicaux et les hiatus chronologiques, repérer les périodes de mutation de vocabulaire ou de rhétorique et les périodes de *statu quo*.

#### 3.2.1. Thorez (1930-1939) : continuum lexical et hiatus chronologique (I)

La mutation du PCF dans les années 30 sous la direction de Maurice Thorez est un phénomène connu de l'histoire politique contemporaine, longuement décrite par l'historiographie. La ligne qu'énonce, et que construit en énonçant, le secrétaire général du PCF évolue durant cette période en trois points fondamentaux. Le parti communiste passe :

- ✓ d'une posture révolutionnaire et marxiste contre l'exploitation capitaliste à une posture républicaine, avec collaboration gouvernementale, de résistance au fascisme.
- ✓ du défaitisme révolutionnaire léniniste au patriotisme jacobin.
- ✓ d'une politique sectaire de refus d'alliance avec la SFIO à une union de toute la gauche (le Front populaire) voire de tous les Français (le « front français »).

Le corpus de notre thèse qui comprend l'exhaustivité des discours de Thorez de l'avant-guerre – 186 discours, 394.642 occurrences – a pu témoigner de cette mutation. Respectivement, ont été enregistrées et chiffrées :

- ✓ la diminution, au fil du temps, des mots marxistes (« classe », « prolétariat », « bourgeoisie », « prolétaire », « révolution »,

---

4500 € mensuel et 48 % des patrons de l'industrie et du commerce ont voté Sarkozy dès le premier tour (contre 31 % de la population, = respectivement +25 et + 17). Cependant, le succès de Sarkozy ne peut se comprendre, numériquement, sans un certain basculement des classes populaires. 27 % des employés et 20 % des ouvriers ont ainsi voté Sarkozy lors du premier tour. Cela reste faible (-4 et -11) mais apparaît supérieur aux suffrages recueillis traditionnellement par les candidats de droite lors des élections précédentes. [Pour une étude fouillée : « L'élection présidentielle : explication du vote » <http://www.csa-fr.com/dataset/data2007/opi20070422-sondage-sortie-des-urnes.pdf> ]

révolutionnaire », etc.) et l'arrivée de mots républicains (« peuple », « république », « populaire », « liberté »).

- ✓ la progression du lexique patriotique (« nation », « France », « pays », etc.) au détriment par exemple des mots anticoloniaux (« impérialisme », « impérialistes », « colonies », « chauvinisme »).
- ✓ l'apparition puis l'augmentation du lexique de l'union (« union », « front », « rassemblement », etc.) en lieu et place d'un vocabulaire particulier dit sectaire comme « social-fascisme », « opportunistes », « traîtres », « social-traître », « renégat », etc.).

Restait alors l'essentiel : dater ces changements, guidé par une histoire du PCF qui insiste sur l'idée de « grand tournant » autour de l'année 1935.

C'est grâce aux travaux d'André Salem sur les séries textuelles chronologiques [Salem 1988 ; 1991 ; 1997] et la forme particulière que prennent le plus souvent les AFC du vocabulaire sur de telles séries, que nous avons pu montrer que la mutation du PCF devait se comprendre moins en termes de tournant subit qu'en terme de maturation lente.

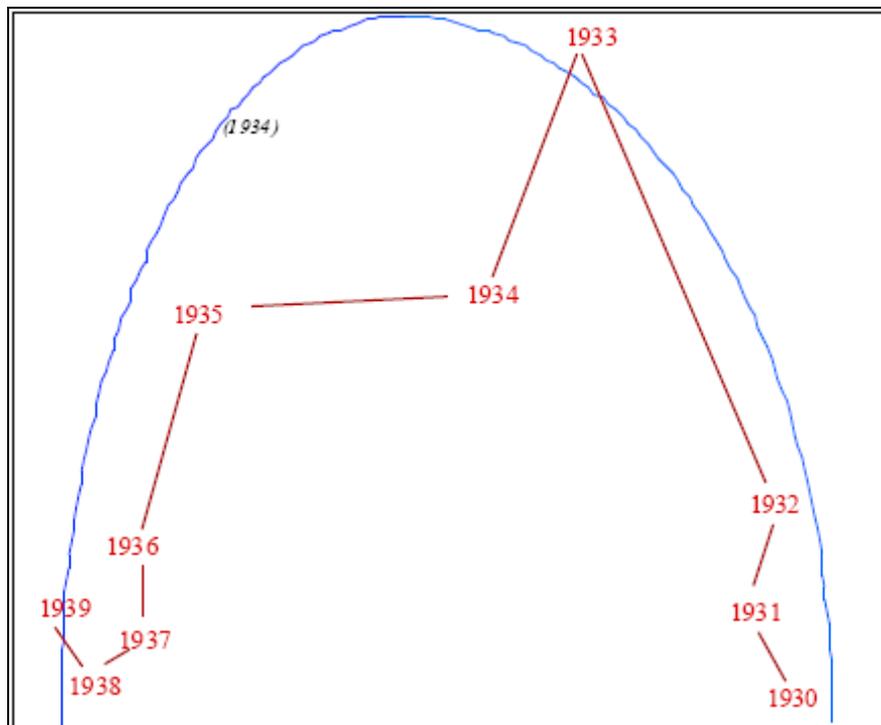


Illustration 14 : AFC du vocabulaire dans le discours de Thorez (1930-1939)

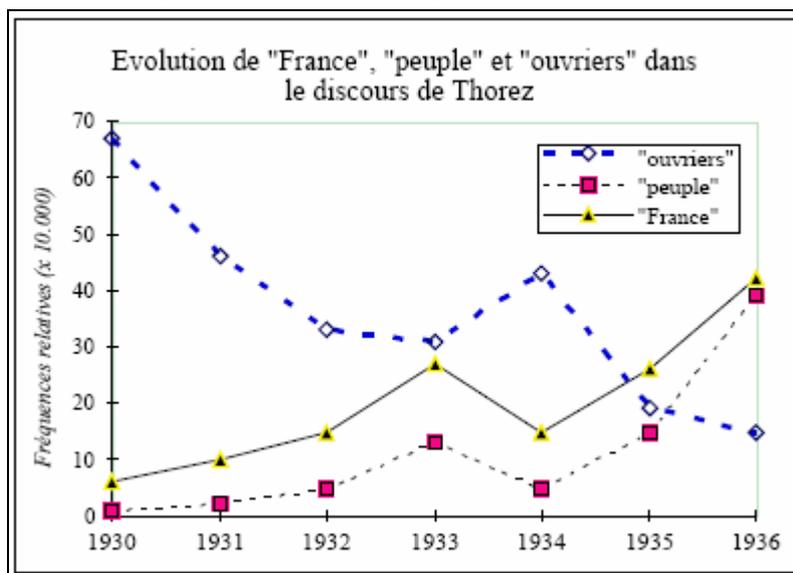
Contrairement aux AFC sur corpus contrastifs pour lesquels la lecture vectorielle se fait par un jeu d'oppositions / attractions selon les différents facteurs (ou les différents quadrants), les AFC sur les séries textuelles chronologiques distribuent, le plus souvent, sur les graphiques, les différentes parties (ici les années) selon une forme parabolique (ici représentée grossièrement en bleu). Connue sous le nom d'effet Guttman, ce phénomène intervient lorsque l'axe premier a un poids déterminant, qu'une structure uni-dimensionnelle peut être mise à jour dans la matrice de données multi-dimensionnelles, et que les parties ont un ordonnancement logique, ou plutôt, ici, chrono-logique.

De fait ici, cet axe premier peut effectivement être qualifié d'axe chronologique et une « gradation » temporelle peut être repérée<sup>81</sup> : de 1930 à 1939, les années se distribuent en effet continûment le long de la parabole, preuve d'un discours en mouvement ; mouvement que l'on découvre précoce et continu sur l'ensemble de la période.

Le sel de l'analyse a alors consisté à expliquer pourquoi la plupart des historiens se sont exagérés le « grand tournant » du PCF voulant voir dans l'année 1935 une forme de rupture [par exemple Peschanski 1989].

Ici aussi l'AFC nous renseigne. C'est l'année 1934, que l'historiographie prend comme point de départ de l'évolution du PCF, qui constitue une anomalie dans le graphique (et dans la période) : sa position attendue (représentée en noir sur le graphique) quitte en effet la parabole idéale pour être attirée par les discours du début de la période dans le quadrant inférieur droit (1930, 1931...). Aussi, loin d'être un point de départ fiable, l'année 1934 offre-t-elle un discours caricatural par rapport auquel l'année 1935 pourra d'autant mieux contraster. C'est donc par un effet de zoom sur une période trop restreinte que la rupture entre 1934 et 1935 est constatée lorsqu'il conviendrait d'insister sur un changement de ligne engagé dès le début des années 1930 et déjà très significatif en 1933.

Concrètement – car les AFC sur l'ensemble du vocabulaire sont toujours délicates à analyser –, la distribution chronologique de trois mots très significatifs (« ouvriers », « peuple », « France ») se présente comme suit.



**Illustration 15 : Distribution chronologique de « ouvriers », « peuple », « France » chez Thorez (1930-1939)**

<sup>81</sup> Dès 1991, André Salem a donc décrit le phénomène. Nous renvoyons aussi le lecteur à un article récent de bonne vulgarisation : Flament et Milland, « Un effet Guttman en ACP », *Mathematics and Social Sciences*, 4<sup>e</sup> année, n°171, 2005, p. 25-49. Les auteurs résument ainsi la problématique : « Le recours à des méthodes multidimensionnelles dans le traitement des données permet d'extraire des structures de données unidimensionnelles, lorsqu'elles existent, et l'on parle alors d'effet Guttman » (p.25)

La progression ou diminution des mots, ici mesurée par l'indice le plus élémentaire et le moins contestable (fréquences relatives), commence déjà entre 1930 et 1931, puis 1931 et 1932, encore entre 1932 et 1933. Seulement, l'année 1934 brise la mutation engagée pour revenir à une pratique lexicale antérieure. Et ce double constat – l'évolution précoce et le retour en arrière de 1934 – pourrait être généralisé pour la plupart des mots sensibles du discours (i.e ceux qui marquent la mutation historique du parti communiste d'un parti prolétarien des origines au parti populaire de l'après guerre)<sup>82</sup>.

Au-delà du cas de Thorez, que nous venons d'illustrer globalement, puis que nous venons de nuancer avec l'extraordinaire année 1934, l'idée rencontrée tout le long de notre travail, dès lors que nous avons affaire à des corpus chronologiques, est la suivante. Un locuteur ne peut pas plus abandonner un discours ou un vocabulaire anciens *ex abrupto* qu'il ne peut s'inventer un discours ou un vocabulaire nouveaux *ex nihilo*. Il compose nécessairement avec son identité linguistico-politique qu'il fait évoluer progressivement – sinon totalement linéairement. Plus encore que pour le factuel politique, le discours politique ne saurait avoir de ruptures chronologiques. La lecture systématique et quantitative des corpus chronologiques montre qu'une pratique discursive doit s'interpréter en tant que continuum. En ce sens, la logométrie est précieuse pour l'historien qui a tendance à ne voir l'événement – l'éventuelle « rupture » – qu'une fois advenu, comme l'on ne regarde les fleurs qu'une fois écloses. Sur de grands corpus et avec une telle méthode, il est au contraire possible de voir les tendances profondes et repérer le bourgeonnement précoce d'une pensée, d'une ligne, d'un discours. Précisément, en Histoire, l'approche événementielle ou histoire-bataille – celle qui insiste sur l'événement fondateur, la date charnière, le moment-clef – a été en partie abandonnée au profit d'une histoire plus profonde qui insiste sur les mouvements lents de l'économie, des mentalités, des sociétés. Dans cette perspective, sur un objet aussi réduit que peut l'être un corpus chronologique de 10 ou 20 ans, ce sont le plus souvent des changements nécessaires mais graduels, des évolutions réelles mais progressives qui ont été observés. Il existe dans les séries textuelles chronologiques un continuum discursif comme il existe un continuum historique ; le temps lexical ou temps discursif s'impose au locuteur mais à un rythme fait autant de lenteur que d'accélération, autant de mue progressive que de saut chronologique ; ainsi l'identité discursive, qui ne saurait s'inventer subitement, se trouve-t-elle toujours en mouvement sans jamais se dénaturer complètement.

### 3.2.2. Jospin versus Chirac : continuum discursif et hiatus chronologique (II)

Une des plus belles sorties machine qu'il nous a été donné d'interpréter a été produite dans l'ouvrage {II : 231} puis les articles {14} et {16}. Tout à la fois,

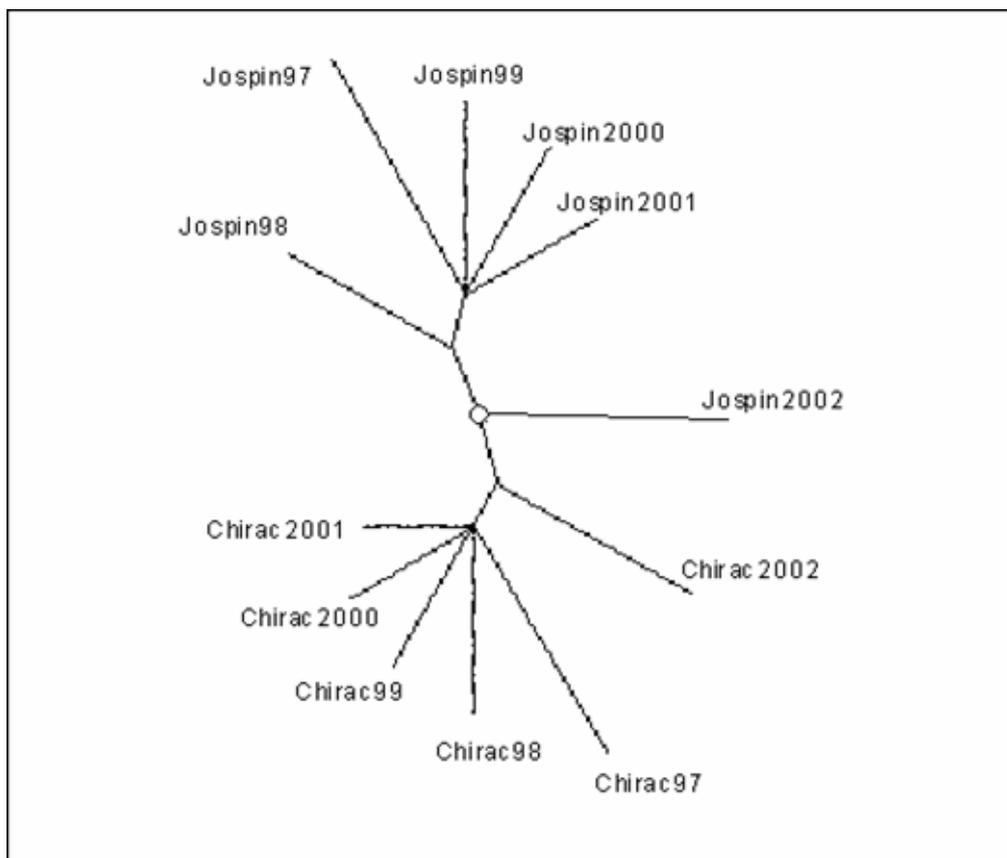
- ✓ elle croisait les deux problématiques majeures de notre réflexion sur le langage politique qui viennent d'être présentées : la question de l'identité linguistico-politique des locuteurs d'une part et la question du temps discursif qui amende cette identité pour imposer au gré de l'actualité historique de lentes mutations au

---

<sup>82</sup> Pour poursuivre l'analyse socio-linguistique, analyser ce que signifie la mutation du PCF et comprendre l'année 1934, voir {I : 407-481, particulièrement 468-481 ; et 4}.

discours d'autre part. (Transparaissait aussi en filigrane la question de la contrainte générique pesant sur les discours).

- ✓ Elle mettait à l'épreuve la méthode logométrique dans un de ses outillages les plus puissants et les plus pointus : le calcul de la distance intertextuelle selon diverses méthodes, et sa représentation par des arbres non plantés qui allie technique classificatoire et approche topologique. La satisfaction était ici autant personnelle que collective car ces outils venaient d'être développés au laboratoire avec la parution de [*Corpus 2*, 2003], et l'implémentation dans Hyperbase de la méthode de Luong avec lequel nous co-signions un article {14} afin de vulgariser l'analyse arborée en SHS. Depuis, plusieurs chercheurs ont adopté cette approche niçoise initiée par [Luong 1988], puis [Barthélemy et Luong 1998] comme [Calvet et Véronis 2008] ou [Labbé et Monière 2003].
- ✓ Enfin, elle permettait, à chaud, de donner une grille de compréhension scientifique au traumatisme citoyen que la société française venait de vivre : la qualification, en France, en 2002, de l'extrême droite xénophobe au second tour de l'élection présidentielle au détriment du candidat de la gauche républicaine. Elle donnait aussi à réfléchir sur un phénomène majeur, récent dans notre vie politique, et qui ne cesse d'interroger historiens et politologues : la cohabitation au sommet de la V<sup>ème</sup> République, en la personne du président de la République et du Premier ministre, de la majorité politique du pays et de l'opposition : inévitablement l'analyste du discours politique contemporain est à un moment donné aux prises avec l'actualité politique.



**Illustration 16 : Analyse de la distance intertextuelle du vocabulaire dans le corpus-cohabitation (1997-2002)**

Trois conclusions doivent être reprises :

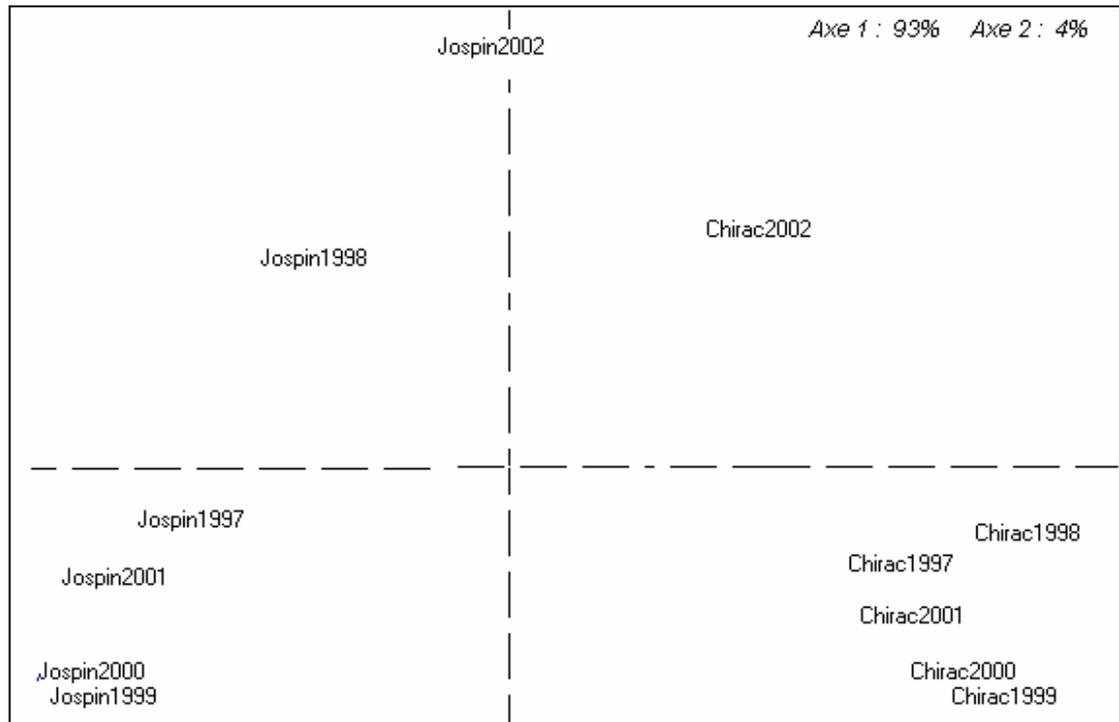
- ✓ L'identité discursive (ici chiraquienne vs jospinienne) existe et est prégnante dans les corpus textuels : c'est elle qui toujours permet de décrypter un corpus contrastif, et offre la meilleure typologie des textes. La méthode sans difficulté reconnaît lexicalement (ici) comme grammaticalement (ailleurs) Chirac et Jospin, et classe leur discours en deux branches différentes. Politiquement, on peut gloser sur la ressemblance entre la social-démocratie molle de Jospin et la droite libérale inavouée de Chirac : même en période de cohabitation, durant laquelle les deux têtes de l'exécutif sont parfois obligées de défendre la même position, une distinction discursive perdue, et le citoyen peut chaque jour l'entendre.
- ✓ Cependant, ces identités tendent à diminuer au fil du temps. A mesure que le temps passe les deux identités très éloignées en 1997 se rapprochent ; cela est particulièrement vrai pour Chirac dont les branches qui séparent son discours de celui de Jospin raccourcissent continûment de 1997 à 2002<sup>83</sup>. Nous pensons ici percevoir l'effet émoullissant de la cohabitation pour la vie politique française : le clivage gauche / droite, au fondement de notre vie politique démocratique depuis 1789, tend à s'estomper ; les discours tendent à s'imiter ; sur l'arbre ou dans la matrice de données {consultable en II : 231}, la distance diminue.
- ✓ Enfin, ponctuellement mais de façon extraordinaire, le discours de Jospin en 2002 apparaît mathématiquement extraordinaire. Là où la classification que propose l'arbre renvoie toujours un discours à son locuteur, cette génétique implacable des textes ne sait à qui attribuer le discours du Premier ministre en 2002. Rattaché au centre topologique de l'arbre, le discours de Jospin apparaît sans paternité, comme un discours monstrueux ou mutant. Sous le coup de la campagne électorale, Jospin trahit son identité discursive et par là devient inaudible pour l'auditeur. En 12 ans d'étude, c'est la seule fois que nous avons enregistré une telle trahison, et peut-être tenons-nous là la raison de l'infortune spectaculaire de Jospin à l'élection ?

Outre l'interprétation socio-linguistique, cette étude a permis, méthodologiquement, de souligner la performance de la méthode arborée qui se trouve aujourd'hui utilisée par les praticiens d'Hyperbase. Dans une perspective herméneutique, elle apparaît, d'un certain point de vue, supérieure à l'AFC permettant, selon les mots de son concepteur, « des classifications locales » plus fines que l'AFC, comme autant de pistes interprétatives.

Sur la même matrice que précédemment (l'indice de distance intertextuelle calculé selon le vocabulaire de la cohabitation), l'AFC donne le graphique ci-dessous.

---

<sup>83</sup> Cela est aussi parfaitement vrai, symétriquement, pour Jospin même si l'année 1998 perturbe un peu l'analyse.



**Illustration 17 : AFC de la distance intertextuelle du vocabulaire dans le corpus cohabitation (1997-2002)**

Sans avoir à mobiliser les ellipses de confiance telles qu’elles son implémentées par exemple dans le logiciel de Ludovic Lebart, DTM, l’opposition Chirac / Jospin apparaît clairement. L’axe 1, ultra-majoritaire, met en scène cette opposition et distribue d’un coté les discours du président de la République et de l’autre ceux du Premier ministre ; quant à Jospin-2002 au mitan de l’axe, il refuse de se situer. Jusqu’ici donc, AFC et Arborée produisent la même interprétation. Mais localement au sein des nuages chiraquien et jospinien, qu’il serait bien imprudent d’interpréter, l’AFC touche à ses limites, là où l’arbre classait nettement les années par ordre chronologique et offrait ainsi une grille d’analyse.

Au terme de ce parcours méthodologique, nous militons pour un usage conjoint de l’analyse factorielle et de l’analyse arborée, afin de limiter le risque de sur-interprétation dont l’une et l’autre méthodes, séparément, sont grosses.

### 3.2.3. De de Gaulle à Chirac : dichotomie discursive

Le travail sur Thorez rappelé ici (comme d’autres travaux sur Blum { **I** : 483-544}, Flandin { **I** : 545-606 } ou Tardieu { **I** : 607-739}), a consisté à mesurer l’influence du temps sur un discours : le locuteur était observé en diachronie pour comprendre la trajectoire chronologique de son parler. Le travail sur Jospin *versus* Chirac durant la cohabitation qu’on vient de voir a permis, lui, d’observer l’évolution de deux identités, en prise l’une avec l’autre. Sur une trame chronologique commune, deux personnalités linguistico-politiques étaient confrontées, et la période de la cohabitation a été une période privilégiée car l’actualité politique qui s’impose aux deux locuteurs est strictement identique.

Dans ces perspectives diachroniques, notre étude sur les présidents de la V<sup>ème</sup> République relève, quant à elle, d'un autre type d'approche : une période historique longue est observée mais grâce à des locuteurs successifs ; ou dit autrement la succession des présidents incarne la trame chronologique mise à l'étude. Comme indiqué plus haut, nous touchons là aux limites de la méthode puisque les mouvements observés même s'ils semblent dus, dans certains cas, à un changement d'époque sont aussi le fait d'un changement de personnalité politique.

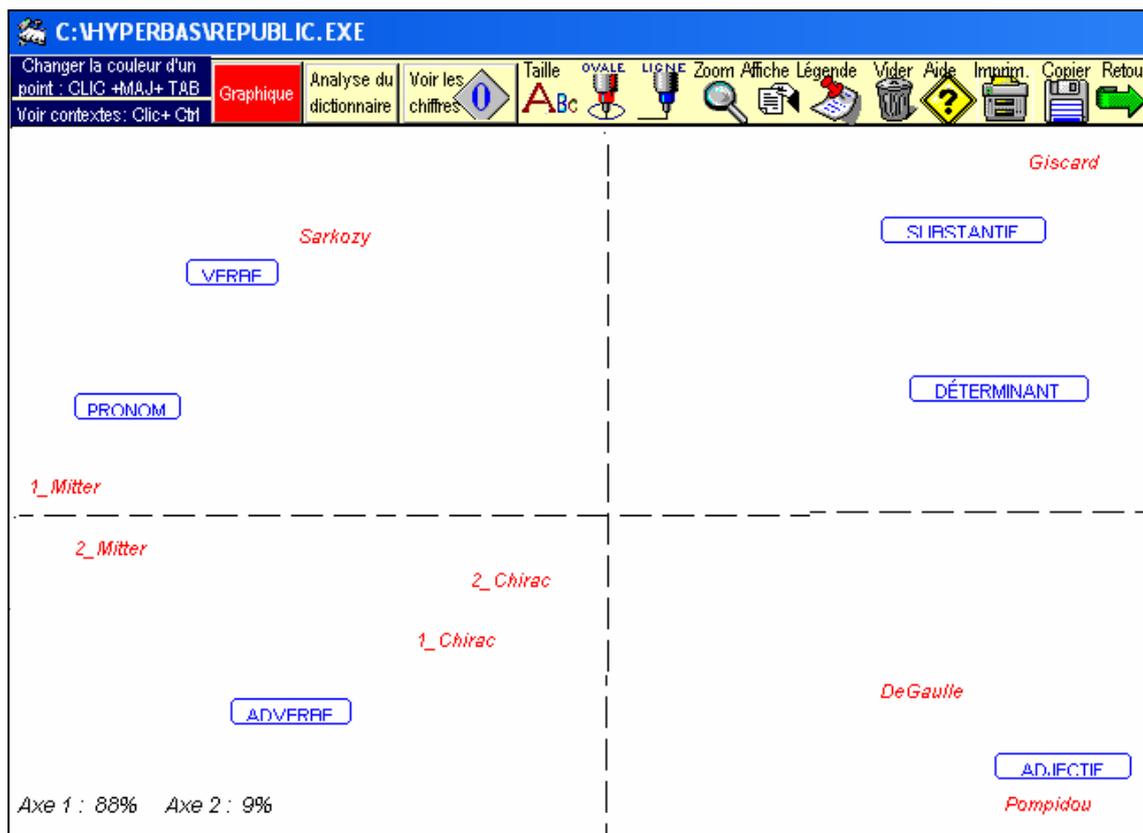
Du reste, sur le corpus présidentiel (1958-2008), ce n'est plus l'évolution continue que l'on avait appris à reconnaître sur nos séries textuelles chronologiques qui a été constatée, mais une césure ou une dichotomie autour des années 1980 : cette rupture chronologique, là où il y a habituellement continuité temporelle, se trouve sans aucun doute accentuée par la personnalité discursive de Mitterrand qui accède à la présidence en 1981<sup>84</sup>.

Quelles qu'en soient les raisons, en tout cas, il s'agit d'un phénomène historique majeur que l'historiographie n'a pas souligné suffisamment [voir par exemple la récente histoire de la France contemporaine de Agulhon, Nouschi, Olivesi et Schor 2008]. Tous les indices lexicaux et grammaticaux relevés montrent en effet qu'il existe une coupure discursive (et politique ? et sociétale ?) entre les années 1958-1960-1970 (soit les présidences de de Gaulle, Pompidou et Giscard) et les années 1980-1990-2000 (soit les présidences de Mitterrand, Chirac et aujourd'hui Sarkozy). La démonstration pourrait se faire sur le lexique, ponctuellement (quelques mots forts) ou généralement (tous les mots) (cf. **II** : section I, chapitre 3}, mais le plus spectaculaire concerne la distribution des parties du discours {**II** : Section I, chapitre 2 ; **19**}. L'AFC suivante reprend les données de {**II**} auxquelles ont été ajoutés ici d'une part le deuxième mandat de Chirac et les deux premières années de Sarkozy<sup>85</sup>.

---

<sup>84</sup> Aussi se propose-t-on de bien distinguer les *corpus chronologiques* (comme le corpus présidentiel 1958-2008) des *séries textuelles chronologiques* au sens strict (comme le corpus Thorez 1930-1939). Dans le premier cas, la variable chronologique est certes centrale mais non unique. Dans le second cas, il y a notamment une unité d'auteur, et seul le temps fait évoluer le corpus.

<sup>85</sup> Ces ajouts ne modifient en rien l'analyse. Tout au contraire, ils la confirment dans ses deux principaux points : rupture discursive autour des années 1980 ; « centralité » du discours de Chirac qui tient un discours de synthèse. Cependant par rapport aux conclusions de {**II**}, le discours de Sarkozy méritera une analyse plus poussée : il semble que Sarkozy allie un discours verbal classique de type mitterrandien et un certain goût pour les noms à l'image de Giscard.



**Illustration 18 : AFC des parties du discours dans le corpus présidentiel (1958-2008)**

De ce constat massif qui touche à la structure rhétorico-grammaticale des discours, nous avons conclu que fondamentalement, en 50 ans, le discours change de nature (linguistique) en changeant de fonction (politique) ; à terme, c'est la démocratie (ou *logocratie*), c'est-à-dire ce régime qui depuis la Grèce règle par le discours la relation entre le peuple et l'élu, qui se trouve modifiée.

La signification du passage d'un discours nominal [réurrence des noms, des adjectifs et des déterminants] à un discours verbal ou pronominal [réurrence des pronoms (avant tout le « je »), des verbes (avant tout les énonciatifs et les modaux) et des adverbes] peut en effet se comprendre selon différents modèles linguistiques ou rhétoriques : tous semblent aller dans le même sens socio-linguistique.

Le discours cesse progressivement d'être référentiel pour devenir phatique (Jakobson), il ne cherche plus à renvoyer au monde mais à construire une relation, (nécessairement asymétrique) entre le président et ses administrés.

Le discours abandonne le dictum au profit du modus (Bally), et favorise le symbole sur le sens : le contenu politique s'évanouit dans la mise en forme de l'agir politique ; l'énoncé s'efface au profit de l'exhibition du processus énonciatif.

Le discours favorise désormais l'éthos et le pathos au détriment du logos (Aristote), et le polémique au détriment du didactique (Guespin) : la démonstration laisse la place à la monstration du narcissisme présidentiel, et l'argument devient le plus souvent argument d'autorité. En mots élémentaires, le passage du nom (la « France », la « république ») au pronom (« je »), met au cœur du débat le président, son énonciation,

ses passions et sa puissance, son image auto-construite par le discours, au désavantage de l'idée<sup>86</sup>.

Pour « le pain, la paix, la liberté » disait le slogan électoral du Front populaire dans les années 30. Il résumait ainsi en trois substantifs majeurs un projet social et économique, un idéal de politique étrangère dans une Europe en armes, et un programme de résistance au fascisme. « Yes we can » ou « Ensemble tout est possible » répliquent aujourd'hui Obama et Sarkozy, laissant à chacun le soin de s'imaginer un programme *sans nom* ; plein d'allant sans doute, mais sans nom<sup>87</sup>.

Cette mort – jusqu'à nouvel ordre évidemment – de l'idée politique par saignée nominale au profit de la modalité politique par pronominalisation (i. e. personnalisation : « je »), verbalisation (i. e. : « dire », « répéter », « penser », « devoir », « falloir », « vouloir », etc.) et adverbialisation (i. e. les adverbes en *-ment* ou les adverbes déictiques) a alors été expliquée par deux phénomènes français qui coïncident avec la chronologie mise à jour :

- ✓ la cohabitation d'abord (9 ans au total en moins de 20 ans, entre 1986 et 2002) qui remplace le débat d'idées par la polémique entre les hommes. Gauche et droite gouvernent désormais ensemble, c'est-à-dire assument un même programme qu'il n'y a plus lieu d'explicitier. Et le président et le Premier ministre sont condamnés à se chicaner *ad hominem* d'autant plus violemment que seuls leur style, leur éthos, leur épouse semblent les distinguer encore. C'est avec la cohabitation que la V<sup>ème</sup> République se présidentialise puisqu'un homme seul – le « je » du président qui subsume désormais tout le discours – peut se maintenir au pouvoir contre l'Assemblée nationale. Le moi présidentiel remplace alors le projet politique.
- ✓ le tout télévisuel ensuite, avec l'ouverture au marché des ondes en 1985, la multiplication des chaînes spectacles, le non stop hertzien ou numérique actuel. La télévision n'est pas seulement l'instrument d'une doxa unique, elle réclame, pour le débat qui nous intéresse, une incarnation politique en la personne du leader. Elle ne saurait montrer sur ces écrans des idées politiques. Elle a seulement besoin de vedettes : le génie politique est aujourd'hui télégénie. Après [Cotteret 1991] et depuis notre ouvrage en 2002, [Charaudeau 2007] a montré que cette incarnation télégénique aboutissait logiquement au *peopolisme* qui n'est point déviance mais parachèvement. De manière plus désespérée encore, et plus profonde, [Debray 2007] conclut quant à lui à une *obscénité démocratique* – on exhibe des corps bronzés et des visages refaits – qui n'est plus démocratie.

---

<sup>86</sup> Par ailleurs, nous montrerons *infra* que le parler politique contemporain passe en 50 ans du *récit* au *discours* (Benveniste) : 3.3.2. *Temps verbaux – pragmatique du présent dans le discours de Chirac*.

<sup>87</sup> L'illustration est volontairement caricaturale et dépasse le cadre géographique du travail. Des contre-exemples pourraient sans difficulté être produits : le « désir d'avenir » de Royal, par exemple, en 2007 est un syntagme nominal. C'est donc ici qu'on mesure l'intérêt de l'approche logométrique : au-delà des exemples et des contre-exemples, c'est une tendance de fond qui a été repérée, décrite et interprétée. Les citations sont toujours, seulement, illustratives, mais en logométrie elles illustrent des faits dont on a mesuré la significativité dans le corpus.

Au-delà de ces explications auxquelles nous renvoyons, il est probable que la transformation du discours que nous avons enregistrée témoigne, de manière plus générale encore, et dans tous les pays, de la mort des idéologies<sup>88</sup>. Avec l'effondrement du modèle soviétique, concomitant, lui aussi, avec notre chronologie, un système de pensée unique, économique et politique s'est imposé dans le monde occidental, et, dans les démocraties parlementaires / présidentielles. Dès lors, l'alternative politique s'exprime aujourd'hui plus en termes de personnalités politiques rivales qu'en termes de concepts, d'idées, de contre-modèles. Et l'alternance concerne plus, à la marge, les modalités de l'agir politique que ces personnalités proposent (*je veux sincèrement, il faut plus rapidement, nous devons absolument...*), que fondamentalement les idées-programmes que les partis de gouvernement, en accord sur l'essentiel, pensent et taisent désormais en commun.

## Conclusion

La démarche logométrique est une démarche différentielle ; à ce titre elle est particulièrement adéquate, par exemple, comme le montre [Pincemin 2010], pour la *sémantique différentielle unifiée* [Hébert 1996] qui se développe maintenant depuis plusieurs années notamment dans la lignée des travaux de François Rastier. Pour nous, c'est le corpus qui constitue la norme, au sein duquel l'on étudie les différences ou compare les variations qui font sens.

Dès lors, comme il est au centre de beaucoup d'études linguistiques – quoique sur des diachronies plus vastes – et qu'il est l'essence même des études historiques, le différentiel chronologique (ou la variation diachronique) a été naturellement au cœur de notre travail sur le discours politique du XX<sup>ème</sup> siècle.

Du reste, le paramètre chronologique est aux origines des études logométriques (anciennement lexicométriques). Avec l'attribution auctoriale des textes, leur datation a toujours été une préoccupation importante de la génétique textuelle assistée par la statistique et l'informatique. Plus concrètement, l'essentiel des travaux de St Cloud, qui furent notre première source d'inspiration, a consisté à identifier des pratiques discursives puis à mettre à l'épreuve ces pratiques sur des temporalités politiques (la période révolutionnaire, le 19<sup>ème</sup> siècle, le 20<sup>ème</sup> siècle). Quant aux travaux niçois d'Etienne Brunet, qui furent notre deuxième source d'inspiration, ils ont dressé de vastes fresques chronologiques du vocabulaire français littéraire de *Rabelais à Gracq* [Brunet 2009] ou *de 1789 à nos jours* [Brunet 1981].

---

<sup>88</sup> On sait que le concept de « mort des idéologies » ou « fin des idéologies » est un concept discuté et discutable. On sait aussi que celui-ci a été précocement développé dans un cadre plus général que le nôtre (et plutôt à propos de la mort du fascisme après guerre que celle du communisme à la fin du siècle). On lira à cet effet, dès les années 1960, D. Bell, *The end off ideology. on the exhaustion of political ideas in the fifties*, Harvard University Press, 1962 (nouvelle édition 2000) ou en France, pour une vision chrétienne, G. Cottier, *La mort des idéologies et de l'espérance*, Cerf, 1970 ; et en contrepoint A. Spire. *Les idéologies toujours présentes*, La dispute, 2004. Plus récemment, on lira. D. et I. Parenteau, *Les idéologies politiques*, PU du Québec, 2008, et consultera les actes à paraître du colloque organisé autour de cet ouvrage , « Les idéologies en question » (22 et 23 octobre 2009, St Jean – Canada).

En ce qui nous concerne, comme indiqué immédiatement *infra*, la mort des idéologies désigne modestement l'effondrement du modèle soviétique et de l'idéal socialiste, c'est-à-dire le passage d'un monde bipolaire au triomphe d'une pensée politique et économique unique en occident. Loin des thèses de Francis Fukuyama, ce constat historique apparaît ponctuel ; et toute mort appelle renaissance.

Rappelons enfin que les séries textuelles chronologiques, comme le corpus Thorez ou le corpus Chirac évoqués ici, résolvent pour partie nos interrogations sur la légitimité du parallélisme entre textualité et corporalité, entre l'objet texte et l'objet corpus. Si le texte est un objet avant tout linéaire et le corpus avant tout un objet sériel (cf. *supra* 1.4. *Le corpus comme texte*), les corpus chronologiques sont des *séries linéaires*, dans lesquelles la *discontinuité* intrinsèque d'un corpus comprenant en son sein plusieurs textes peut être mise en balance avec son déroulement, sa progression, sa *continuité* temporelle. A ce titre, par leur richesse, les séries textuelles chronologiques resteront l'un des principaux objets empiriques et théoriques de nos recherches à venir<sup>89</sup>.

### **3.3. Du discours politique à la langue**

En 12 ans de recherche, les corpus étudiés n'ont jamais été considérés comme *des échantillons de la langue*. Ceci par conviction et par nécessité.

D'abord, fondamentalement, parce que nous doutons que la langue puisse être échantillonnée dans des corpus représentatifs de l'ensemble du système et de ses possibles. Notamment en sémantique, le système est un système ouvert et les *effets de sens*, comme autant de phénomènes particuliers, sont produits parfois plus en dépit du code que grâce à lui : le sens d'un texte n'est pas le résultat d'une grammaire consignable dans un manuel mais le fruit d'une interprétation dans le cadre d'une *contextualisation immédiate* que formalise, autant que possible, à un niveau supérieur et global, un corpus bien construit (cf. *supra* 1.2. *Le corpus comme contexte*) et, au niveau le plus bas et local, les co-occurrences {cf. **30**, **31**, **32**}. La lexicologie, aussi, montre que le lexique, malgré les dictionnaires, est un ensemble non fini que les néologismes viennent sans cesse enrichir, et dans lequel la signification des mots se trouve mise en débat par l'usage textuel, co(n)textuel, social. Bref, aucun corpus, aussi bien échantillonné soit-il, ne peut prétendre rendre compte de la langue dans sa variété et sa totalité (avec ses idiolectes, ses formes orales et écrites, ses registres, sa version standard ou ses langues de spécialités, ses variations diachroniques et diatopiques, génériques, stylistiques, etc.), sauf à réduire la langue à quelques règles syntaxiques synchroniques.

La deuxième raison relève des limites de notre formation universitaire, menée jusqu'en 1998 sous l'autorité jalouse de Clio. C'est par manque de compétences fondamentales en linguistique, acquises suffisamment tôt, que nous n'avons jamais eu la prétention de rendre compte de la langue saussurienne telle que la conçoit Bally : interpréter les discours politiques d'une époque donnée apparaît en soi un objectif suffisamment ambitieux auquel nous nous sommes toujours limité.

Reste que l'observation de grands corpus de textes politiques français contemporains peut interroger le linguiste sur la signification en langue de traits discursifs massifs. Particulièrement en effet, la méthode logométrique, en mettant à jour des éléments têtus, récurrents, systématiques du discours peut éclairer leur valeur

---

<sup>89</sup> Ainsi au moment où sont écrites ces lignes, nous rédigeons un article, en collaboration avec André Salem, à paraître et jamais mentionné dans ce mémoire { **35** : *Le discours communiste : ruptures et continuité. Les adresses finales des congrès du PCF (1945-2008)* } dans lequel nous essayons de travailler l'impact du temps sur une production discursive pluri-décennale aussi balisée que les discours de clôture d'un parti politique.

systemique. Il n'y a pas de langue sans discours, comme il n'existe pas de noumène sans phénomène, d'essence sans existence, de puissance sans acte. Mais en retour, l'observation empirique d'un corpus de données textuelles attestées doit s'ouvrir vers une compréhension de réalités qui le transcendent. Comme indiqué, nous ne prétendons pas ici apporter de réponse : puisse-t-il seulement que notre travail ait amorcé, en certains points, le dialogue entre les spécialistes du discours et les spécialistes du système tel [Vion 2006 et 2007] qui a bien voulu s'appuyer sur nos sorties machine pour nourrir sa réflexion sur la modalisation.

### 3.3.1. L'adverbialisation du discours contemporain

Jacques Chirac durant 12 ans (1995-2007) tient un discours difficile à caractériser par l'approche logométrique. Le plus simple serait de qualifier son discours de discours *moyen* au sein du corpus présidentiel, au sens où l'usage qu'il fait des traits linguistiques (mots, lemme, morphosyntaxe, enchaînements syntaxiques) se situe quantitativement à mi-chemin des extrêmes que pourraient représenter le discours de Mitterrand et celui de de Gaulle (cf. par exemple supra les *illustrations 3* et *10* sur le vocabulaire et sur les parties du discours).

Toutefois, une caractéristique majeure a été repérée : la récurrence de l'adverbe.

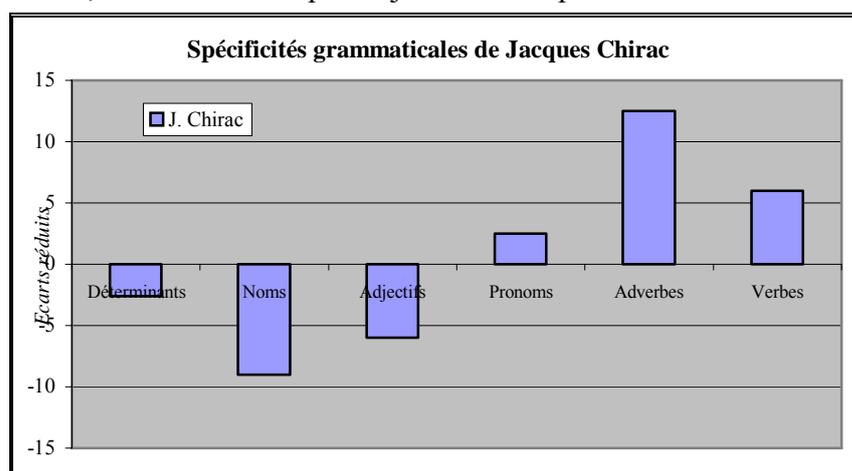


Illustration 19 : Spécificités grammaticales de Jacques Chirac dans le corpus présidentiel (1958-2002)

Comme il est massif (écart réduit de +12), et qu'il épouse une tendance diachronique plus générale qui sépare les discours des années 1950-60-1970 de ceux des années 1980-1990-2000, le constat mérite qu'on s'y arrête : historiquement, sur une soixantaine d'années, le discours présidentiel de la V<sup>ème</sup> République s'adverbialise {cf. une des grandes conclusions de II et supra 3.2.3. *De de gaulle à Chirac*} ; mieux sans doute, si l'on veut bien comparer ces résultats avec ceux entrevus sur le discours de l'exécutif de la III<sup>ème</sup> République, cette tendance s'inscrit dans une évolution séculaire<sup>90</sup>.

<sup>90</sup> La comparaison ici est délicate. D'abord parce qu'il n'y a pas, sous la III<sup>ème</sup> République, de rôle institutionnel comparable à celui de nos présidents ; la comparaison la plus légitime est sans doute avec les présidents du Conseil (Blum, Flandin, Tardieu). Ensuite parce que notre étude doctorale ne bénéficiait pas de lemmatiseur/étiqueteur et abordait le discours seulement via le lexique et les mots graphiques.

De plus, sans avoir les moyens ici de le prouver, cette tendance à l'adverbialisation nous paraît caractériser l'ensemble de la communication politique aujourd'hui, bien au-delà du discours des présidents (voir les études sur le discours ministériel ou parlementaire étudiés par [Marchand 2007] ou [Labbé et Monière 2003] ou les discours des campagnes électorales que nous avons pu étudier {27, 30, 31, 32, 34, B}). Toujours plus loin encore, sans doute pourrait-on constater que le langage quotidien contemporain, particulièrement celui des jeunes locuteurs, est marqué par ce même phénomène ; la littérature universitaire actuelle elle-même (dont la nôtre) n'échappe pas à ce(tte) mode, avec par exemple le sur-emploi d'adverbes d'intensité en *-ment* (« vraiment », « véritablement », « réellement », etc.) ou de « notamment » et « seulement » comme l'illustre [Vion 2007] en disséquant la prose d'un vice-président d'Université.

De fait, l'invention d'adverbes – sans parler ici encore de leur sur-utilisation statistique – est un des processus créatifs les plus productifs de la langue. Facile morphologiquement dans le cas des adverbes en *-ment*, il répond aussi à la contrainte économique du système qui favorise la rapidité dans l'échange d'information : dans le discours politique en tout cas, contraint par des formats courts de communication – parfois quelques secondes seulement au JT pour convaincre – cette économie devient impérieuse. Potentiellement – nous voulions écrire « de manière potentielle » –, tout adjectif, par exemple, peut être adverbialisé par l'ajout d'un simple suffixe (simplicité et économie morphologique), sans que les locuteurs – ici les citoyens – ne doutent du sens à donner au néologisme (transparence sémantique). Si « petitement » a déjà une entrée dans le Robert, « minusculement », sans y être encore, est obvie pour chacun. Le néologisme nominal relève le plus souvent, pour l'initié, du domaine technique – puis de l'extension progressive de ce domaine à la langue standard –, la création adverbiale, elle, peut être immédiatement partagée par tous ; il n'y a pas lieu que l'homme politique s'en prive. Et l'on trouvera par exemple dans la bouche de Chirac « écologiquement » (absent du dictionnaire) comme existe depuis longtemps « économiquement ».

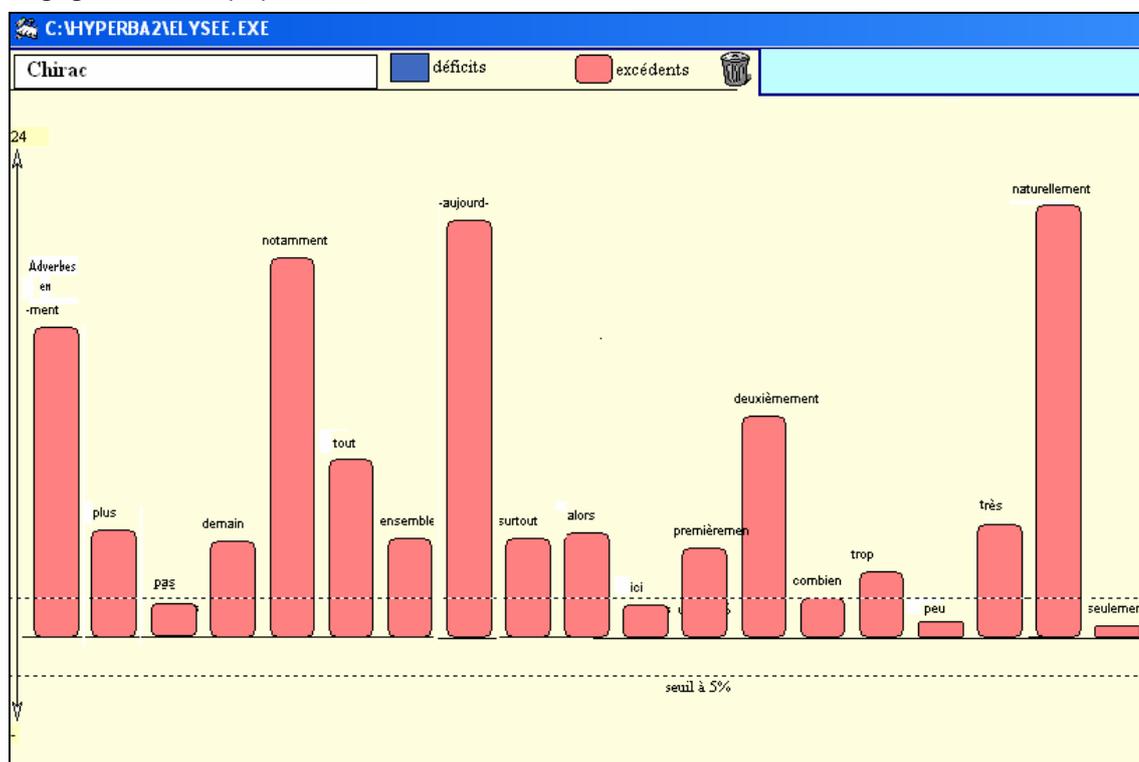
Mais si l'usage de l'adverbe dans le discours de Chirac est pour nous remarquable, c'est parce que tous les adverbes, sauf exception, semblent concernés : *par une utilisation statistique homogène – une sur-utilisation – Chirac donne ainsi une unité à une partie de discours qui ne semble pas en avoir*<sup>91</sup>.

L'adverbe est en effet le casse-tête des grammairiens ; certains linguistes, pour ne rien dire de certaines langues, refusent de le considérer comme une catégorie opératoire. « Classe introuvable » [Wilmet 1998 : 426], « catégorie poubelle » [Chervel 1977 : 251], « ensemble extrêmement touffu et aux contours parfois fuyants » [Dalbera 1980 : 39] : l'adverbe semble rassembler les inclassables de la grammaire française, aussi bien des déictiques (« ici ») que des connecteurs (« ainsi »), aussi bien des mots d'ordination (« premièrement ») que les négatifs (« ne », « pas »), aussi bien la grande classe des adverbes en *-ment* (qu'il faudrait elle-même ventiler) que celle, minuscule, des interjectifs comme « oui » ou « non », etc.

---

<sup>91</sup> Mitterrand (comme Sarkozy aujourd'hui) sur-utilise aussi statistiquement l'adverbe. Mais cette sur-utilisation statistique est très majoritairement due à l'usage immodéré des adverbes de négation « ne... pas ». A ce propos, et sans anticiper par trop la démonstration subséquente, notons que Chirac sur-utilise massivement le « ne... pas » par rapport à un de Gaulle, un Pompidou ou un Giscard. Mais cette sur-utilisation massive est en partie masquée par la sur-utilisation plus massive encore de Mitterrand.

C'est donc dans ces conditions que l'illustration suivante intrigue en pointant le sur-usage adverbial varié et général de Jacques Chirac ; et notons, d'un point de vue méthodologique, que si l'évolution de l'étiqueteur *Cordial*, dans sa version 14 (avril 2008) permet désormais sans trop de fautes de distinguer certains types d'adverbes, il a fallu encore opérer ici manuellement certains regroupements pour préciser le travail engagé en 2004 {II}.



**Illustration 20 : Le sur-usage des adverbes dans le discours de Chirac (par rapport au corpus présidentiel 1958-2002)<sup>92</sup>**

Dès lors, si l'on veut bien exclure l'explication accidentelle, qui mettrait sur le compte d'une coïncidence linguistique la sur-utilisation conjointe, dans un même discours, des différents types d'adverbes, il convient d'essayer de rechercher une explication globale ou unifiante. La tâche apparaît vaine puisque aucun linguiste ne s'y est jamais résolu ; tentons là néanmoins aussi loin que possible.

Comme le montre [Dalbera 1980 : 58], la clef de compréhension de l'unité de la catégorie adverbiale ne peut être recherchée du côté de la syntaxe :

« Il est injustifiable, syntaxiquement, de regrouper des classes, au comportement si différencié sous l'étiquette unique d'adverbes avec, pour seule justification leur invariabilité »

De manière tout aussi évidente, l'explication ne peut pas plus se trouver du côté de la morphologie. Les adverbes sont certes des mots invariables – il s'agirait là de la définition positive la plus sûre, pourtant nuancée par des contre-exemples – mais leur construction et origine sont les plus diverses avec notamment la distinction des adverbes hérités du latin et ceux qui ne le sont pas.

<sup>92</sup> Il va sans dire que nous pourrions trouver quelques contre-exemples. Par exemple « bientôt » ou « comment » sont (faiblement) sous-utilisés par Chirac à hauteur respective de -0,8 et -1,2.

Trouvée nulle part donc dans les grammaires que nous avons pu consulter, gageons qu'une des clefs de compréhension de l'unité adverbiale se situe en sémantique pragmatique ou plus précisément dans le cadre général des théories de la modalisation du discours.

— *De la modalisation adverbiale des discours*. La modalisation du discours ne s'épuise pas dans l'adverbe, mais l'adverbe semble en être une composante essentielle et particulière. Plus précisément, [Vion 2005, 2006, 2007] oppose *expression modale* (d'essence nominale ou verbale) et *modalisateur adverbial*; Jacques Chirac privilégiant, on l'a compris, systématiquement le second sur la première.

Nous reprendrons ici la démonstration en l'illustrant par des exemples issus du corpus présidentiel.

Lorsque Pompidou dit :

(1) *Il est certain* que si la pénurie de pétrole devait se prolonger pendant des mois et des années, tous les pays finiraient par en subir les conséquences, directes ou indirectes. (Pompidou, 20 décembre 1973, entretien télévisé).

Chirac répond :

(2) Le dispositif de surveillance et de précaution qui a été mis en place en France [à propos de la maladie de la vache folle] est aujourd'hui *certainement* le plus complet et le plus efficace de tous les pays européens, et même du monde. (Chirac, 10 novembre 2000, conférence de presse).

L'effet en discours des deux modalités, d'apparence identique, est différent. Dans le premier cas, l'expression modale *il est certain* fait bloc, par le biais de la relative hypotactique, avec le dictum *tous les pays finiraient par en subir les conséquences*. En pareil cas, « dictum et modus étant complémentaires, ils participent de manière congruente au sens de l'énoncé de sorte que la valeur modale [ici la certitude] se retrouve dans l'énoncé global » [*ibid.* : 201]<sup>93</sup>. En revanche, dans le second cas, l'adverbe de modalisation *certainement* qui affecte le dictum *le dispositif est le plus complet et le plus efficace* « ne complète[nt] pas le sémantisme de [l']énoncés au[x]quel[s] [il] participe[nt] » [Vion 2005 : 147], il se contente de faire entendre le point de vue subjectif de l'énonciateur sur son énoncé. Surtout, le linguiste précise que la valeur de *certainement* n'est « pas explicitée linguistiquement [et] se manifestera par le cotexte ou le contexte » [Vion 2007 : 201]. C'est à cet endroit précis – par ce renvoi au co-texte ou au contexte – que nous touchons à l'essentiel : nous y reviendrons longuement. En l'occurrence, loin de présenter une certitude, l'adverbe *certainement* – contrairement à *il est certain* – marque un doute ou au moins une probabilité que l'on ne saurait évaluer sans la prise en compte d'éléments extérieurs, notamment – c'est le cœur de notre propos – sans la prise en compte de la crédibilité de l'énonciateur. Au final, le résultat est peu banal puisque la même affirmation non modalisée (« Le dispositif de surveillance en France est aujourd'hui le plus complet ») apparaîtrait plus *certaine* que celle qui exhibe ce *certainement* rhétorique qui engage / construit – et n'a d'autre fonction que d'engager / construire – l'éthos du locuteur.

Sans la recouvrir, cette réflexion sur l'expression modale – que l'on qualifiera grossièrement d'objectivante – *versus* le modalisateur adverbial – que l'on qualifiera

---

<sup>93</sup> R. Vion exprime la même idée différemment ailleurs en précisant : « ... la modalité de la certitude qui, au sein d'un même mouvement énonciatif, se retrouve dans le sémantisme de [l']énoncé » [Vion 2005 : 147].

grossièrement de subjectivant – fait écho à un autre distinguo classique entre deux types de modalisation. Y compris dans le cadre restreint du seul adverbe modalisateur – nous laissons donc là l’expression modale –, il conviendrait en effet de distinguer les adverbes qui modalisent l’énonciation et ceux qui modalisent l’énoncé (ou une partie de l’énoncé) ou, dit autrement, les adverbes d’évidence pragmatiques et ceux qui ne le sont pas.

Dans, (3) « *Malheureusement*, cela relève du rêve et non pas de la réalité » (Chirac, 14 juillet 1995), l’adverbe d’énonciation *malheureusement* porte sur la totalité du dictum, et il devient facile de distinguer le dit et le dire, le contenu de l’énoncé d’un côté et la modalité – le jugement – que l’énonciateur émet sur ce contenu de l’autre. En président ici à l’ensemble de la proposition, *malheureusement* renvoie directement à l’instance énonciative, c’est-à-dire, *a minima*, au sujet de l’énonciation et à une doxa (prétendument<sup>94</sup>) partagée par le locuteur et l’allocutaire.

En revanche dans (4) « [La croissance] est plus facile lorsque le monde se développe *rapidement* » (Chirac, 14 juillet 1996) l’adverbe final ne détermine qu’une partie de l’énoncé. Ici, qui plus est, le prédicat de *rapidement* est un verbe conjugué à la troisième personne (et non à la première ou la deuxième personne) et sous une forme pronominale réfléchie de sens médio-passif (*se développer*) : en pareil cas, le sujet de l’énonciation ne semble pas devoir être concerné par le phénomène adverbial.

Pourtant, reconnaissons d’abord qu’il est difficile techniquement de faire une approche quantitative automatique de ce genre de distinguo sur un grand corpus. Les meilleurs étiqueteurs ne permettent pas de différencier un adverbe d’énonciation ou pragmatique d’un adverbe qui ne le serait pas. Et sauf d’avoir recours à un étiquetage manuel interminable sur plusieurs millions de mots, et soumis à des arbitrages douloureux, l’opération reste inaccessible à nos études.

Surtout, sans la nier, prétendons d’une part avec la théorie pragmatique que cette différence n’est pas aussi marquée que d’apparence. Nous pensons en effet que même lorsqu’il modalise seulement l’énoncé, l’adverbe reste, selon les mots de [Jaubert 1990 : 101] ce « reflet fugace de la subjectivité » qui « filigrane l’image d’un énonciateur dosant son engagement » : il renvoie toujours, *terminus a quo*, au sujet énonciateur.

Nous prétendons d’autre part et avant tout, que dans le cas du discours de Chirac, *l’effet de masse* que provoque l’accumulation d’adverbes fait sens – *le nombre fait sens* (cf. *supra* 2. *Une méthode : la logométrie*) – et abolit une partie de la distinction mentionnée. Dans une phrase comme :

(5) « Si nous voulons *réellement* lutter contre le chômage avec tout ce qu’il implique en matière d’exclusion *notamment*, nous devons au fil des temps mais *rapidement* changer d’état d’esprit. » (Chirac, 14 juillet 1995, entretien télévisé).

il nous paraît incertain d’essayer d’attribuer une valeur pragmatique différente aux trois adverbes soulignés, dans un ensemble qui fait bloc. Particulièrement ici, le fait que *réellement* puis *rapidement* porte sur des verbes eux-mêmes modaux (*vouloir* et *devoir*)

---

<sup>94</sup> Sans aller trop vite dans la démonstration (cf *infra* le développement sur le « naturellement »), posons que l’adverbe a souvent fonction pour l’énonciateur de supposer une doxa pour mieux l’imposer ; de supposer une complicité entre l’énonciateur et l’auditoire pour mieux la créer. L’adverbe intervient dans le cadre d’une relation intersubjective asymétrique : il renforce la place de l’énonciateur et sollicite le destinataire en anticipant, de manière auto-réalisatrice, sur sa réception du message.

conjugués à la première personne et qui président à la phrase brouille l'analyse<sup>95</sup>. De même, si l'on reprend maintenant l'exemple précédent de (4) dans lequel *rapidement* avait été considéré comme un adverbe d'énoncé sans lien avec l'énonciation, en le replaçant dans son contexte élargi, notre première conclusion pourra être nuancée :

(4-bis) La croissance ce n'est pas un mythe, la croissance il ne faut pas l'attendre, il faut la faire, elle est entre les mains de chacun d'entre nous. Alors naturellement elle est plus facile lorsque le monde se développe *rapidement* et elle est plus difficile dans d'autres périodes, mais elle est entre les mains de chacun d'entre nous. Ce que je veux, c'est que les Français aient *petit à petit* conscience que la croissance c'est le résultat de leur propre comportement. (Chirac, 14 juillet 1996).

Dans ce paragraphe ouvert par les adverbes de négation *ne... pas* (deux fois), par le modal *falloir* (deux fois) et le pronom première personne *nous*, puis continué par un *naturellement* chiraquien prototypique et à nouveau le pronom *nous*, enfin terminé par un *je veux* sans ambiguïté pragmatique, dans ce paragraphe sur-modalisé donc ou sur-embayé, la valeur du *rapidement* n'appartient plus seulement, nous semble-t-il, à l'énoncé mais reflète lui aussi, après d'autres éléments et par effet de récurrence – comme une cuti de rappel sans cesse rééditée – une énonciation sur-représentée<sup>96</sup>.

En d'autres mots donc, pour revenir au modèle de [Vion 2007], l'adverbe modalisateur renvoie, directement ou indirectement, de manière évidente ou de manière tenue, aux *ici et maintenant* du discours, à l'instance énonciative, « au co-texte ou au contexte » : à Chirac.

Les adverbes déictiques si nombreux dans le discours chiraquien sont des caricatures de cette réalité adverbiale qu'il est inutile d'illustrer. Et l'on renverra seulement le lecteur, à titre d'exemple, au contexte d'utilisation des 676 occurrences d'« aujourd'hui » qui font de ce mot le deuxième plus spécifique du discours Chirac (après « naturellement » et avant « jeunes » et « notamment » !).

Mais les adverbes de négation, par exemple, relèvent pour nous, aussi, la plupart du temps, de cette analyse. Si Ducrot semble avoir établi, dans un premier mouvement, une trichotomie utile pour les chercheurs (négation métalinguistique / négation descriptive / négation polémique), [Nølke 1991 et 1992] montre finalement que la négation a toujours, peu ou prou, une valeur polémique et en tout cas polyphonique :

La polémique est intrinsèque à l'emploi de *ne... pas*. [...] Il y a une seule négation *ne... pas*, et elle est polémique. [Nølke 1992 : 65-66]

Or en intégrant / contestant le discours de l'autre, la négation ne fait que marquer plus encore la prégnance de l'énonciateur sur son discours ; en renvoyant à une altérité et à un ailleurs, c'est bien le je-ici-maintenant d'un énonciateur tout puissant qui se trouve centralement confirmé comme cadre, point de départ et point d'horizon du discours. Insistons encore : dire « ne... pas » pour Chirac (ou Sarkozy ou Mitterrand<sup>97</sup>),

<sup>95</sup> Dans le détail, toutes les interprétations nous paraissent possibles notamment pour le *réellement* en début de phrase : (*nous voulons réellement*) + *lutter* ; *nous voulons* + (*réellement lutter*) ; mais encore, selon l'intonation du locuteur, de manière pragmatique *réellement* + (*nous voulons lutter*).

<sup>96</sup> Reflet : [Jaubert 1990 : 104] développe l'image : « Par petites touches cumulables, les différents morphèmes de la modalisation [ici l'adverbe] font miroiter la subjectivité fugace d'un énonciateur, manipulant dans l'élasticité du langage le sens de ses énoncés [...] Mais ce miroitement n'exclut pas les effets réflexifs plus appuyés, à commencer par ceux de la deixis ».

<sup>97</sup> Rappelons que Mitterrand et Sarkozy utilisent plus encore la négation que Chirac qui l'utilise déjà beaucoup.

c'est se poser en énonciateur omnipotent – peut-être pourrait-on reprendre la notion de sur-énonciateur ? – , auquel nous sommes obstinément renvoyés, même quand, *surtout quand*, il entre en (pseudo-)dialogue avec l'adversaire.

Du reste, c'est par la dimension polyphonique, dialogique, polémique de l'adverbe que Vion discute notre travail en reprenant le « naturellement » chiraquien que nous avons longuement mis en évidence, puisqu'il s'agit, entre tous, du premier mot typique de Chirac {II : 160-170 ; 258-259}. Et nous acceptons sans réserve les éléments de la discussion sous condition de bien rappeler, dans la logique de ce développement, que la polyphonie est une des formes de mise en scène énonciative les plus efficaces pour renvoyer au metteur en scène (i. e : au locuteur, à Chirac) : la polyphonie n'est pas cacophonie, elle ne tient discursivement que par la grâce de l'énonciateur qui se trouve par là conforté dans son discours<sup>98</sup>. Pour Vion, en effet, le « naturellement » est un élément dialogique essentiel car complexe voire pernicieux du discours chiraquien. C'est, en termes plus précis, un « leurre dialogique » [*ibid* : 217] qui permet à Chirac de sous-entendre un discours déjà émis avec lequel on entrerait en dialogue, une doxa supposément discutée antérieurement et prétendument admise que le locuteur pourrait ainsi reprendre, un déjà-là – la *nature* supposée des choses – qu'on présente comme connu et partagé et auquel il devient légitime de faire allusion. Effet dialogique donc. Mais leurre dialogique car le « naturellement » chiraquien intervient précisément là où il y a silence et artifice – et non naturel – , contradiction<sup>99</sup>, *a priori* contestable et contesté, voire mensonge. Nous avons donné de nombreux exemples dans {II : 160-170 et l'exemple conclusif 258}. Reprenons-en un seul ici.

A propos des essais nucléaires français dans les eaux du Pacifique, et de la tension diplomatique entre la France et l'Australie qui s'ensuivit, Chirac peut déclarer :

Nos essais n'ont jamais naturellement fait de mal aux moindres poissons. (Chirac, 10 septembre 1995, entretien télévisé).

Le « naturellement » semble ici renvoyer à une commission d'enquête ou un rapport d'experts : rapport d'experts que précisément la France – et particulièrement Jacques Chirac – s'est toujours refusée à commanditer. Il semble renvoyer à un consensus de la communauté scientifique et politique : consensus qui précisément n'existait pas. Plus loin encore, il semble renvoyer à une évidence naturelle incontestable comme si les essais atomiques *in situ* ne pouvaient entrer en contradiction avec la vitalité de la faune et de la flore locales.

En termes simples donc, le « naturellement » de Jacques Chirac simule un débat ou des vérités préalables que l'on aurait débattues lorsqu'il ne renvoie qu'à la personnalité de Chirac, lorsqu'il ne tient que par l'autorité que l'on prête à l'énonciateur-président. Et nous avons essayé de montrer dans des propos difficiles {II :

---

<sup>98</sup> Pas de doute sur cet élément central de la démonstration. Vion rappelle par exemple que le « discours cité n'est pas seulement le discours de l'autre : il est également celui du locuteur qui intègre cet autre discours dans le sien. Cet enchâssement implique donc la présence énonciative du locuteur qui, en dépit d'un effacement apparent, modifie l'orientation de cette parole. » [Vion 2007 : 209-210]. Dans son article sur la polyphonie [Nølke 1999: 9] rappelle que l'énonciateur est toujours « le metteur en scène du jeu polyphonique présenté ».

<sup>99</sup> Nous avons indiqué assez précisément que le « naturellement » chiraquien pourrait être souvent remplacé par « mais ». Sauf que le « mais » montrerait la contradiction ou au moins la difficulté de l'argumentation, là où le « naturellement » cherche à afficher la cohérence de l'argumentation {II : 162-163}.

168-169 et 258-259} que le phénomène devenait spéculaire ou réflexif : le « naturellement » ne tient que par l'autorité de Chirac et l'autorité (contestée) de Chirac ne tient finalement que par la capacité dont il se dote, lui-même, par le discours, de dire aux Français ce qu'il est naturel ou non de penser.

Les « certainement », les adverbes déictiques, les adverbes de négation, le « naturellement » : la valeur globale sémantico-pragmatique des adverbes, en ce qu'ils n'ont d'autres fonctions que de marquer / construire l'énonciation, pourrait sans doute être généralisée.

Le gros des adverbes, ceux à fonction quantifiante (« peu », « beaucoup », « très »...) comme ceux à fonction caractérisante (*grosso modo* les adverbes en *-ment*), n'a de sens que par rapport à une norme ou une axiologie dont le seul locuteur manipule l'échelle : dans le discours politique, plus qu'ailleurs sans doute, aucun jugement de valeur, d'intensité, de manière n'a d'innocence et tous engagent / construisent l'orateur qui les émet. Si le jugement « globalement positif » de Georges Marchais, à propos du bilan historique du régime soviétique, reste dans les mémoires, c'est que par l'adverbe le dirigeant du PCF, pourtant soucieux de montrer à ce moment là<sup>100</sup> ses efforts d'autocritique, avouait tout au contraire son incapacité idéologique à renoncer à un modèle politique qui s'écroulait. Lorsque le président Chirac, en déplacement en Avignon début 2002, glisse à son auditoire :

Nous devons faire plus confiance aux entreprises (Chirac, 12 février 2002, discours en Avignon)

le principal du positionnement politique de Chirac, comme le seul intérêt de la phrase, tiennent dans le seul adverbe. Personne en France, en 2002, ne refusait de faire confiance à l'entreprise. En revanche, un duel politique était engagé entre le président de la République et le Premier ministre, tous deux candidats à l'élection présidentielle du printemps, et un lourd débat s'ouvrait sur le bilan du gouvernement Jospin soupçonné par la loi sur les 35 heures d'avoir défié, pour des raisons idéologiques, le monde économique. Par ce seul adverbe donc, Chirac se construit une image d'un candidat responsable économiquement, soucieux de la compétitivité des entreprises et comparativement supérieur à son adversaire. Certes le bilan économique de Jospin n'était pas négatif factuellement (chômage en baisse, compte public en équilibre, balance commerciale excédentaire, croissance économique), mais Chirac pouvait incarner par la grâce d'un discours auto-réalisateur le *plus* ou le *mieux*.

Et encore : les 448 « notamment » du discours de Chirac qui font de ce mot d'apparence anodine le quatrième mot le plus caractéristique du discours (après « naturellement », « aujourd'hui » et « jeunes ») ont-ils une portée ou seulement un intérêt ? Non sans doute en première lecture, et peut-être s'agit-il là simplement d'un tic de langage. Pourtant par petites touches, les occurrences de « notamment » font entendre la voix d'un orateur maître de son argumentation c'est-à-dire susceptible d'exemplifier sa démonstration ; plus subtilement, ils renvoient à l'autorité supérieure d'un président omniscient ayant une vue générale de la situation en France, qui ne saurait expliquer au peuple l'ensemble des phénomènes mais choisit, dans sa grandeur, d'en dévoiler quelques bribes par quelques exemples « notables » ou bien sentis. Bref,

---

<sup>100</sup> La phrase date du congrès du PCF en 1974. Elle poursuit Georges Marchais qui fut interrogé régulièrement sur la question tout le long des années 1970 et 1980.

les « notamment » ont un intérêt sémantico-pragmatique majeur : ils construisent dans la trame du discours – et n’ont d’autre fonction que de construire – l’image d’un président tout à la fois pédagogue, prudent dans ses exemples, soucieux d’être concret et de rester accessible, et celui d’une autorité supérieure omnisciente ayant bien en main les destinées de la nation.

Au final<sup>101</sup>, comme nous le présentons dans {II}, l’adverbe apparaît bien comme la clef du discours chiraquien par ailleurs vide de contenu et lisse de traits linguistiques remarquables. C’est l’adverbe qui paragraphe après paragraphe, syntagme après syntagme – environ une fois tous les dix mots du discours, c’est-à-dire plusieurs fois par phrase – permet de déplacer le débat de l’énoncé vers l’énonciation, du dictum vers le modus, du monde (tel que représenté dans le discours) vers le locuteur. C’est l’adverbe, par la forme de modalisation du discours qu’il propose, qui, plus subtilement que le pronom à la première personne (du singulier), permet de décentrer le propos du thème du discours pour le centrer sur la personne du président qui l’énonce et développer ainsi une rhétorique de l’éthos. Mot après mot, ou plutôt adverbe après adverbe, comme autant de scintillements, le propos chiraquien « miroite » [Jaubert 1990 : 104] son énonciation, et ne repose que sur la personnalité (la légitimité, la crédibilité, la sincérité, l’honnêteté, l’autorité, la présence qui peut être interprétée comme de la prestance) de celui qui parle. Et en retour – il y a là un cercle fondamental au cœur de notre propos – le discours adverbial a pour principale vocation de construire la personnalité charismatique de quelqu’un qui semble autorisé – parce qu’il se l’autorise dans et par le discours – à émettre un avis sur toute chose. (Ici, nous ne pouvons rappeler que rapidement, encore une fois, que la construction discursive de cet éthos présidentiel était particulièrement nécessaire et problématique pour Chirac – d’où l’insistance – du fait d’une conjoncture historique hors norme. Les conditions fratricides de l’élection en 1995, les grandes grèves de 1996 qui paralysent le pays, la dissolution manquée de 1997 qui jette le président durant 5 ans dans l’opposition, les démêlés judiciaires encore ont privé Jacques Chirac d’une légitimité naturelle durant son septennat : celle-ci – particulièrement du fait de la dissolution manquée<sup>102</sup> – demande donc, jour après jour, à être reconquise par le discours.)

[Amossy 1999 et 2006] a bien résumé récemment le débat sans vainqueur entre Searle et Bourdieu. Pour le premier et la philosophie du langage anglo-saxonne, c’est le langage, par sa dimension performative, qui construit le monde et bâtit l’autorité du locuteur ; pour le second et la sociologie française du dernier quart du XX<sup>ème</sup> siècle, c’est le statut du locuteur, sa place institutionnelle, le rituel social qui donne sens au langage. Cette double réalité ne doit pas se vivre comme une opposition mais comme

---

<sup>101</sup> On l’a compris, nous avons tenté, par défi, une explication unifiante de l’adverbialisation du discours. Nous ne prétendons pas, cependant, avoir épuisé tous les adverbes et tous leurs usages. Le « y » entre autres exemples - sur-utilisé lui aussi par Chirac - aurait mérité une analyse à part. Rappelons la conclusion de Wilmet après sa tentative, qu’il déclare vaine, de définition de l’adverbe : « Que reste-t-il comme élément définitoire ? Rien. Le roi est nu » [Wilmet 1998 : 427].

<sup>102</sup> La défaite électorale sans précédent de 1997 est sans doute le point explicatif crucial. Les présidents tirent, depuis 1962, leur légitimité et leur autorité de l’élection au suffrage universel. Mais, ici, c’est cette légitimité élective de Chirac, conquise en 1995, qui semble annulée par la dissolution manquée de 1997. Par ailleurs et concrètement, comme indiqué infra, la dissolution manquée de 1997 prive *de facto* Chirac de son pouvoir de dissoudre la Chambre à nouveau ; c’est-à-dire, en un mot, supprime la prééminence constitutionnelle du président sur le Parlement.

une complémentarité ou comme une boucle. Dans le corpus de Chirac, nous avons cru pouvoir établir que l’adverbe était le rouage linguistique nécessaire et privilégié de ce langage agi et agissant, légitime et légitimant, produit et réalisant, d’un président en titre qui se forge, par son discours, une stature de président.

### 3.3.2. Temps verbaux – pragmatique du présent dans le discours de Chirac

Entre langue et discours, l’étude des temps verbaux est devenue une entrée consacrée dans les corpus textuels, notamment parce qu’elle prétend pouvoir, depuis les travaux de Benveniste puis Weinrich, établir une typologie robuste des textes.

De fait, en 12 ans de travail, une des sorties machine les plus heuristiques qui nous ont été données d’observer, avec la représentation arborée de la distance intertextuelle de Chirac *versus* Jospin durant la cohabitation (cf. *supra* illustration 16), fut l’usage contrasté que font de Gaulle, Pompidou, Giscard, Mitterrand et Chirac des principaux temps de la langue française.

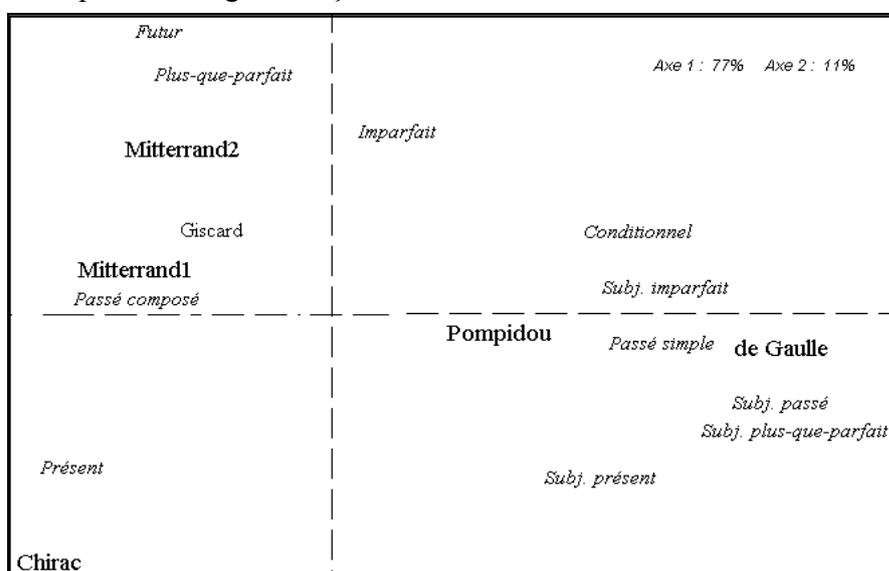


Illustration 21 : AFC des temps verbaux dans le corpus présidentiel (1958-2002)

Dans ce graphique, c’est le positionnement particulier de Chirac qui nous a intéressé {II : 112-122} et que nous aimerions ici analyser plus avant. (Avec cette question simple mais impérieuse statistiquement : comment expliquer le sur-usage quasi exclusif du présent de l’indicatif dans le discours chiraquien ?). Mais une considération générale s’impose au préalable de cette réflexion focalisée sur le seul discours de Chirac.

Nuancée par la plupart des auteurs contemporains, contestée violemment par exemple par [Reboul et Moeschler 1998 : 99-121], l’opposition entre *discours* et *récit* apparaît bel et bien comme une grille de lecture importante du graphique. C’est notamment l’usage du passé composé (et du présent) d’un côté et du passé simple de l’autre qui explique la géographie du graphe<sup>103</sup>. D’un point de vue mathématique, passé

<sup>103</sup> Une autre grille de lecture, complémentaire, est aussi intéressante : l’opposition *grosso modo* entre l’indicatif d’une côté et le subjonctif / conditionnel de l’autre. La superposition de ces deux

composé et passé simple contribuent en effet fortement à la construction de l'axe premier (qui concentre 77% de l'information), respectivement à hauteur de 156/1000 et 286/1000, là où, de manière idéale pour la réflexion benvenistienne, l'imparfait refuse non seulement de se situer sur l'axe premier (il est au milieu du graphe dans une lecture de gauche à droite) mais de contribuer à sa constitution (valeur quasi nulle de 4/1000). Nous retrouvons ainsi de part et d'autre de l'axe horizontal les présidents favorisant *l'histoire*, le *récit*, la *narration* et ceux favorisant le *discours*. Comme toujours dans notre travail, la preuve est statistique et l'on prendra les extraits suivants à titre illustratif (en sachant que l'on pourrait ponctuellement illustrer en sens inverse).

Lorsque de Gaulle raconte, Mitterrand polémique :

De fait, le moment vint où chacun y vit clair. Car les auteurs de la crise se trouvèrent, par ses dimensions mêmes, conduits à poser la question du régime, autrement dit celle du chef de l'Etat. (De Gaulle, 9 septembre 1968, conférence de presse ; à propos des événements de mai 1968 vieux de seulement 3 mois)

Non, vous savez, personne, à vrai dire, ne m'entraînera dans des querelles partisans. Beaucoup cherchent naturellement, à m'y mêler mais moi, je ne le veux pas. Je compte mener à bien la politique dont je suis responsable - qui ne peut pas plaire à tout le monde - mais je veux aussi sur l'essentiel [...] vraiment rassembler les Français. Moi, je ne m'en suis jamais pris aux personnes. J'ai été pendant longtemps dans l'opposition, on le sait, j'avais souvent la dent dure, mais je ne m'en suis jamais pris aux personnes et je déplore que les passions qui s'avivent aujourd'hui n'épargnent plus rien et offrent à l'opinion publique un spectacle contestable, pour le moins. (Mitterrand, 12 février 1984, entretien télévisé)

Lorsque Pompidou conte et commémore l'ami anglais, Chirac discourt :

Lorsque la France fut littéralement submergée par la puissance hitlérienne, l'Angleterre seule et directement menacée refusa de s'incliner et sauva la liberté du monde. (Pompidou, 15 mai 1972, discours prononcé en l'honneur de Elisabeth II)

Tous ensemble, nous pouvons faire de 1996 une année décisive, une bonne année pour la France. Notre pays joue un rôle important dans le monde. Nous y sommes respectés. [...] L'Europe nous est plus nécessaire que jamais. Elle nous garantit la paix. La France veut y prendre toute sa place. [...] Pour toutes ces raisons, mes chers compatriotes, je suis sûr de notre avenir. Je le sais, certains doutent et souffrent. Il y a des femmes et des hommes en profonde détresse. Ce soir, c'est d'abord à eux que va ma pensée. [...] Mes chers compatriotes, en cette soirée de vœux, je vous souhaite, simplement et chaleureusement, une bonne année, une année sereine, une année heureuse. Nous sommes au début du chemin, mais nous sommes sur le bon chemin. Vive la République ! Vive la France ! (Chirac, 31 décembre 1995, allocution télévisée).

Evidemment, la pertinence de cette typologie pourrait être complétée – comme invite à le faire Benveniste – par l'usage des personnes : nous savons en effet, sans avoir à revenir là-dessus, que les présidents qui acceptent le passé simple sont ceux-là même qui sous-utilisent la première personne, et qu'inversement chez Mitterrand ou Chirac le sur-usage du présent ou du passé composé est concomitant au sur-usage du *je* ou du *nous* (sur le sujet de l'énonciation voir le graphique produit dans {II : 119}). La seule exception concerne Giscard d'Estaing qui met rarement sa propre personne en *je(u)*, et pourtant conjugue son discours passé plutôt au passé composé.

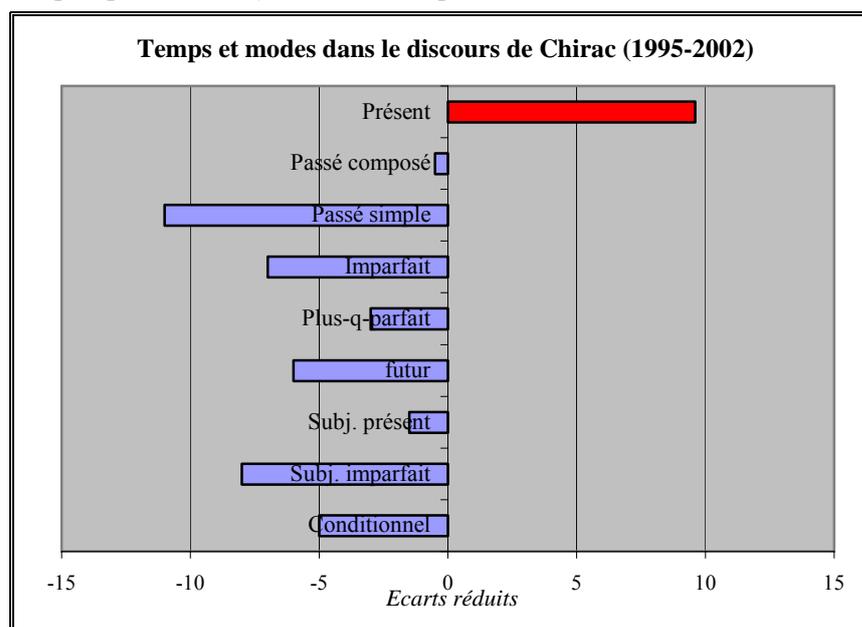
---

oppositions nous permettait de conclure, simplement, au passage chronologique d'un niveau de langue soutenu et en partie littérisé à un niveau de langue ordinaire.

Concluons donc, au-delà de cette anomalie giscardienne, que la logométrie sur un grand corpus politique comme le nôtre, comme sur les discours littéraires [Brunet 2009] est à même de prouver / chiffrer la pertinence de la dichotomie benvenistienne ; un examen qualitatif affiné apporterait sans doute d'infinies nuances mais l'examen quantitatif est, dans un premier mouvement, sans ambiguïté. Pour ce qui nous concerne, nous passons bien à grands traits, en 50 ans, d'un discours-récit, nominal, impersonnel, non-embrayé, à l'énonciation distendue ou masquée et le cas échéant *conjugué au passé simple*, à un discours-discours sur-embrayé, à l'énonciation tendue voir cassante, *conjugué au passé composé* ou au présent<sup>104</sup>.

Quoi qu'intéressante cette réflexion générale doit laisser la place à une réflexion plus particulière : c'est le zoom sur le cadrant comptant Jacques Chirac (cadrant inférieur gauche) de l'AFC précédente qui mérite le plus l'attention de l'analyste. Il semble indiquer la relation privilégiée sinon quasi exclusive de Chirac avec le présent de l'indicatif. Et ce constat linguistique marqué est d'autant plus important pour Chirac que son discours refuse habituellement, comme indiqué, les traits linguistiques trop appuyés ou les signatures verbales caricaturales. Ici pour la seconde et dernière fois (après le sur-usage des adverbes), nous tenons une caractéristique discursive chiraquienne véritablement forte, récurrente, originale.

En termes de spécificités, la distribution des temps verbaux chez Chirac (par rapport au corpus présidentiel) est en effet spectaculaire.



**Illustration 22 : Temps verbaux chez Chirac par rapport au corpus présidentiel ( 1958-2002)**

<sup>104</sup> Rappelons encore et toujours la relativité de nos conclusions. Le discours politique est globalement, par essence, de type *discours*. Et ceci même chez de Gaulle ou Pompidou. Seulement, par contraste et ici en diachronie, nous pouvons conclure que les années 50- 60-70 favorisaient une certaine *narration* là où celle-ci est quasiment exclue chez Mitterrand, Chirac ou Mitterrand. En d'autres termes, comme on le sait, la typologie de Benveniste ne peut être considérée comme une typologie absolue (la *narration* peut infiltrer le *discours* et le *discours* infiltrer la *narration*) mais comme un outil de description comparative.

On rappellera d'abord – car la statistique, particulièrement chez les littéraires, mérite d'être employée avec déontologie – que le graphique comme l'AFC précédente exagèrent visuellement la réalité du discours. C'est *par contraste* – par rapport aux autres – que Chirac semble utiliser seulement le présent. En réalité, on trouvera des temps autres que le présent dans le discours chiraquien. Exactement : 52 % des formes verbales sont au présent de l'indicatif dans le discours de Chirac contre 48 % qui ne le sont pas (passé composé = 8 % ; imparfait = 5 % , futur = 4 %, subjonctif présent = 3 %, infinitif = 22 %<sup>105</sup>, etc.).

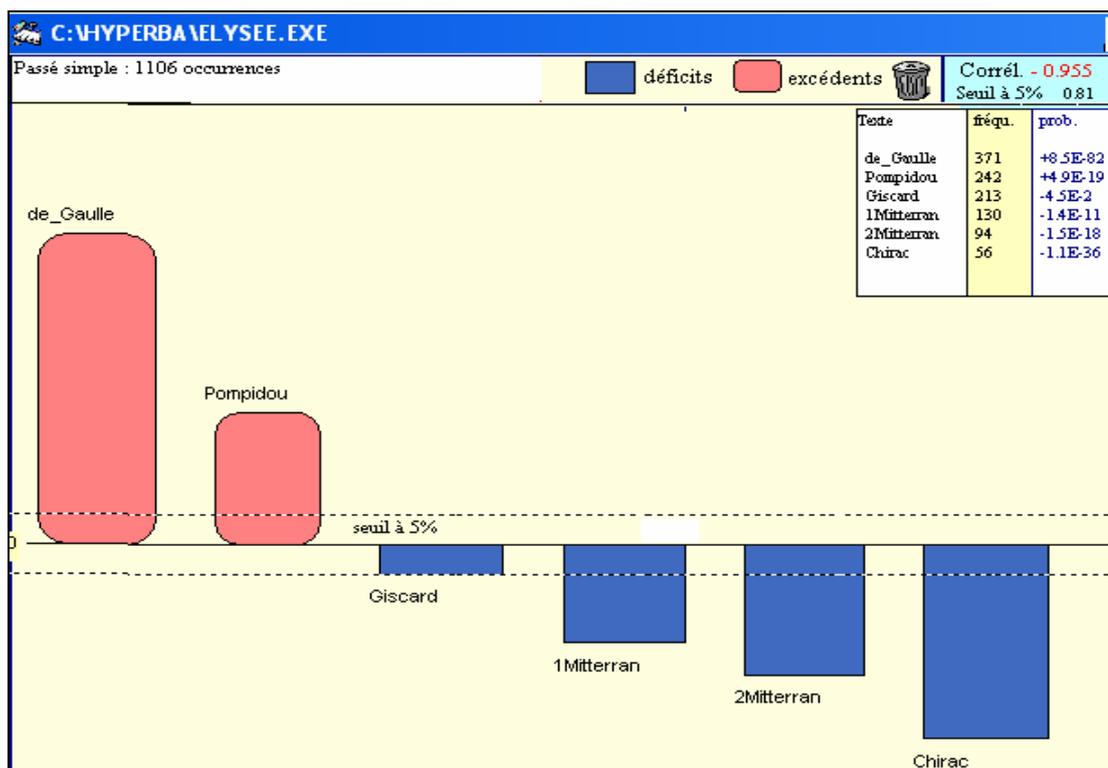
Cependant, au-delà de l'effet visuel amplifié, le constat reste considérable et il convient d'essayer de comprendre pourquoi le discours de Chirac joue d'un tel artifice : pourquoi sur-favorise-t-il le présent au détriment de tous les autres temps ?

Il ne faut pas exclure d'abord que nous enregistrons ici un phénomène langagier profond qui dépasse le seul discours politique et plus encore nos seuls présidents. L'objection est récurrente pour ceux qui travaillent sur corpus et nous l'avons souvent entendue en colloque lors de la présentation de nos travaux : les observations enregistrées – ici une mutation de l'usage des temps verbaux qui mène Chirac au sur-usage du présent – témoignent-elles de la liberté particulière des locuteurs étudiés ou bien des tendances profondes d'une époque entière voire du système lui-même qui s'imposent auxdits locuteurs ?

De fait, sans doute, le français contemporain tend vers une simplification des temps verbaux dont bénéficie le présent chiraquien dans les années 2000. Le subjonctif imparfait par exemple disparaît au fil du corpus entre 1950 et 2000 comme il disparaît de nos discussions quotidiennes et de nos œuvres littéraires contemporaines [Brunet 2009] ; et avec le subjonctif imparfait, c'est de manière générale la concordance des temps qui se trouve sinon abolie en tout cas simplifiée. La distribution du passé simple dans le corpus est particulièrement probante de cette évolution chronologique générale des temps en français. De tous les indices de corrélation chronologique calculés sur diverses unités linguistiques (mots graphiques, lemmes, parties du discours, codes grammaticaux, enchaînements syntaxiques) et sur divers corpus (corpus entre-deux-guerres, corpus présidentiel, corpus électoraux), celui du passé simple sous la V<sup>ème</sup> République a été le plus proche de 1 (ou plutôt de -1, avec -0,955) dans l'ensemble de nos travaux :

---

<sup>105</sup> Nous le verrons, Chirac utilise beaucoup de modaux (« pouvoir », « falloir », « devoir »...). Et la structure [verbe modal (au présent de l'indicatif) + infinitif] explique en grande partie le nombre important d'infinitifs.



**Illustration 23 : Distribution du passé simple dans le corpus présidentiel (1958-2002)**

Le déclin ne peut être plus régulier à mesure que le temps passe, comme s'il était inéluctable ou « structurel ». Particulièrement éclairant est, localement, le déclin entre Mitterrand1 et Mitterrand2 : un même locuteur, expérimenté qui plus est, semble condamné à subir les aléas du temps et se prive du passé simple dans les années 90 plus qu'il ne le faisait dans les années 80.

C'est donc dans ce cadre général qu'il est possible de comprendre comment le présent de l'indicatif s'impose à la fin du siècle chez Chirac, comme le simple triomphe toujours du complexe. Les atouts linguistiques du présent sont connus. Outre sa relative simplicité morphologique, son élasticité lui permet de couvrir tout le spectre temporel du passé lointain au futur proche : avec un seul temps, le locuteur peut dire la pluralité temporelle ; comme la pluralité aspectuelle. Et de fait, dans le corpus chiraquien, nous pourrions produire des exemples de présent d'énonciation, momentané, duratif, d'habitude, dispositionnel, de vérité générale, gnomique, de passé récent, de futur proche, scénique, tabellaire, prophétique, historique ou narratif, de sous-phrase conditionnelle, de phrase conditionnante etc. [Dans le sillage du présent, le passé composé, lui aussi, s'impose dans le discours contemporain pour les mêmes raisons : simplicité morphologique (*être* ou *avoir* + un participe passé unique) ; couverture temporelle importante (un passé (non révolu) qui vient embrasser le présent)].

Pourtant, l'explication systématique par l'évolution historique de la langue française ne saurait suffire : le locuteur Chirac, comme l'attestent quelques passages, maîtrise le futur antérieur, l'imparfait du subjonctif ou évidemment le passé simple. Le sur-usage massif qu'il fait du présent n'est pas seulement le choix passif d'un locuteur soumis aux contraintes d'une langue qui aurait évolué mais relève de la liberté pleine et entière d'un orateur politique qui choisit la stratégie rhétorique qui lui semble la plus efficace pour communier avec l'opinion publique.

C'est ici encore du côté de la pragmatique, et de manière surplombante du côté de la philosophie du langage austinienne, que l'explication la plus cohérente peut être proposée. Le présent dans le discours de Chirac n'est pas un présent momentané ou duratif, ou que savons-nous encore ? C'est un présent performatif. Ou plus précisément, le seul temps qui agréé à Chirac est le présent car le présent est la condition de la performativité du langage. Inversement, supposons que la principale valeur du présent dans le discours politique est d'être – toujours ? – performatif. Et à ce titre nous restons étonné du silence des grammaires françaises qui insistent certes pour dire que les temps verbaux n'ont pas seulement une valeur temporelle mais aussi une valeur aspectuelle, mais qui n'intègre pas dans cette dernière la *performativité* pourtant différente des oppositions aspectuelles classiques entre l'accompli et l'inaccompli, le tensif et l'extensif, le global et le sécant, le perfectif et l'imperfectif, le telique et l'atélitique, le semelfactif et l'itératif, l'inchoatif et le terminatif, etc.

La dimension performative du présent chiracien est facilement plaidable par le détour sémantique des principaux verbes employés. Le discours politique en général, le discours de Chirac en particulier ont ceci de remarquable que l'essentiel de leurs verbes sont soit des verbes énonciatifs ou para-énonciatifs (*dire, penser, parler, croire, savoir...*) soit des modaux (*pouvoir, falloir, devoir, vouloir...*)<sup>106</sup>. Par ordre d'importance, le tableau suivant récapitule les 15 premiers verbes qu'utilise Chirac : à eux seuls ils représentent pas moins de 60% des formes verbales du discours.

**Tableau 1 : Les 15 premiers verbes du discours de Chirac**

Rang 1	<b>Etre</b>	Rang 6	<b>Falloir</b>	Rang 11	<b>Croire</b>
Rang 2	<b>Avoir</b>	Rang 7	<b>Vouloir</b>	Rang 12	<b>Prendre</b>
Rang 3	<b>Faire</b>	Rang 8	<b>Aller</b>	Rang 13	<b>Penser</b>
Rang 4	<b>Dire</b>	Rang 9	<b>Savoir</b>	Rang 14	<b>Parler</b>
Rang 5	<b>Pouvoir</b>	Rang 10	<b>Voir</b>	Rang 15	<b>Demander</b>

Si l'on veut bien maintenant conjuguer ces verbes majoritairement au présent de l'indicatif et souvent à la première personne du singulier ou du pluriel, l'effet rhétorique est garanti. *Etre* et *avoir*, par leur complexité, échappent en partie à l'analyse, de même que *aller* (le plus souvent utilisé comme un auxiliaire de futur immédiat), mais la dimension performative, actantielle, illocutoire ou auto-réalisatrice d'un « je dis... » ou d'un « je vous demande... », d'un « nous pensons... » ou d'un « je veux... », d'un « nous devons... » ou d'un « je crois que... », ou encore d'un « on doit... » ou d'un « il faut... » est évidente<sup>107</sup>. Il ne s'agit pas ici d'entrer dans le détail de la distinction

<sup>106</sup> Et l'on ne caricature pas la réalité du discours politique en soulignant que beaucoup de structures verbales combinent un modal et un énonciatif : « je veux dire... », « il faut dire... », « nous devons répondre... », « il faut savoir... », etc. Ou alors, « je pense que nous devons... », « je dis qu'il faut... ». On trouvera par exemple : « Alors, je ne dis pas qu'il ne faut pas le faire. Je dis que, là encore, on ne peut pas le faire sans le dire et qu'il faut alors que les Français soient conscients que l'on change le pacte républicain » (Chirac, 14 décembre 2000, JT télévisé).

<sup>107</sup> Et le tableau ne rend qu'imparfaitement compte de la réalité du discours politique. A ces 15 premiers verbes, il faudrait ajouter : *répéter, questionner, redire, espérer, souhaiter, répondre, ajouter, permettre, refuser*, etc.

performatifs primaires / performatifs secondaires (ou semi-performatifs), ni de rappeler comment Austin après avoir défini les verbes performatifs *stricto sensu* (*promettre, jurer, ordonner, etc.*), a fini par conclure que tout discours était un acte de parole, que tout acte de parole avait une dimension illocutoire, et que la dimension illocutoire s'idéalise dans la performativité. Dès les années 1960, les plus grands linguistes sont en effet entrés dans ce débat. Et aussi bien [Benveniste 1966 : 267-276] que [Ducrot 1966 : 7-34] ont conclu que si le concept de performativité devait être discuté, il apparaissait dans une définition précise ou extensive, restrictive ou recouvrante, de la plus grande pertinence pour les Sciences du langage<sup>108</sup>. Notre propos est seulement de montrer comment les verbes chiraquiens (énonciatifs ou modaux), conjugués au présent, se réalisent dans le discours en se disant ou s'auto-suffisent / s'auto-établissent comme acte de parole.

— Lorsque je dis « je dis », je dis. C'est-à-dire que je suis, précisément, en train de dire. Et si l'on veut venir sur le terrain des performatifs purs, on pourra considérer que « je dis » dans un discours public, pour un homme politique investi d'une fonction officielle comme un président de la République, a une valeur proche de « je déclare ».

— Lorsque je dis « je pense », je pense ; c'est-à-dire que je suis en train de penser ; en train de penser à voix haute évidemment. C'est-à-dire que « je pense » est proche de « je dis » ou de « je déclare ».

— Lorsque je dis « je veux », je veux (ou feins de vouloir), c'est-à-dire que j'exprime une volonté, qui sans cette expression publique, ici, dans le discours, ne serait pas une volonté politique ; c'est bien au moment où elle est énoncée dans le discours (« je veux... ») que la volonté intime de Chirac devient une volonté politique, un événement historique (c'est-à-dire pour nous un objet de travail). Et s'il faut venir sur le terrain des performatifs purs, « je veux » est quasi synonyme de « je souhaite » (version basse) ou « j'ordonne » (version haute) ou « je promets » (version intermédiaire).

— Lorsque je dis « il faut », il faut ; c'est-à-dire, à nouveau, que c'est le fait de dire publiquement « il faut » qui rend souhaitable, nécessaire ou obligatoire telle ou telle action que Chirac réclame. Bien que conjugué à la troisième personne, « il faut », fonctionne dans la bouche du président comme « je veux », « je souhaite », « j'exige », « j'ordonne ».

— Lorsque je dis « je peux », je peux. Ou du moins je demande à l'auditoire de croire que je peux, à l'image de ce sketch où un fakir-charlatan répond crânement à une question sur ses capacités réelles à réaliser tel exploit : « oui, oui, je peux le faire ». Et le public d'applaudir la performance (verbale).

— Lorsque je dis « nous devons », nous devons. A l'instar de « je veux », mais à la première personne du pluriel, c'est l'expression ou action verbale qui transforme une pensée ou un désir intime du président en programme politique officiel.

— etc.

A ce titre, comme on le sait, – et il s'agit là d'une vertu essentielle pour le discours politique – l'énonciation performative ne peut se juger qu'en terme de félicité ou infélicité et non de vériconditionnalité : Chirac ne sera jamais mis devant ses contradictions ; Chirac ne ment jamais ; il arrive seulement que la réalité que son

---

<sup>108</sup> Au plus haut niveau notamment, en cela qu'elle demande une redéfinition de la dichotomie langue / parole [Ducrot, 1966 : 7-14]

discours construit ne soit pas à la hauteur de la réalité qu'il prétendait construire<sup>109</sup>. Nous retrouvons ici un élément pragmatique essentiel du discours de Chirac longuement développé à propos de l'adverbe. En disant que les énoncés performatifs sont, en discours, auto-réalisateurs et auto-suffisants, on veut dire que leur pertinence ne peut se juger en terme d'adéquation avec le monde extérieur, avec la réalité, avec la vérité. La pertinence d'un énoncé performatif ne doit en effet se juger qu'au regard de la sincérité, la crédibilité, l'autorité du locuteur qui les énonce. De manière élémentaire donc, en deçà de la performativité proprement dite, le verbe chiraquien ne décrit pas objectivement un monde en activité, mais renvoie de manière circulaire au discours, c'est-à-dire, plus précisément, à l'énonciateur qui se met en scène en train de *dire, parler, demander, répéter, questionner, penser, affirmer, répondre* ou qui se montre en train de *promettre, espérer, vouloir, souhaiter, ordonner*. Traitant de *la valeur aspectuelle du présent*, [Mellet 2006 : 170] ne peut mieux exprimer, de manière générale, ce que nous cherchons à démontrer chez Chirac :

La narration [au présent de l'indicatif] n'a d'autre vérité que son énonciation et l'acte énonciatif – la prise de parole – ne se justifie pas autrement que dans l'exhibition de sa propre production.

Dans la liste des 15 verbes les plus employés, les plus problématiques pour la démonstration sont ceux qui en réalité la scellent : « faire » et « prendre ». Ces verbes, en discours et au présent de l'indicatif – ou à l'infinitif dans la cadre d'une structure modale elle-même au présent : *il faut prendre..., nous devons faire..., etc.* – ne décrivent que rarement une action factuelle mais *performent* une action linguistique. Le retour au texte, toujours nécessaire en logométrie, est parlant. « Prendre » est souvent compris dans les expressions verbales du type *prendre un engagement, prendre un exemple, prendre une décision, prendre des mesures, prendre le parti*. Ainsi retrouvons-nous des extraits comme :

Je prends un exemple. Prenons un paquet de biscuits qui coûte 4,80 frs - je dis n'importe quoi, peu importe. Cela représente, si je retiens ce chiffre, 0,731 euros, disons 73 centimes nouveaux. (Chirac, 14 juillet 2001, entretien télévisé)

ou

Je prends aussi un autre engagement : à la fin de chaque année, le gouvernement présentera un rapport d'étape qui fera le point de son action par rapport aux engagements pris. Il engagera sa responsabilité sur ce bilan.<sup>110</sup> (Chirac, 12 avril 2002, *Le Figaro*)

---

<sup>109</sup> Ainsi, un certain nombre de « je vous dis que... » ou « je pense que... » n'aboutissent pas, notamment à cause des interruptions du discours par les journalistes. Plus sérieusement, la modalisation verbale en « je veux » ou en « il faut » soustrait le contenu du discours de l'impératif de réalisation. Le « je veux être le président du pouvoir d'achat » de Sarkozy reste inattaquable linguistiquement (et politiquement) même en période de crise économique et de remontée du chômage, sauf à remettre en cause la sincérité du candidat qui a exprimé ce vœu (cf. immédiatement *infra*). Sarkozy voulait réellement (et veut encore sans doute) être le président du pouvoir d'achat... ou aurait voulu l'être.

<sup>110</sup> Cet engagement de Chirac durant la campagne électorale de 2002 ne sera pas suivi d'effet (infélicité) et les gouvernements Raffarin ou Villepin n'engageront pas leur responsabilité en fin d'année sur leur bilan. Dans le détail, on pourrait s'amuser. Chirac prend l'engagement (engagement 2, car il s'agit d'un « autre engagement » laissant supposer un engagement antérieur dans le discours) que les engagements du gouvernement (engagements 3) seront contrôlés. En l'occurrence, l'engagements 2 de Chirac n'ayant pas été réalisé (infélicité), on ne sait rien, à ce jour, des engagements 3 du gouvernement...

Le verbe « faire » est lui aussi utilisé, dans un sens que l'on qualifiera de figuré, dans des expressions particulières comme *faire l'hypothèse*, *faire le bilan*, *faire la proposition*, *faire confiance*, *faire appel*, *faire un commentaire*, etc. Et on lira les exemples suivants avec la plus grande attention tant la performativité s'y trouve partout, au-delà des seuls passages soulignés :

Je fais toujours la même analyse et je la fais, je dirais, avec encore plus d'inquiétude parce que la mondialisation comporte en elle-même un risque considérable d'exclusion pour les individus mais aussi pour les pays. (Chirac, 14 juillet 1999, entretien télévisé)

ou

Madame Chabot, je ne ferai aucun commentaire sur l'image du RPR. Je fais toute confiance à ses dirigeants pour réussir et je le souhaite. (Chirac, 14 juillet 1997, entretien télévisé)

ou

C'est à chacune et chacun d'entre vous que je fais appel. Je vous invite à vous rassembler le plus largement possible pour construire une société d'initiative et de responsabilité. (Chirac, 27 mai 1997, allocution radio-télévisée)

ou

Je souhaite également que les propositions du délégué à l'aménagement du territoire, que je salue ici, [...] puissent faire l'objet, rapidement, de décisions concrètes de la part du gouvernement et je lui fais confiance dans ce domaine. (11 janvier 1997 discours à Tulle)

De proche en proche, le lecteur aura peut-être deviné le sens du propos. Jusqu'ici la démonstration n'a été qu'un exercice d'école et notre conviction, qu'il faut maintenant énoncer, dépasse largement les exemples des verbes donnés : le discours de Jacques Chirac (particulièrement lorsqu'il est conjugué au présent) est tout entier un performatif ; plus loin encore, nous considérons que tout discours politique doit être considéré comme un meta-performatif<sup>111</sup>.

Après sa définition générale du discours, déjà citée, qui pressent lointainement, nous semble-t-il, l'idée de performativité<sup>112</sup>, c'est Benveniste à nouveau qui nous met sur la voie en donnant une définition de l'énoncé performatif qui recouvre entièrement, constatons-nous, l'idée de discours politique.

Un énoncé performatif qui n'est pas acte n'existe pas. Il n'a d'existence que comme acte d'autorité. Or, les actes d'autorité sont d'abord et toujours des énonciations proférées par ceux à qui appartient le droit de les énoncer. Cette condition de validité, relative à la personne énonçante et à la circonstance de l'énonciation, doit toujours être supposée remplie quand on traite du performatif. Là est le critère et non dans le choix des verbes. Un verbe, quelconque de parole, même le plus commun de tous, le verbe *dire*, est apte à former un énoncé

---

<sup>111</sup> Concrètement, nous pouvons faire présider à tout discours émis sur la scène politique par un locuteur politique une affirmation tacite (ou parfois explicite) préalable du type : « je déclare que + discours », « je promets + discours », « je jure que +discours », ...

<sup>112</sup> Rappelons : « toute énonciation supposant un locuteur et un auditeur, et chez le premier l'intention d'influencer l'autre en quelque manière » [Benveniste 1966, t.1 : 242].

performatif si la formule : *je dis que...*, émise dans les conditions appropriées, crée une situation nouvelle [Benveniste 1966 : 273]

C'est précisément, en effet, *ces conditions appropriées* de la performativité que le discours politique, notamment présidentiel, présuppose idéalement. Le locuteur politique s'exprime seulement parce qu'il jouit socialement de l'autorité pour le faire : sans quoi on ne lui donnerait pas la parole. Surtout, il entend par son discours créer une situation nouvelle : sans quoi il ne prendrait pas la parole. En toute modestie, l'homme politique parle car il espère changer le monde en parlant, et la société l'écoute car elle admet qu'il puisse changer le monde en parlant. Le discours politique n'a pas d'autre vocation ; pas d'autre raison d'être ; pas d'autre définition.

Dans ces conditions, l'objectif devient donc seulement, pour conclure, de comprendre pourquoi Chirac exagère cette prétention performative du discours politique, notamment par le sur-usage systématique du présent de l'indicatif. Nous avons donné deux explications qui toujours renvoient à la situation politique du pays et à la formation idéologique de Chirac.

(i) Chirac est condamné à agir en parlant car il ne peut agir autrement. Au-delà du fait que l'homme politique depuis longtemps ne fait plus la guerre mais *déclare la guerre*, ne construit pas des infrastructures mais *ordonne leur construction*, ne défie pas arme à la main l'adversaire mais *débat* avec lui dans des rituels logomachiques<sup>113</sup>, Jacques Chirac se trouve confronté à une situation historique sans comparaison. Privé de majorité parlementaire durant la quasi-totalité de son septennat, il ne peut faire voter les lois, ni mener la politique du pays : il s'agit là d'un non sens dans la constitution gaullienne<sup>114</sup>. Disgracié par sa dissolution manquée de 1997, désavoué par le peuple, il ne peut plus faire jouer son principal pouvoir constitutionnel, celui de dissoudre (à nouveau) la Chambre, et en cela peser sur elle par son autorité. Quant à la nomination du Premier ministre par le président de la République, elle n'apparaît plus que symbolique au printemps 1997, et Lionel Jospin sera imposé pendant 5 ans à Jacques Chirac par une Assemblée hostile. Bref, le seul pouvoir politique de Chirac fut de pouvoir s'exprimer, et de performer par le langage<sup>115</sup>. Dans le détail, l'analyste restera fasciné en observant ces discours durant lesquels le président vient solennellement rendre compte, à la télévision, d'une politique dont il n'est plus responsable, se situant face à la presse ou l'opinion publique en commentateur extérieur de la politique de la France, ou plutôt en donneur de conseils (« je veux... », « il faut... », « moi, je dis qu'il faudrait... », « je répète que nous devons... ») : donneur de conseils qui n'a pas les moyens politiques de rendre ses conseils effectifs sinon de gager sur leur performativité.

(ii) Chirac n'est pas seulement victime d'une situation historique. Il choisit aussi délibérément d'agir en parlant ; d'agir en parlant, dans le *hic et nunc* du discours et de

---

<sup>113</sup> Nous nous inspirons ici du constat de [Chilton 2004 : 30] : « And only through language tied into social and political institutions can one declare war, declare guilty or not guilty, prorogue parliaments, or raise or lower taxes. »

<sup>114</sup> Avant Chirac, Mitterrand a certes expérimenté la cohabitation, mais comme situation d'exception, imposée par l'histoire, et comme période temporaire. Chirac, lui, est à l'initiative de la dissolution de 1997, et sera en cohabitation pendant 5 ans durant son septennat. Après Chirac, le législateur a rendu peu probable cette monstruosité constitutionnelle en ramenant la durée de la présidence à 5 ans et en faisant précéder les élections législatives par l'élection présidentielle.

<sup>115</sup> On objectera à cette affirmation que le président garde la possibilité de décréter les pleins pouvoirs. Ou encore qu'il garde dans le domaine de la politique étrangère un rôle important.

circonscrire son propos dans un présent performatif et une parole auto-suffisante. Car il en va de sa conception, désabusée (on ne peut pas changer le monde) et conservatrice (il ne faut pas révolutionner l'ordre social en place), de la politique. C'est cette idée que nous avons mise en avant dans {II} pour expliquer le sur-usage du présent de l'indicatif, et qui nous permettra ici de terminer. Chirac refuse ou dénonce explicitement la politique et l'idéologie en tant que projection future ou commémoration passée. Chirac dit ne pas vouloir faire cette politique qui s'enracine dans une généalogie révolue et se dilue dans des projets utopiques à venir. Il entend seulement gérer le moment présent. Gérer de manière pragmatique l'instant présent. (cf. les extraits dans II : 157-158). Cette gestion de l'actualité, en guise de programme politique établi ou d'identité idéologique revendiquée<sup>116</sup>, trouve une consécration dans le présent performatif des discours. Chirac agit en parlant car il n'entend pas agir autrement qu'en parlant, lui (l'énonciateur), ici (à la télévision devant les Français), maintenant (dans la seule actualité d'un propos éphémère conjugué au présent).

---

<sup>116</sup> Les biographes de Jacques Chirac n'arrivent pas à percer les secrets de sa généalogie idéologique [Madelin 2002 : 800]. Chirac, lui-même, dans ses mémoires (tome 1) parues récemment [Chirac 2009], revendique deux généalogies antinomiques : le radical-socialisme et le gaullisme. Quant au programme politique chiraquien, il a historiquement oscillé entre ultra-libéralisme thatchérien et l'interventionisme étatique, entre souverainisme anti-européen et européenisme militant, jacobinisme centralisateur et girondisme décentralisateur [Bacqué 2001 : 72]. Particulièrement, l'attelage électoral que Chirac forme en 1995 avec Alain Juppé, Alain Madelin et Philippe Seguin est apparu hétéroclite à tous : Alain Madelin ministre de l'Economie et des Finances démissionnera du gouvernement Juppé dès les premiers mois du septennat ; Philippe Seguin se retirera progressivement du combat politique jusqu'à son décès en janvier 2010.

## Conclusion générale

Les corpus textuels, la logométrie, le langage politique contemporain : en trois points notre travail depuis 12 ans s'adonne aux joies et *aux limites de l'interprétation*.

Le sens n'est qu'interprétation mais l'interprétation n'a de sens, comme le rappelle [Eco 1992], que lorsqu'elle règle de manière harmonieuse le régime de relations entre l'*intentio auctoris*, l'*intentio operis*, l'*intentio lectoris*.

(i) Il y a dans le discours politique une intention d'auteur qu'on ne peut ignorer. Plus que d'auteur psychologique, étranger à notre travail, nous voulons parler du sujet social ou idéologique qui s'exprime debout dans l'actualité ; et autant que d'intention avouée, superficielle, explicite du message, nous voulons parler d'une intention historique profonde que l'analyste doit découvrir. De manière apaisée, après les débats virulents des années 1970, nous reprenons en effet à l'école française d'Analyse du discours l'idée simple et difficilement contestable qu'un discours est toujours, *via* son instance énonciative, c'est-à-dire avant tout son locuteur, historiquement situé. Et que le contexte socio-politique, dans lequel se trouve immergé l'auteur, de la production d'un texte informe ses contours et son contenu. A ce titre, l'histoire qui a cessé d'être première dans notre parcours reste essentielle dans l'acte interprétatif ; particulièrement cette histoire langagière des concepts et cette histoire des événements linguistiques qui de Pocock à Skinner, de Koselleck à Guilhaumou montrent, dans un mouvement international puissant, l'historicité du langage et la dimension langagière de l'histoire. A ce titre encore, l'histoire contemporaine apparaît comme un terrain d'étude particulièrement probant tant elle semble marquée par l'hypertrophie de la parole politique sinon en elle-même en tout cas dans ses moyens médiatiques démultipliés de diffusion, de circulation, d'amplification. Toujours de manière plus pointue enfin, le discours présidentiel sous la V<sup>ème</sup> République s'avère un objet privilégié d'analyse, tant la fonction du locuteur qui le porte reste constitutionnellement ambiguë. Sans beaucoup de précédents historiques, sans guère de comparaisons géographiques, nos présidents français, à côté du Parlement, à côté du gouvernement, détiennent un pouvoir extraordinaire : pouvoir *performatif* pourrait-on dire en référence aux développements antérieurs sur le présent de l'indicatif – plus que pouvoir symbolique, exécutif ou législatif –, dans lequel la force politique est le verbe et le verbe est action.

(ii) Cependant, le texte a une intention propre que l'AD nouvelle formule a tendance à oublier. Cette *intentio operis* du texte exige une analyse linguistique et un outillage méthodologique adéquats. Ici, on cherche à appréhender les textes (nombreux), recueillis en corpus, en s'appuyant le plus fidèlement possible sur leur matérialité linguistique constitutive. Nous ne croyons pas qu'il existe un sens littéral des textes. La philologie nous a appris à relativiser le signifiant du texte, et la sémantique (interprétative) montre que son signifié est toujours appréciation, construction, interprétation. Cependant, les textes que l'on se décide à saisir en corpus et que l'on

analyse ont bien une forme matérielle empirique donnée de laquelle le traitement démarre, et dont on ne peut se détacher qu'après le plus précis des examens.

Aucune théorie de la réception ne peut faire l'économie de cette restriction préliminaire. Tout acte de liberté du lecteur vient *après* et non *avant* ... [Eco 1992 : 12. Souligné par l'auteur<sup>117</sup>]

Les mots graphiques d'abord qui constituent la surface la moins contestable du texte que le lecteur (et l'ordinateur) peut appréhender, les parties du discours, les catégories grammaticales ou, de manière déjà plus problématique, les lemmes sont des unités textuelles identifiables, repérables, quantifiables, que l'on peut indexer, localiser, compter dans un grand corpus, et sur lesquelles, dès lors, l'interprétation peut s'initier. Plus loin, l'organisation de ces unités textuelles en réseaux de *co-occurrences*, en *séquence* (ici au sens de [Brunet 2006 et 2008]), l'appréhension de leurs *enchaînements*, leurs rythmes d'apparition *en rafale* [Lafon 1984], leurs *scansions fréquentielles d'emplois* selon l'objectif désormais accompli de [Tournier 1980], ou leur combinaison en *motifs* selon l'objectif programmatique des derniers travaux de [Mellet *et alii* 2009 et Mellet et Longrée 2009] sont autant de parcours dans le texte qui permettent de nourrir l'interprétation. Particulièrement, nous avons peu évoqué nos travaux les plus récents sur les co-occurrences {30, 31, 32}. Encore en chantier à ce jour<sup>118</sup> et s'inscrivant dans une longue tradition allant de [Firth 1957] à [Viprey 1997], ils réclament encore réflexion avant de pouvoir prétendre être synthétisés. Notons ici de manière non structurée et sous forme de simples alinéas quelques idées-forces, et retenons seulement, dans la logique du développement, que le traitement constitue un effort pour aller de la matérialité du texte (la co-occurrence comme constat statistique de la co-présence matérielle de deux unités dans une fenêtre co(n)textuelle donnée) vers une interprétation sémantique (la co-occurrence comme champ thématique, isotopie, corrélat sémantique).

- ✓ la co-occurrence doit être abordée dans une perspective textuelle. Abandonnée aujourd'hui en grande partie au TALN, à la traductologie, à la phraséologie, on sait pourtant depuis M. A. K. Halliday qu'elle constitue une clef majeure des études textuelles : elle témoigne au premier chef de l'organisation textuelle, de la textualité ou de la texture<sup>119</sup>.
- ✓ Le passage de l'occurrence à la co-occurrence ou du fréquentiel au co-fréquentiel est un saut non pas quantitatif mais qualitatif qui fait passer nos études de la lexicographie

---

<sup>117</sup> On sait que c'est autour de cette citation que Eco va jusqu'à affirmer, dans une provocation revendiquée, que les textes sont dotés d'un sens littéral. S'il ne peut nous convaincre jusque là, c'est que, dans un glissement significatif, les exemples qu'il donne pour illustrer son affirmation concernent, subitement, *les items lexicaux* [*ibid.* : 12], et non plus les textes ou les énoncés. Eco peut alors faire référence au sens (littéral) des mots consignés dans les dictionnaires. Mais existe-t-il des dictionnaires de textes ?

<sup>118</sup> Cf. {36 : *La cooccurrence, une relation asymétriques ?*}, article soumis et accepté mais encore à paraître, et jamais mentionné dans ce mémoire. Co-signé avec Luong, Brunet, Longrée, Mellet et Poudat, il traite du rapport asymétrique qui peut exister dans une relation co-occurrence : A peut attirer statistiquement B sans que B attire obligatoirement A. Selon les mots du résumé, l'article *propose une nouvelle piste de calcul dont la nouveauté consiste à mesurer et comparer les associations lexicales au sein d'un corpus en recourant aux matrices de dissimilarités et aux calculs de distances, puis à la représentation graphique par l'analyse arborée.*

<sup>119</sup> Pour un point terminologique et conceptuel, dans la littérature française vs. anglo-saxonne, entre *textualité* vs. *texture* voir [Viprey 2009]. Pour un historique critique sur les co-occurrences ou collocations comme centre de la « Corpus Linguistics » (la *London School* ; Palmer, Firth, Halliday, Sinclair) voir [Léon 2008].

(recension de formes dans un corpus ; ventilation desdites formes dans un dictionnaire alphabétique ou dans un dictionnaire de fréquences) à une lexicologie pleine et entière (science du vocabulaire, compréhension sémantique des mots du texte).

- ✓ En effet, si le sens naît en / du co(n)texte (postulat fort de nos études), la co-occurrence prend une position majeure par la définition stricte que nous en donnons : *la co-occurrence est la forme minimale du co(n)texte*. Dire que *A* co-occure avec *B*, c'est contextualiser (c'est-à-dire sémantiser) minimalement l'un par l'autre. Comme nous tenions, par ailleurs, avec les *corpus réflexifs*, la forme maximale du contexte, nos études contextualisantes se trouvent, d'un point de vue théorique et pratique, avantageusement bornées aux deux extrémités.
- ✓ Enfin et surtout, les enjeux théoriques et méthodologiques autour du traitement de la co-occurrence restent importants. Que cela soit au niveau des calculs eux-mêmes (calculs multiples et encore aujourd'hui soumis à examen), des représentations graphiques ou ergonomiques des relations, réseaux, profils de co-occurrences (représentations variées et encore aujourd'hui discutées dans les différents logiciels) ou bien de la signification linguistique profonde du phénomène, les interrogations l'emportent sur les certitudes. Nous ne poserons, arbitrairement ici, qu'un seul problème théorique (illustré par nos études sur *travail, famille, patrie* dans le discours de Sarkozy {30, 31, 32}). Si dans un texte donné, *A* co-occure statistiquement avec *B* et que *B* co-occure avec *C*, quel type de relation existe-t-il entre *A* et *C* ? Ou en d'autres termes, si dans le discours de Sarkozy, *A* (« patrie ») est contextualisé par *B* (« famille ») et que *B* (« famille ») est contextualisé par *C* (« travail »), peut-on dire que *A* (« patrie ») est – indirectement – contextualisé par *C* (« travail »), et peut-on dire, dès lors, que le triptyque pétainiste *travail, famille, patrie* est (re)constitué dans le discours sarkozien ? (Et si on admet cette contextualisation indirecte de *A* par *C*, quel calcul statistique proposer pour la chiffrer ? Quelles représentations graphiques avancer pour la visualiser ?).

(iii) Reste enfin, *l'intentio lectoris* qui inscrit notre démarche dans le cadre d'une sémantique interprétative et non dans une sémantique formelle. Si c'est, comme nous le croyons, la lecture qui *in fine* fait la richesse d'un texte, l'enjeu alors devient patent pour les Lettres, les Arts, les Sciences humaines et sociales : contrôler et expliciter la lecture, contrôler et expliciter les parcours de lecture. En amont de tout traitement automatique, nous avons insisté sur le corpus – sa constitution, son rôle – qui orientera la lecture textuelle. Ici, l'hypothèse du chercheur – celle qui préside à la sélection des textes et à leur organisation en corpus – précède la description / interprétation desdits textes. Dans les termes de l'herméneutique, le corpus peut ainsi être défini comme une matérialisation de *l'horizon d'attente* du lecteur : on se propose donc d'explicitier rigoureusement cet horizon. En aval du traitement linguistique ensuite, la liberté du lecteur se traduit par un bond interprétatif toujours périlleux<sup>120</sup> : on se propose donc non pas de nier ce bond ou cette liberté du lecteur mais de l'encadrer autant que possible pour s'éloigner autant que possible d'une lecture magique. Au milieu enfin, nous militons pour une définition raisonnable de la logométrie comme une pratique herméneutique – une *herméneutique matérielle numérique* – qui ne peut prétendre objectiver le contenu du texte – puisque l'on sait la relativité du signifiant et du signifié du texte, puisque l'on sait le caractère construit des unités textuelles – mais qui espère objectiver les parcours interprétatifs de lecture. Plus loin que la logométrie proprement dite, c'est l'Analyse du discours elle-même qui semble devoir être définie

---

<sup>120</sup> S'il fallait préciser simplement l'endroit où ce bond interprétatif est définitivement consommé, nous dirions qu'il intervient au moment où nous passons du « que dit ce texte ? » à « que veut-il dire ? » ou « que peut-on dire de ce qu'il semble nous dire ? ».

comme une « méthode interprétative par excellence » [Guilhaumou 2006 : 40] dont le but est d'encadrer la subjectivité créative du lecteur.

Un objet, une méthode, une finalité : les trois entrées dans notre travail passé constituent un programme de recherche d'avenir.

(i) Le corpus reste en grande partie aujourd'hui à modéliser ; du reste, il s'agit là du projet collectif de notre laboratoire de rattachement, *Bases, Corpus et Langage*. Le passage au numérique en effet permet et exige de reprendre une réflexion aussi ancienne que l'étymologie du mot ; et ce passage – une révolution culturelle au-delà des simples avancées techniques – est aujourd'hui à peine engagé. Dans notre domaine, comme le notent Adam et Viprey, dans leur appel à contribution pour [CORPUS 2009. *Corpus de textes, textes en corpus*], il semble qu'il y ait actuellement un écart entre des pratiques surabondantes qui se réclament des corpus (textuels) et une théorisation encore déficitaire. Victime de son succès ne serait-ce que par la vulgarisation du syntagme « linguistique(s) de corpus », la notion de corpus reste à travailler. Particulièrement, nous semble-t-il, l'heure est de savoir si la linguistique accepte le corpus – notamment les corpus numériques – comme un objet ; non comme un moyen, un champ d'expérimentation, un réservoir d'exemples mais un objet linguistique *en lui-même* ou *pour lui-même, dans lequel* ou *par lequel* s'épanouit le sens ; bref, le corpus comme condition du déploiement de la sémantique interprétative. Si l'on connaît des linguistes assez hardis pour déclarer le texte comme un des objets de la linguistique, peu, nous semble-t-il, revendiquent le même statut pour le corpus. De fait, objecte-t-on, le corpus n'est pas le produit naturel du locuteur mais celui de l'analyste. Mais le texte n'est-il pas souvent le fruit des éditeurs ? La phrase, le fait du grammairien ? Du reste, si les corpus sont bien artificiels – des objets construits –, certains le sont sous l'autorité de l'auteur lui-même, comme cette sélection éditoriale de discours de [Chirac 2007], ou le site officiel de l'Élysée qui met plus ou moins en valeur tels ou tels discours du président : ici, c'est toute la réflexion autour de la tension entre *archive, corpus* et *discours* qui peut être reprise [Guilhaumou, Maldidier, Robin 1994]. La linguistique de corpus en tout cas, telle que nous l'entendons, serait celle qui prendrait à bras le corps la « corporalité », c'est-à-dire les effets de sens, de cohésion, de cohérence du corpus, ou les conséquences que produit l'immersion du mot, de la phrase, du texte dans un corpus ; les effets réellement produits par *la mise en corpus* [Adam et Viprey 2009 : 18]. Les pistes sont certes aujourd'hui tracées mais l'essentiel reste à réaliser en la matière. La notion de *corpus réflexif* notamment que nous avons avancée pour régler le régime entre textes dans le corpus, c'est-à-dire au niveau de l'intertextualité, se trouve retravaillée actuellement par exemple par [Florea 2009] à un niveau intratextuel : cela réclame confirmation. Plus généralement, les réflexions sur l'architextualité – notion floue depuis Genette mais que des macro-corpus numériques structurés pourraient matérialiser – réclament sans aucun doute un approfondissement, notamment à la lumière des avancées méthodologiques de l'ADT auxquelles nous espérons contribuer, et de celles de la philologie numérique.

(ii) La méthode utilisée n'a pas d'âge. Dans ses fondements, nous ne pouvons pas la dater. A ce titre, il est peut-être ambitieux de vouloir la renouveler. Les concordanciers par exemple sont presque aussi vieux que les livres et sont répandus en Occident dès le XIII<sup>ème</sup> siècle. Les dictionnaires alphabétiques des formes remontent eux à l'Antiquité. Plus récemment, la création d'index (des noms propres, des lieux, des

notions) est devenue la norme dans l'édition moderne. Quant au décompte des unités textuelles, il a sans doute toujours été un fantasme pour l'interprète, et les ordres de grandeur ont toujours servi à décrire le texte comme ils décrivent le monde<sup>121</sup>. De manière moins générale, on considèrera que la logométrie – anciennement la lexicométrie – est solidement constituée en corps de doctrine dans les années 1980<sup>122</sup>. Il s'agit donc seulement aujourd'hui de faire fructifier l'héritage et de prolonger un appareillage informatique et statistique qui a fait la preuve de son efficacité, mais adapté essentiellement d'une part au corpus brut et d'autre part au schéma d'urne. En une phrase qui résume ces deux limites : le texte était jusqu'ici avant tout considéré comme *un sac de mots graphiques* là où on peut le considérer désormais, aussi, comme *un espace d'unités linguistiques établies*. Nous l'avons dit, en effet, deux pistes de recherche doivent être poursuivies. D'abord, sous le double effet de l'enrichissement des données que permettent les nouveaux formats de saisie des textes, comme XML, et le perfectionnement des étiqueteurs automatiques ou des analyseurs morpho-syntaxiques, c'est l'unité de décompte qui demande à être redéfinie. Ou plutôt (car nous avons une idée des unités quantifiables d'un texte), l'enjeu est de pouvoir mettre en correspondance ou en dialogue les traitements quantitatifs que l'on est capable aujourd'hui de faire sur les différents états de texte (texte brut, texte lemmatisé, texte étiqueté / grammaticalisé) : nos modestes propositions dans {23} demandent prolongement. Ensuite sous l'influence de la linguistique textuelle, les traitements quantitatifs doivent s'affronter à la linéarité du texte, c'est-à-dire que des traitements quantitatifs susceptibles de prendre en considération *l'organisation spatiale* du texte doivent prendre le relais des approches probabilistes traditionnelles. Ces deux programmes de recherche (diversification des unités textuelles / approche topologique des textes), quoiqu'indépendants l'un de l'autre, se trouvent exposés conjointement dans la proposition d'avenir de [Mellet et alii 2009 : 109] :

Nous partons du postulat qu'un texte est d'abord un ensemble (E) d'unités linguistiques qui ne sont pas indépendantes les unes des autres, et qui est muni d'une structure ou, plus exactement, de plusieurs structures imbriquées dont l'union constitue cet ensemble.

(iii) Le langage politique enfin est un sujet inépuisable que nous n'avons jusqu'ici qu'effleuré dans notre travail et que les années futures ne suffiront pas à épuiser. Aussi loin que remonte la *polis* : le *logos*. Aussi loin que remonte la démocratie : le discours, la Pnyx, l'agora. Mieux, Athènes la machiste et l'esclavagiste n'a pas inventé la démocratie : elle nous a légué la logocratie c'est-à-dire une organisation sociale et un système de gouvernement des hommes par le discours (même

---

<sup>121</sup> Concordandiers, dictionnaires des formes, index, traitement quantitatif : inutile d'insister sur l'apport de l'ordinateur pour ces pratiques ancestrales. Du texte à l'hypertexte, de la lecture à l'hyperlecture, du décompte manuel à une statistique textuelle moderne : les avancées sont décisives.

<sup>122</sup> On ne reconstituera pas ici une généalogie précise. On sait seulement l'importance des travaux de [Muller 1973 et 1976], des thèses de [Brunet 1978] et [Lafon 1984], de la création du laboratoire de Saint-Cloud et des 10 premiers numéros de la revue *Mots*, du rôle des logiciels historiques comme Lexico ou Hyperbase. C'est peut-être la thèse d'Etat de [Salem 1993], reproduite pour partie dans [Lebart et Salem 1994] qui scelle la discipline. Rappelons que cette généalogie – à laquelle il faudrait intégrer l'œuvre majeure et déterminante de Benzécri, puis la littérature récurrente autour des JADT – est essentiellement française et que le monde anglo-saxon apparaît bizarrement en retard au début du XXI<sup>ème</sup> siècle en matière de statistique textuelle, comme le montre l'artisanat des travaux actuels de [Biber 2004], [Partington et Morley 2004] ou [Teubert 2009].

sophistique), la raison (même déraisonnable), la rhétorique (même manipulatrice). Socialement, notre objet d'étude n'est donc pas anecdotique, et ce n'est point s'aveugler sur le *linguistic turn* que d'affirmer qu'il est constituant et constitutif de la vie des sociétés ; particulièrement sans doute de nos sociétés contemporaines où la doxa se trouve à chaque instant reconstruite par les discours que véhiculent sans cesse la radio, la télévision, internet. Même dans le champ étroit du seul XX<sup>ème</sup> siècle français, même travaillé par de nombreux auteurs avant nous, le sujet reste donc en devenir, ne serait-ce que parce qu'il est un objet en constant renouvellement. Renouvellement en quantité d'abord comme l'atteste par exemple la production quotidienne voire bi-quotidienne d'un discours par l'Elysée (messages, allocutions, interviews, tribunes, etc.). Renouvellement en qualité surtout comme semble l'attester les constats diachroniques que nous avons déjà pu faire sur les dernières décennies, et dont l'intrigue contemporaine – notamment le parler Sarkozy qui s'annonce original – continue de nous tenir. Objet majeur donc qui appelle approfondissement, le langage politique l'est enfin, aussi, pour nous, parce qu'il s'agit d'un objet linguistique privilégié que les linguistes négligent encore et qu'on gagnera dès lors à appréhender. Rares sont en effet les problématiques du texte et du discours, voire celles de la langue qui ne prennent une dimension particulière dans le langage politique. Le dialogisme, la polyphonie, la circulation des discours par exemple, peu évoqués jusqu'ici dans notre travail, ne prennent-ils pas une lumière plus crue dans l'arène politique que partout ailleurs ? L'argumentation dans le discours et dans la langue, toujours en marge de nos réflexions, connaît-elle un champ d'étude plus pertinent que le langage politique ? Et encore, n'est-ce pas de manière sur-qualifiée que le langage politique peut témoigner de la performativité du langage, de la subjectivité dans la langue, de l'intentionnalité des discours ? La question générale, qui traverse toute les linguistiques, de la *référence* ne prend-elle pas une résonance particulière dans des textes politiques parfois instituants, souvent déterminants ? Ou enfin (sans avoir l'idée d'être exhaustif) : les problématiques de l'énonciation (de manière concrète la valeur par exemple d'un *nous* (chez Thorez) ou d'un *on* (chez Sarkozy)) ou plus généralement l'approche pragmatique ne s'exemplifient-elles pas mieux qu'ailleurs dans le *hic et nunc* si contraignant d'un locuteur politique ?

Et gageons que cette acuité linguistique n'est pas étrangère à l'importance sociale mentionnée précédemment : le discours politique est pour nous un programme de recherche fascinant, gros de recherches doctorales à encadrer, parce que le locuteur politique actualise la langue et le monde dans un même mouvement en parlant. Sa performance linguistique et sa performance politique ne font qu'une.

## Bibliographie

- Achard P.(1995). Formation discursive, dialogisme et sociologie. *Langages*, 117, pp.82-96.
- Adam J.-M. (1990). *Eléments de linguistique textuelle*. Bruxelles : Mardaga.
- Adam J.-M. (1999). *Linguistique textuelle. Des genres de discours aux textes*. Paris : Nathan.
- Adam J.-M. (2001). Discours et interdisciplinarité. *Cahiers Ferdinand de Saussure*, 54, pp. 201-218.
- Adam J.-M. (2005). Les sciences de l'établissement des textes et la question de la variation, in Adam J.-M. et Heidmann, U. (éds.), *Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité*. Genève : Slatkine, pp. 69-92.
- Adam J.-M. (2006). Autour du concept de texte. Pour un dialogue des disciplines de l'analyse de données textuelles. *Conférence d'ouverture aux JADT 2006* [texte en ligne sur *Lexicométrie* ([http://www.cavi.univ-paris3.fr/lexicometrica/jadt/JADT2006-PLENIERE/JADT2006\\_JMA.pdf](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/JADT2006-PLENIERE/JADT2006_JMA.pdf))]
- Adam J.-M. (2008 – éd. revue et augmentée). *La linguistique textuelle. Introduction à l'analyse textuelle des discours*. Paris : Colin.
- Adam J.-M et Heidmann U. (éds.) (2005). *Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité*. Genève : Slatkine Erudition.
- Agamben G., Badiou A. et al. (2009). *Démocratie, dans quel état ?* Paris : La fabrique.
- Agulhon M. , Nouschi, A., Olivesi, A. et Schor R. (2008). *La France de 1848 à nos jours*. Paris : Armand Colin.
- Aijmer B. & Altenberg K. (éds). (2002). *Advances in Corpus in Corpus Linguistics*. Amsterdam : Rodopi.
- Amossy R. (dir.). (1999). *Images de soi dans le discours. La construction de l'ethos*. Genève : Delachaux et Niestlé.
- Amossy R. (2006). *L'argumentation dans le discours*. Paris : Colin.
- Anscombe J.-Cl. et Ducrot O. (1983). *L'argumentation dans la langue*. Bruxelles : Mardaga.
- ARGUMENTATION & ANALYSE DU DISCOURS*. (2009). « Ethos discursif et image d'auteur », numéro 3, sous la direction de Michèle Bokobza Kahan et Ruth Amossy.
- Austin J. (1962, trad. fr. 1970, rééd 1991). *Quand dire c'est faire*. Paris : Seuil.
- Bachelard G. (1934). *La formation de l'esprit scientifique*. Paris : Librairie philosophique.
- Bacqué R. (2001). *Chirac ou le démon du pouvoir*. Paris : Michel.
- Baroni R. et Macé M. (éds.) (2007). *Le savoir des genres*. Rennes : Presses Universitaires de Rennes.
- Barthélemy J.-P. et Luong X. (1998). Représenter les données textuelles par les arbres. *4èmes journées internationales d'analyse statistique des données textuelles (JADT'98)*, éd. par Sylvie Mellet, UNSA, Nice, 1998, pp. 49-73.
- Barthes R. (1973). *Le Plaisir du texte*. Paris : Seuil.
- Benveniste E. (1966 et 1974). *Problèmes de linguistique générale*. Paris : Gallimard, tomes I et II.
- Biber D. (1988). *Variation accross speech and writing*. Cambridge : Cambridge University Press.
- Biber D. (1995). *Dimensions of Register Variation : A Cross-linguistic Comparison*. Cambridge : Cambridge University Press.
- Biber D. (2004). Conversation text types : A multi-dimensional analysis, in G. Purnelle, C. Fairon, A. Dister (éds), *JADT04*. Louvain : Presses universitaires de Louvain, pp. 15-31.

- Biber D., Conrad S. & Reppen R. (1998). *Corpus linguistics. Investigating language, Structure and Use*. Cambridge : Cambridge University Press.
- Bommier-Pincemin B. (1999-a). *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*. Thèse de doctorat, Paris IV.
- Bommier-Pincemin B. (1999-b). Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative, in Condamines A. et al (éds.), *Corpus et traitement automatique des langues : pour une réflexion méthodologique*. Cargèse : Actes de l'atelier thématique TALN, pp. 26-36.
- Bouquet S. (1997). *Introduction à la lecture de Saussure*. Paris : Payot.
- Bouquet S. (2005). Après un siècle, les manuscrits de Saussure reviennent bouleverser la linguistique. *Texte !* <[http://www.revue-texto.net/Saussure/Sur\\_Saussure/Bouquet\\_Apres.html](http://www.revue-texto.net/Saussure/Sur_Saussure/Bouquet_Apres.html)>.
- Bourdieu P. (1982). *Ce que parler veut dire : l'économie des échanges linguistiques*. Paris : Fayard.
- Branca-Rosoff S. (1999). Types, modes et genres : entre langue et discours. *Langage et société*, 87, pp. 5-24.
- Bronckart J.-P. (1997). *Activité langagière, textes et discours*. Lausanne-Paris : Delachaux et Niestlé.
- Brunet E. (1978). *Le Vocabulaire de Giraudoux. Structure et Evolution*. Slatkine : Genève.
- Brunet E. (1981). *La vocabulaire français de 1789 à nos jours*. Paris-Genève : Champion-Slatkine.
- Brunet E. (2000). Qui lemmatise dilemme attise. *Scolia*, 13, pp. 7-32.
- Brunet E. (2001). *Manuel d'utilisation d'Hyperbase*. Nice.
- Brunet E. (2007). Le corpus conçu comme une boule in Rastier F. et Ballabriga M. (éds), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation*. Toulouse : Put, pp. 85-94.
- Brunet E. (2009). *Comptes d'auteurs. Tome I : Etudes statistiques de Rabelais à Gracq*. Paris : Champion.
- Brunet E. (2010 – à paraître). *Comptes d'auteurs. Tome II*. Paris : Champion.
- Calas F. (2006). *Cohérence et discours*. Paris : Pups.
- Calvet J.-L. et Véronis J. (2008). *Les mots de Nicolas Sarkozy*. Paris : Seuil.
- Charaudeau P. (2008). *Entre populisme et peopolisme : comment Sarkozy a gagné*. Paris : Vuibert.
- Charaudeau P. et Maingueneau D. (dir.) (2002). *Dictionnaire d'analyse du discours*. Paris : Seuil.
- Charolles M. (1983). Coherence as a principle in the interpretation of discourse. *Text*, 3-1, pp. 71-99.
- Charolles M. (1995). Cohésion, cohérence et pertinence du discours. *Travaux de linguistique*, 29, pp. 125-151.
- Chervel A. (1977). *Histoire de la grammaire scolaire*. Paris : Payot.
- Chilton P. (2004). *Analysing Political Discourse. Theory and Practice*. London : Routledge.
- Chilton P. and Schäffner Chr. (eds.) (2002). *Politics as Text and Talk: Analytic Approaches to Political Discourse*. Amsterdam : John Benjamins.
- Chilton P. and Wodak R. (eds.) (2005). *A New Research Agenda in Critical Discourse Analysis: Theory and Interdisciplinarity*. Amsterdam : John Benjamins
- Chirac J. (2007). *Mon combat pour la France. Textes et interventions 1995-2007*. Paris : Odile Jacob.
- Chirac J. (2009). *Chaque pas doit être un but. Mémoires T1*. Paris : Nil éditions.
- Combettes B. (1983). *Pour une grammaire textuelle*. Bruxelles : De Boeck-Duculot.
- CORPUS (2002). « Corpus et recherches linguistiques », numéro 1, coordonné par Sylvie Mellet.
- CORPUS (2003). « La distance intertextuelle », numéro 2, coordonné par Xuan Luong.
- CORPUS (2007). « Interprétation, contextes, codage », numéro 5, coordonné par Bénédicte Pincemin.
- CORPUS (2009). « Corpus de textes, textes en corpus », numéro 8, coordonné par Jean-Michel Adam et Jean-Marie Viprey.

- Cotteret J.-M. (1991). *Gouverner, c'est paraître*. Paris : PUF.
- Courdesses L. (1971). Blum et Thorez en mai 1936. *Langue française*, 9, pp. 22-33.
- Dalbera J.-Ph. (1980). Esquisse d'une classification syntaxique des adverbes. *Travaux du cercle linguistique de Nice*, 2, pp. 39-60.
- Dalbera J.-Ph. (2002). Le corpus entre données, analyse et théorie. *Corpus*, 1, pp. 89-105.
- Debray R. (2007). *L'obscénité démocratique*. Paris : Flammarion.
- Détrie C., Siblot P., Verine B. (2001). *Termes et concepts pour l'analyse du discours. Une approche praxématique*. Paris : Champion.
- Ducrot O. (1966). De Saussure à la philosophie du langage ; préface à John R. Searle, *Les actes de langage. Essai de philosophie du langage*. Paris : Hermann, pp. 7-34.
- Ducrot O. (1980). *Le Dire et le Dit*. Paris : Minuit.
- Eco U. (1992). *Les limites de l'interprétation*. Paris : Grasset.
- Fabre C., Habert B. & Issac F. (1998). *De l'écrit au numérique : constituer, normaliser, exploiter les corpus électroniques*. Paris : InterEditions/Masson.
- Fairclough N. (1995). *Critical Discourse Analysis*. Boston : Addison Wesley.
- Fairclough N. (2001 2nd edition). *Language and Power*. London: Longman.
- Fairclough N. (2003). *Analysing Discourse: Textual Analysis for Social Research*. London : Routledge.
- Firth J. (1957). A Synopsis of Linguistic Theory 1930-1955. *Studies in Linguistic Analysis*, pp. 1-32.
- Flament C et Milland L. (2005). Un effet Guttman en ACP. *Mathemactis and Social Sciences*, 171, pp. 25-49.
- Florea M.-L. (2009). Tabularité : des textes aux corpus. *Corpus*, 8, pp. 177-196.
- Gaudin-Bordes L. et Salvain G. (dir.) (2008). *Les registres. Enjeux stylistiques et visées pragmatiques*. Louvain : Bruylant-Academia.
- Gerstlé J. (1979). *Le langage des socialistes*. Paris : Stanké.
- Guilhaumou J. (1993). A propos de l'analyse de discours : les historiens et le « tournant linguistique ». *Langage & Société*, 65, pp. 5-38.
- Guilhaumou J. (1986). L'historien du discours et la lexicométrie. *Histoire & Mesure*, vol I, 3/4, pp. 27-46.
- Guilhaumou J. (2002). Le corpus en analyse de discours. Perspective historique. *Corpus*, 1, 2002, pp. 21-50
- Guilhaumou J. (2004). Où va l'analyse de discours ? Autour de la notion de formation discursive. *Texto* ([www.revue-texto.net](http://www.revue-texto.net)).
- Guilhaumou J. (2006). *Discours et événement. L'histoire langagière des concepts*. Besançon : Presses Universitaires de Franche-Comté.
- Guilhaumou J. (2007) L'analyse du discours du côté de l'histoire. Une démarche interprétative. *Langage & Société*, 121-122, pp. 177-188.
- Guilhaumou J, Maldidier D. & Robin R. (1994). *Discours et archive. Expérimentations en analyse du discours*. Liège : Mardaga.
- Habert B. (2005). Portrait de linguiste(s) à l'instrument. *Texto* ([http://www.revue-texto.net/Corpus/Publications/Habert/Habert\\_Portrait.html](http://www.revue-texto.net/Corpus/Publications/Habert/Habert_Portrait.html)).
- Habert B. & Nazarenko A. & Salem A. (1997). *Les linguistiques de corpus*. Paris : Colin.
- Hébert L. (1996). Une sémantique différentielle unifiée. *RS/SI, Association canadienne de sémiotique*, vol. XVI, n°1-2, pp. 275-285.
- Hjelmslev L. ([1943] 1968-1971). *Prolégomènes à une théorie du langage*. Paris : Minuit.
- Halliday M.A.K. (1973). *Explorations in the functions of language*. London : Edward Arnold.

- Halliday M. A. K. & Hasan R. (1976). *Cohesion in English*. Londres : Longman.
- Harris Z. S. (1952.). Discourse Analysis. *Langue*, 28:1, pp. 1-30. Traduit en français dans *Langages* (1969), « L'Analyse du discours », n°13, pp. 8-45.
- Jaubert A. (1990). *La lecture pragmatique*. Paris : Hachette.
- Kastberg Sjöblom M. (2006). *L'écriture de J.M.G. Le Clézio. Des mots aux thèmes*. Paris : Champion.
- Kennedy G. (1998). *An introduction to corpus linguistics*. London & New York : Longman.
- Krieg A. (2009). *La notion de "formule" en analyse du discours : cadre théorique et méthodologique*. Besançon : Presses Universitaires de Franche-Comté.
- Labbé C et Labbé D. (2003). La distance intertextuelle. *Corpus*, 2, pp. 95-118.
- Labbé D. (1977). *Le discours communiste*. Paris : Presses de la Fondation des sciences politiques.
- Labbé D. (1983). *François Mitterrand : essai sur le discours*. Grenoble : La pensée sauvage.
- Labbé D. et Monière D. (2003). *Le discours gouvernemental. Canada, Québec, France (1945-2000)*. Paris : Champion.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève-Paris : Slatkine-Champion.
- LANGAGES (1995). « Les analyses du discours en France », n°117, coordonné par D. Maingueneau.
- LANGAGES (2006). « Unité(s) du texte », n°163, coordonné par D. Legallois.
- LANGAGES (2008). « Construction des faits en linguistique : la place des corpus », n°171, coordonné par M. Cori, S. David, J. Léon.
- Laks B. (2008). Pour une phonologie de corpus. *Journal of French Language Studies*, 18, pp. 3-32.
- Le Bart Chr. (2001). *Le discours politique*. Paris : PUF.
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Legallois D. (2006). Le texte et le problème de son et ses unités : proposition pour une déclinaison. *Langages*, 163, pp. 3-9.
- Léon J. (2008). Aux sources de la « Corpus Linguistics » : Firth et la London School. *Langages*, 171, pp. 12-34.
- Loiseau S. (2006). *Sémantique du discours philosophique chez Deleuze : du corpus aux normes*. Thèse de doctorat : Université Paris X Nanterre.
- Longrée D., Luong X. et Mellet S. (2004). Temps verbaux, axe syntagmatique, topologie textuelle : analyses d'un corpus lemmatisé, in G. Purnelle, C. Fairon, A. Dister (éds), *JADT04*. Louvain : Presses universitaires de Louvain, vol. 2, pp. 743-752.
- Longrée D., Luong X. et Mellet S. (2006). Distance intertextuelle et classement des textes d'après leur structure : méthodes de découpage et analyses arborées, in J.-M. Viprey (textes réunis par), *JADT'06*. Besançon : Presses universitaires de Franche-Comté, vol. 2, pp. 643-654.
- Luong X. (1988). *Méthodes d'analyse arborée. Algorithmes, applications*, Thèse d'Etat, 1988, Paris V.
- Madelin Ph. (2002). *Jacques Chirac. Une biographie*. Paris : Dunod.
- Maingueneau D. (1991). *L'analyse du discours : initiation aux lectures de l'archive*. Paris : Hachette.
- Maingueneau D. (1996). *Les termes clés de l'analyse du discours*. Paris : Seuil.
- Maingueneau D. (2002). Linguistique et littérature : le tournant discursif. *Vox Poetica* (<http://www.vox-poetica.org/t/maingueneau.html>).
- Maingueneau D. (2004). *Le discours littéraire. Paratopie et scène d'énonciation*. Paris : Armand Colin.
- Maingueneau
- Malrieu D. et Rastier F. (2001). Genres et variations morphosyntaxiques. *Traitement Automatique des langues*, vol. 42, n°2, pp. 548-577.
- Marchand P. (2007). *Le grand oral. Le discours de politique générale de la Ve République*. Bruxelles : De boeck.

- Mazière F. (2005). *L'analyse du discours. Histoire et pratiques*. Paris : Puf.
- McCombs M. & Shaw D. (1972). The agenda-setting function of mass-media, *Public Opinion Quarterly*, vol. 36, n°2, pp. 176-187.
- Mellet S. (1998). Présent et présentification : un problème d'aspect, in S. Vogeller, A. Borillo, C. Vetter & M. Vuillaume (éds.) *Temps et discours*. Louvain-la-Neuve : Peeters (BCILL 99), pp. 203-213.
- Mellet S. (2000). Chronique de linguistique générale et française. Le présent. *Travaux de linguistique*, 40, pp. 97-111.
- Mellet S. (2001). Valeur aspectuelle du présent : un problème de frontière. *Cahiers Chronos*, 7, « Le présent en français » (éd. : P. Le Goffic), Amsterdam – Atlanta : Rodopi, pp. 27-39.
- Mellet S. (2002). Corpus et recherches linguistiques : introduction. *Corpus*, 1, pp. 5-13.
- Mellet S. (2006). La valeur aspectuelle du présent, in J.F. Marillier, M. Dalmas & I. Behr (éds), *Text und Sinn* (Festschrift für Marcel Vuillaume). Stauffenburg Verlag, pp. 167-180.
- Mellet S. *et alii* (2009). Représentations du texte pour la classification arborée et l'analyse automatique de corpus. Application à un corpus d'historiens latins. *Mathématiques et Sciences humaines*, 187, pp. 107-121.
- Mellet S. et Longrée D. (2009). Syntactical Motifs and Textual Structures. Considerations based on the Study of a Latin historical Corpus, in S. Mellet et D. Longrée, *New approaches in text linguistics*. Amsterdam : John Benjamins, pp. 161-173.
- Mellet S et Purnelle G. (2002). Les atouts multiples de la lemmatisation : l'exemple du latin, in Morin A. et Sébillot P. (éd), *JADT 2002*. Rennes : IRISA, vol 2, pp. 529-538.
- Muller Ch. (1973). *Initiation aux méthodes de la statistique linguistique*. Paris : Hachette.
- Muller Ch. (1977). *Principes et méthodes de la statistique lexicale*. Paris : Hachette.
- Nølke H. (1990). Formes et emplois des énoncés négatifs : polyphonie et syntaxe de *ne... pas*. *Revue romane*, 25/2, pp. 1-17.
- Nølke H. (1992). *Ne... pas* : négation descriptive ou polémique. *Langue française*, 94, pp. 48-67.
- Nølke H. (1999). La polyphonie : analyses littéraire et linguistique. *Tribune*, 9, pp. 5-19.
- Partington A. (1998). *Patterns and Meanings : Using Corpora for English Language Research and Teaching*. Amsterdam : John Benjamin's.
- Partington A., Morley J. et Haarman L. (éds.) (2004). *Corpora and Discourse*. Proceedings of ComConf 2002 Università degli Studi di Camerino, Centro Linguistico d'Ateneo Sept 27th-29th 2002. Bern, Berlin, Bruxelles, Frankfurt /M., New York, Oxford, Wien : Peter Lang.
- Partington A. & Morley J. (2004). At the heart of ideology : Word and cluster / bundle frequency in political debate, in Lewandowska-Tomaszyk B. (dir.) *Practical Applications in Language and Computers (PALC 2003)*. Bern : Lang, pp. 179-192.
- Perrineau P. (éd.) (2008). *Le vote de rupture. Les élections présidentielle et législatives d'avril-juin 2007*. Paris : Presses de Sciences Po.
- Peschanski D. (1989). *Et pourtant ils tournent : vocabulaire et stratégie du PCF (1934-1936)*. Paris : Klincksiek.
- Peschanski D. (sous la direction de...) (1984). *Histoire moderne et contemporaine et informatique*, n°4 spécial.
- Pincemin B. (2004). Lexicométrie sur corpus étiquetés, in G. Purnelle, C. Fairon, A. Dister (éds.), *JADT04*. Louvain : Presses universitaires de Louvain, vol. 2, pp. 865-873. [En ligne sur le site Lexicometrica : [http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT\\_084.pdf](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_084.pdf)].
- Pincemin B. (2008). Modélisation textométrique des textes, in S. Heiden & B. Pincemin (éds.) *Actes des 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*. Lyon : Presses Universitaires de Lyon, vol. II, pp. 949-960.
- Pincemin B. (2010 –sous presse). Sémantique interprétative et textométrie. *Topics del Seminario*, 23, (numéro coordonné par Carine Duteil-Mougel et Viviana Cardenas).

- Poudat C. (2006). *Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres*. Thèse de doctorat : Université d'Orléans.
- Poudat C. (2010- à paraître). *L'article scientifique de linguiste*. Paris : Champion
- Prost A. (1974). *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*. Paris : PUF.
- Prost A. (1988). "Les mots", in R. Rémond (dir.), *Pour une histoire politique*. Paris : Seuil, pp. 255-286.
- Rastier F. (1998). Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage. *Langages*, 129, pp. 97-111.
- Rastier F. (2001). *Arts et sciences du texte*. Paris : Puf.
- Rastier F. (2005-a). Enjeux épistémologiques de la linguistique de corpus, in G. Williams (éd.), *La linguistique de corpus*. Rennes : Pur, pp. 31-45. [ En ligne sur *Texto !* ([http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html))]
- Rastier F. (2005-b). Discours et texte (première partie). *Texto !* ([http://www.revue-texto.net/Reperes/Themes/Rastier\\_Discours.html](http://www.revue-texto.net/Reperes/Themes/Rastier_Discours.html)).
- Rastier F. (2007). Passages. *Corpus*, 6, pp. 25-54.
- Reboul A. et Moeschler J. (1998). *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*. Paris : Colin.
- Robin R. (1973). *Histoire et Linguistique*. Paris : Colin.
- Salem A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
- Salem A. (1988). Approches du temps lexical. Statistique textuelle et séries chronologiques. *Mots*, 17, pp. 105-143.
- Salem A. (1991). Les séries textuelles chronologiques, *Histoire et Mesure*, VI-1/2, pp. 149-175.
- Salem A. (1993). *Méthodes de la statistique textuelle*, Thèse pour le doctorat d'État ès lettres et sciences humaines, Université de la Sorbonne nouvelle - Paris 3, mars 1993, 3 vol, 998 p.
- Salem A. (1997). Chap. IX-7) Temps lexical, in B. Habert, A. Nazarenko et A. Salem, *Les linguistiques de corpus*. Paris : Colin, pp. 207-211.
- Salem A. (2004). Introduction à la résonance textuelle, in G. Purnelle, C. Fairon, A. Dister (éds), *JADT04*. Louvain : Presses universitaires de Louvain, vol. 2, 986-992.
- Salem A. (2006). Proximité segmentale in J.-M. Viprey (textes réunis par), *JADT' 06*. Besançon : Presses universitaires de Franche-Comté, vol. 2, pp. 843-853.
- Sarfati G.-E. (2003). L'analyse du discours et la question du texte : l'horizon du sens commun et de la doxa, in Amossy R. et Maingueneau D., *L'Analyse du discours dans les études littéraires*. Toulouse : Cerisy-Presses universitaires du Mirail, pp. 429-438.
- Sassier M. (2008). Genre, registre, formation discursive et corpus. *Langage et société*, 124, pp. 39-57.
- Scheer T. (2004-a). Présentation du volume. En quoi la phonologie est vraiment différente. *Corpus*, 3, pp. 5-85.
- Scheer T. (2004-b). Le corpus heuristique : un outil qui montre mais ne démontre pas. *Corpus*, 3, pp. 153-193.
- Searle J. R. (1966). *Les actes de langages. Essai de philosophie du langage*. Paris : Hermann.
- SEMEN (2006). « Catégories pour l'analyse du discours politique », n°21, coordonné par Ph. Schepens.
- SEMEN (2009). « Critical Discourse Analysis I. Les notions de contexte et d'acteurs sociaux », n°27, coordonné par Ph. Schepens et A. Petitclerc.
- Sinclair J. McH. (1991). *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- Teubert W. (2009). La linguistique de corpus : une alternative. *Semen*, 27, pp. 185-211.
- Tognini-Bonelli E. (2001). *Corpus Linguistics at Work*. Amsterdam : John Benjamin's Publishing.
- Tournier M. (1980). En souvenir de Lagado. *Mots*, 1, pp. 5-9.

- Tournier M. (1985). Sur quoi pouvons-nous compter ? Réponse à Charles Muller. *Verbum* (numéro spécial). Nancy : Presses universitaires de Nancy.
- Tournier M. (1986). La lexicométrie socio-politique. *Le courrier du CNRS*, 65, pp. 185-192.
- Tournier M. (1987). La réduction : principe de lexicométrie politique, *brochure de URL "Lexicométrie et textes politiques"*, 14 pages.
- Valette M. (2008). A quoi servent les lexiques sémantiques généralistes ? Discussion et propositions. *Cahiers du Cental*, 5, pp. 43-58.
- Valette M. (2009). *Approche textuelle du lexique*. Mémoire d'Habilitation à diriger des recherches (à paraître).
- Van Dijk T. (1984). "Texte",. in Beaumarchais et al (éds.), *Dictionnaire des littératures de langue française*. Paris : Bordas.
- Van Dijk T. (2009). Texte, Contexte et Connaissance. *Semen*, 27, pp. 127-155.
- Vion R. (2005). Modalités, modalisations, interaction et dialogisme, in J. Bres, P. Haillet, S. Mellet, H. Nølke, L. Rosier (éds), *Dialogisme, polyphonie : approches linguistiques*. De Boeck Duculot, pp. 143-156.
- Vion R. (2006). Modalisation, dialogisme et polyphonie. *Recherches linguistiques*, 28 « Le sens et ses voix. Dialogisme et polyphonie en langue et en discours », numéro spécial sous la direction de L. Perrin, pp. 105-125.
- Vion R. (2007). Dimensions énonciative, discursive et dialogique de la modalisation. *Linguas & Letras*, vol 8, n°15, pp. 193-224.
- Viprey J.-M. (1997). *Dynamique du vocabulaire des Fleurs du mal*. Paris : Honoré Champion.
- Viprey J.-M. (2005-a). Philologie numérique et herméneutique intégrative, in Adam J.-M. et Heidmann U. (éds.), *Sciences du texte et analyse de discours*. Genève : Slatkine, pp. 51-68.
- Viprey J.-M. (2005-b). Corpus et sémantique discursive : éléments de méthode pour la lecture des corpus, in A. Condamines (dir.), *Sémantique et corpus*. Paris : Lavoisier, pp. 245-276.
- Viprey J.-M. (2006). Structure non-séquentielle des textes. *Langages*, 163, pp. 71-85.
- Williams G. (éd.) (2005). *La linguistique de corpus*. Rennes : PUR.
- Wilmet M. (1998). *Grammaire critique du français*. Paris : Hachette.
- Wodak R. (2007). Pragmatics and Critical Discourse Analysis : A cross-discipline Inquiry. *Pragmatic & Cognition*, 15,1, pp. 203-225.
- Wodak R. (2009). *The Discourse of Politics in Action: Politics as Usual*. Basingstoke : Palgrave.
- Wodak R. and Meyer M. (2002). *Methods of Critical Discourse Analysis*. London : Sage Publications.

<b>INTRODUCTION</b>	<b>5</b>
<b>1. UN OBJET : LES CORPUS TEXTUELS NUMERIQUES</b>	<b>10</b>
1.1. Le corpus comme matrice	11
1.2. Le corpus comme contexte	12
1.3. Corpus réflexifs et herméneutique endogène	13
1.4. Le corpus comme texte	15
1.5. Corpus et numérique	19
Conclusion	20
<b>2. UNE METHODE : LA LOGOMETRIE</b>	<b>22</b>
2.1. Des lexies (lexi*) au discours (logo*) : retour sur les unités textuelles	23
Matérialisme textuel et primauté de la forme graphique	24
L'avantage de l'étiquette morpho-syntaxique	27
Croiser les différents niveaux	30
2.2. Quantitatif et qualitatif , macro et micro, global et local	34
2.3. Administration de la preuve et contrôle de l'interprétation	37
Conclusion	44
<b>3. UNE FIN : LE LANGAGE POLITIQUE CONTEMPORAIN</b>	<b>46</b>
3.1. Individuation politique et linguistique ?	48
3.1.1. Thorez <i>versus</i> Blum ou la question des genres de discours	48
3.1.2. Chirac, entre de Gaulle et Mitterrand	55
3.1.3. Sarkozy <i>versus</i> Royal	58
3.2. Temps discursif et temps politique	63
3.2.1. Thorez (1930-1939) : continuum lexical et hiatus chronologique (I)	63
3.2.2. Jospin <i>versus</i> Chirac : continuum discursif et hiatus chronologique (II)	66
3.2.3. De de Gaulle à Chirac : dichotomie discursive	69
Conclusion	73
3.3. Du discours politique à la langue	74
3.3.1. L'adverbialisation du discours contemporain	75
3.3.2. Temps verbaux – pragmatique du présent dans le discours de Chirac	84
<b>CONCLUSION GENERALE</b>	<b>95</b>
<b>BIBLIOGRAPHIE</b>	<b>101</b>

