



**HAL**  
open science

## Analyse du discours politique et Logométrie : point de vue pratique et théorique

Damon Mayaffre

► **To cite this version:**

Damon Mayaffre. Analyse du discours politique et Logométrie : point de vue pratique et théorique. Langage et Société, 2005, 114, pp.91-121. hal-00551483

**HAL Id: hal-00551483**

**<https://hal.science/hal-00551483>**

Submitted on 3 Jan 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.

Fac-similé



# langage & société

REVUE TRIMESTRIELLE

n° **114**

DÉCEMBRE 2005

**APPROCHES INTERDISCIPLINAIRES  
DES PRATIQUES LANGAGIÈRES ET DISCURSIVES**

**Présentation**

*Josiane Boutet et Dominique Maingueneau*

**Sociolinguistique et analyse de discours :  
façons de dire, façons de faire**

*Josiane Boutet et Dominique Maingueneau*

**Sociolinguistique et analyse du discours : quel type d'interface ?**

*Norbert Dittmar*

**Anthropologie sociale et analyse du discours**

*Bertrand Masquelier*

**Analyse du discours politique et logométrie**

*Damon Mayaffre*

**L'analyse du discours institutionnel entre formations  
discursives et problématiques socio-anthropologiques**

*Claire Oger*

**Des villes, des mots, des discours**

*François Leimdorfer*

MAISON DES SCIENCES DE L'HOMME PARIS

## Analyse du discours politique et logométrie : point de vue pratique et théorique

**Damon Mayaffre**  
CNRS-UNSA  
mayaffre@unice.fr

Assigner à l'Analyse du discours de penser l'articulation entre texte et contexte, c'est-à-dire plus largement entre langage et société, c'est pointer – en espérant à terme le réduire – le grincement pluridécennal entre les différentes disciplines académiques qui ont affaire au discours (Linguistique / Histoire, Sociologie, Psychologie, etc.).

Ériger, aujourd'hui, l'Analyse du discours en discipline autonome, comme le fait la littérature récente sur le sujet, telle l'introduction du *Dictionnaire d'analyse du discours*<sup>1</sup>, a évidemment l'avantage de soulager, par exemple, la tension jamais retombée entre Histoire et Linguistique, par un déplacement (ou recentrage) épistémologique séduisant.

Tout au crédit des penseurs de l'AD, ce déplacement est à la fois fertile d'un point de vue pratique et problématique d'un point de vue théorique. La réalité de nos pratiques quotidiennes d'analyste du discours politique ne saurait s'élever au rang de théorie. Elle repose néanmoins sur deux constats élémentaires, donc irréductibles, déjà évoqués dans notre thèse (Mayaffre, 2000) puis réaffirmés dans un ouvrage récent sur le discours présidentiel sous la V<sup>e</sup> République

---

1. P. Charaudeau & D. Maingueneau (dir.) (2002), voir l'avant propos (pp. 7-12) et l'entrée « Analyse du discours » (pp. 41-45).

(Mayaffre, 2004) : le matériau d'étude est linguistique – son traitement ne saurait faire l'économie d'une approche linguistique –, sa compréhension est, au sens large, politique – son interprétation ne saurait faire l'économie de compétences interprétatives sociales ou historiques.

Toute analyse qui se soustrait à ces deux réalités risque le discrédit soit du côté de la Linguistique soit du côté de l'Histoire ; parfois des deux. De fait, certaines analyses du discours semblent oublier la matérialité linguistique de leur objet d'étude (le discours ou, comme nous le préciserons, le corpus textuel) pour s'égarer vers l'appréhension d'autres signes (gestométrie, sémiométrie, analyse scénique des médias ou de la communication) et lorsque le texte n'est pas oublié, il continue chez beaucoup d'historiens à faire l'objet d'approches intuitives du type commentaire de texte littéraire renouant au mieux avec l'herméneutique pré-saussurienne, au pire avec l'inconséquence journalistique.

À l'inverse, d'autres, notamment chez les linguistes ou les littéraires fascinés par le discours politique, s'autorisent, forts de leurs compétences textuelles, des conclusions socio-historiques sans un détour contrôlé par l'Histoire, oubliant que le sens n'est qu'interprétation et que celui d'un discours, inscrit dans la matérialité du texte, ne peut se saisir qu'au regard des conditions historiques de production.

C'est dans ce cadre-là que la lexicométrie de la deuxième génération – que nous appellerons logométrie car elle ne se contente plus de repérer et compter les seules lexies mais étend ses procédures aux lemmes, aux codes grammaticaux, aux enchaînements syntaxiques et finalement à toutes les unités essentielles du discours – nous apparaît comme le plus rigoureux moyen pour mettre en place une linguistique de corpus qui ne s'éloigne jamais des unités linguistiques effectivement émises, tout en ouvrant les portes, grâce à son pouvoir descriptif et sa valeur heuristique, à l'interprétation extra-linguistique<sup>2</sup>.

---

2. L'autre vertu de la logométrie, que nous ne développerons pas ici, est d'articuler le niveau micro ou local (grâce à la lecture qualitative systématique des plus petites unités de signification du discours et le renvoi hypertextuel précis du lecteur aux contextes paragraphiques, phrastiques, syntagmatiques) et le niveau macro ou global grâce au traitement quantitatif exhaustif du corpus et l'analyse des affinités statistiques de tel mot avec tel autre, etc.

### I. Entre Histoire et Linguistique : les limites de l'analyse du discours comme *carrefour*

Il faut bien le constater, la définition modeste de l'Analyse du discours comme *carrefour* entre les sciences humaines ou sociales<sup>3</sup> a provoqué plus de rendez-vous manqués que de rencontres réussies : pour cette raison sans doute doit-elle être dépassée par une prétention académique plus affirmée.

Les rapports emblématiques entre Histoire et Linguistique, dont l'AD devait être le lieu, ont abouti en effet à un dialogue de sourds tenace. Dès l'origine sur la question, R. Robin (1973) intitulait le chapitre introductif de son livre fondateur « *Le malentendu* »<sup>4</sup>. Plus grave, treize années après, l'auteure se désespérait dans *Langages* (n° 81, 1986) : « Postface. L'Analyse de discours entre la linguistique et les sciences humaines : *l'éternel malentendu* ». À terme, l'exil de Robin au Canada et le détournement de sa production scientifique vers d'autres terrains semblent symboliser l'impossible dialogue interdisciplinaire, moins que jamais aujourd'hui abouti<sup>5</sup>. Certes, en trois décennies, des progrès ont été enregistrés et certaines études sociolinguistiques témoignent du fructueux échange pluridisciplinaire pour attester de la perméabilité – sans reniement – des différentes disciplines. Les travaux de Saint-Cloud, la thèse prometteuse, mais sans suite, de D. Peschanski, les travaux de J. Guilhaumou ou de D. Labbé, aujourd'hui nos propres tentatives, prouvent que tous les essais n'ont pas été vains.

De manière encourageante, en 1978, J. Le Goff demandait à un historien et à une linguiste de rédiger, à deux mains, l'entrée *Langage* dans sa somme historiographique<sup>6</sup>. Plus significatif encore, en 1984, la fine fleur des historiens en histoire politique contemporaine écrit sous la direction de R. Rémond un manifeste méthodologique. Et l'on

3. Le terme de « carrefour » est en effet celui qui fut le plus commodément employé entre 1970 et 2000 pour qualifier l'AD. Par exemple, dès le départ : (Maingueneau 1976 : 183).

4. R. Robin (1973) première partie : *Le problème des rapports Histoire/Linguistique*, chapitre 1 : *Le malentendu* : pp. 13-34.

5. D'autres accidents biographiques sont aussi à mentionner dans l'histoire de l'AD en France, comme les décès prématurés de M. Pécheux et D. Maldidier.

6. J. Guilhaumou & D. Maldidier (1978) entrée « Langage » in J. Le Goff, *et alii* (dir.), *La Nouvelle histoire*, Paris, pp. 304-308.

confia au meilleur d'entre eux un chapitre sur *Les mots* (Prost 1988). Certes, A. Prost s'y montre assez prudent – le titre « *Les mots* » est en lui-même révélateur de cette prudence –, mais l'article signifie que la Linguistique ne peut plus être ignorée des historiens qui traitent du texte politique.

Cependant, avouons-le, 30 ans après le livre de R. Robin, le bilan est maigre, et tend, semble-t-il aujourd'hui, à s'amaigrir. Sans doute des échanges autour ou à propos du discours existent-ils entre les disciplines, mais uniquement par capillarité. Les productions sont le fait de contributions individuelles, mais aucune école ne s'est imposée : la disparition *ab intestat* du laboratoire de Saint-Cloud en est le plus triste signe. Dit autrement, du point de vue des historiens ou des sociologues, aucune procédure linguistique ne s'est normalisée. Le détour par la Linguistique est une option parfois – rarement – empruntée, non une voie nécessaire ou naturelle pour celui qui prétend étudier en Sciences humaines un corpus textuel.

La récente livraison d'*Histoire et Mesure* (vol. XVIII, 2003) intitulée « *Mesurer le texte* » est significative dans le cadre particulier mais fondateur de la lexicométrie. Les coordinateurs du numéro titrent prudemment leur introduction : *Des chiffres et des lettres : quelques pistes pour l'historien*. Du chemin pionnier défriché par Robin et labouré pendant deux ou trois décennies par Saint-Cloud, il ne reste que « quelques pistes » à redécouvrir. Retour à la case départ donc.

Et l'ambiance épistémologique actuelle ne dispose pas à l'optimisme. Le recroquevillement monodisciplinaire dans les sciences historiques d'une part – recroquevillement lointainement issu du rejet réactionnaire du marxisme comme grille de compréhension transversale –, la perte du prestige de la Linguistique (structurale) comme science et discipline fédérative d'autre part, ne laissent rien envisager de très prometteur. En ce sens, l'Analyse du discours ne doit sans doute plus être seulement un *carrefour*, une *pratique interprétative* ou un *lieu problématique* pour des chercheurs qui viendraient – mais ne viennent plus – d'ailleurs, mais doit trouver un moyen d'exister indépendamment des grandes disciplines académiques de nouveau repliées sur elles-mêmes.

## 2. Les prétentions de l'Analyse du discours comme discipline

À reprendre la production de D. Maingueneau, la constitution de l'Analyse du discours en discipline est un vœu complexe et lointain, d'abord clairement dénié, ensuite inavoué dans l'esprit même de l'auteur, pour n'être affirmé que tardivement. Si la tentative d'émancipation de l'AD a pris aujourd'hui un corps que l'on ne peut ignorer c'est qu'elle apparaît comme le fruit mature et problématique de 30 ans de réflexion.

Dès les premières lignes de *l'Initiation aux méthodes de l'analyse du discours*, en 1976, Maingueneau qualifie l'analyse du discours de « discipline » (p. 3). Celle-ci est certes « tâtonnante » (*ibid.*) et apparaît seulement comme une sous-discipline dégradée de la Linguistique. Mais à bien lire, si elle apparaît comme sous-discipline de la Linguistique, c'est aussi pour être immédiatement pressentie comme son « double » (*ibid.*). C'est ce dédoublement benvenistien, originellement entrevu donc, d'une linguistique de la Langue par une linguistique du Discours qui semble s'être progressivement imposé à l'esprit de l'auteur.

Dans le cheminement intellectuel de Maingueneau, le deuxième jet, en 1987, comporte à la fois un recul et une avancée. Le titre de l'ouvrage est symptomatique du recul : *Nouvelles tendances en analyse du discours*, a été préféré à *Nouvelles tendances de l'analyse du discours* : l'AD apparaît ainsi uniquement comme une pratique – dont on devine l'aspect pluriel et hétérogène – et non comme discipline constituée. Pourtant, ce retrait évident peut être interprété avant tout comme une sage stratégie et un nécessaire principe de précaution devant la floraison anarchique des analyses portant sur les discours et que l'auteur dénonce ; car l'introduction comme les trois parties du livre montrent la nécessité de définir un cadre théorique, de stabiliser les pratiques, d'émanciper la discipline des autres : en un mot, certes non explicité par l'auteur, de se constituer en champ autonome.

C'est ce qui pourrait ressortir en 1991 du titre décomplexé *l'Analyse du Discours*. Sauf que la livraison de 1991 se profile comme le symétrique de 1987. Le titre est hardi mais le propos reste prudent. Maingueneau n'est toujours pas prêt à faire le pas de l'indépendance. D'autant que sa revendication de l'héritage de l'École

française d'analyse du discours semble l'empêcher d'aller plus loin<sup>7</sup>. Et la livraison de 1995 (D. Maingueneau (éd.), 1995) qui s'inscrit précisément toujours dans l'affiliation à l'École française ne change pas fondamentalement la donne.

Bref, il faudra attendre le tournant des années 2000 pour voir la rupture dans la posture de Maingueneau : c'est l'entreprise même du *Dictionnaire d'analyse du discours*, qui ne cache plus ni dans sa quatrième de couverture, ni dans son introduction, ni à l'entrée *Analyse du discours* son ambition de constituer l'AD, qualifiée plusieurs fois de « *discipline émergente* », en véritable institution scientifique et académique dont les contours pour n'être pas encore totalement définis s'annoncent très ambitieux<sup>8</sup>.

### 3. Quel objet ? Quelle méthode ? Le problème reste entier

L'ambiguïté de l'AD comme discipline vient de l'objet même qu'elle s'est donné et des méthodes qu'elle se refuse.

Le discours n'a jamais trouvé une définition stable ; cinq, six, dix définitions en ont été données<sup>9</sup>. Ou plutôt, la définition minimale du discours qui a fini par s'imposer<sup>10</sup> ne lui donne pas de stabilité disciplinaire performante ni au regard de la Linguistique ni de l'Histoire : d'où la tentation de faire une nouvelle discipline *ad hoc* pour traiter cet objet sans forme.

7. Nous ne développerons pas ici ce constat essentiel. Mais il est évident que l'autonomisation de l'AD, autour des années 2000, coïncide avec, sinon le rejet, en tout cas la mise à distance de l'héritage et du « geste inaugural » de l'École française d'AD. Voir à ce sujet, nécessairement, le point de vue critique de (J. Guilhaumou 2004).

8. Charaudeau & Maingueneau (dir.) (2002) : « L'existence même d'une discipline comme l'analyse du discours constitue un phénomène qui n'est pas anodin : pour la première fois dans l'histoire, la totalité des énoncés d'une société, appréhendée dans la multiplicité de leurs genres, est appelée à devenir objet d'étude » (p. 45). Ou « cette entreprise va au-delà de la simple mise au point d'un ouvrage utile, ... elle consacre l'émergence d'un nouveau champ de savoir, l'aboutissement de plus de quatre décennies d'efforts longtemps demeurés dans l'obscurité pour faire prévaloir des démarches que souvent les tenants des disciplines plus anciennes considéraient marginales ou superflues (...) Une telle entreprise en est encore à ses débuts, mais, pour la première fois dans l'histoire, c'est la totalité des productions verbales, dans leur multiplicité, qui peut devenir objet d'étude » (p. 12).

9. Par exemple (Maingueneau 1976 et 1991, 11-12 et 15).

10. Discours = texte + conditions de production. Définition qui s'est imposée progressivement dans les années 1970. Elle est rappelée encore récemment, pour être critiquée, par A. Jaubert, « Corpus et champs disciplinaire », *Corpus*, n° 1, 2002, pp. 77-78.



Dès l'origine en effet, on se rappelle le laborieux échange, en trois temps, entre Guespin et Robin<sup>11</sup>. L'un et l'autre faisant semblant de s'entendre sur le terme *discours*, au moment même où un fossé les séparait. On crut un moment réduire le fossé, et dépasser les différends, en parlant moins de *discours* que d'*archive* aussi bien du côté des linguistes<sup>12</sup> que des historiens-linguistes<sup>13</sup>. Le malentendu ne fit alors que s'aggraver. Le terme même d'*archive* – fût-il emprunté à Foucault et non à Hérodote – arrivait sur le terrain conceptuel – en son cœur – de l'Histoire : l'AD, devenue analyse de l'archive, prétendait-elle subsumer l'Histoire ? Sur ce soupçon, elle ne pouvait être que rejetée.

Reste – et la précision est importante – que pendant longtemps l'objet de l'Analyse du discours politique fut moins le discours que la formation discursive<sup>14</sup> ; à défaut de s'entendre sur la nature historique ou linguistique du discours politique, les chercheurs s'entendaient sur sa dimension idéologique. Cette approche idéologique des discours, partagée par la plupart des chercheurs influencés alors par le « *freudo-marxisme althussérien qui a explicitement présidé à l'Analyse du discours* » (Rastier 2001 : 245) en France, a aujourd'hui disparu, et il ne reste plus rien pour unifier ou fédérer des approches ainsi menacées par des forces centrifuges. La notion de genre théorisée par Rastier (2001, chap. VIII, pp. 227-275) ou par Adam (1999) et admise par Maingueneau (dès 1987 : pp. 23-26) semble destinée à remplacer la métacompréhension – jusqu'ici idéologique donc – du discours, mais à condition de recentrer l'attention sur le fait linguistique au moment où précisément l'on s'en éloigne : car – d'effet imprévu en

11. Guespin (1971), « Problématique des travaux sur le discours politique », *Langages*, n° 23, p. 10 / (Robin 1973 : 21) / Guespin (1976), *Langages*, n° 41, pp. 4-5.

12. Explicitement : Maingueneau (1991) : *L'analyse du Discours. Introduction aux lectures de l'archive*.

13. Explicitement : Guilhaumou & Maldidier (1986), « Effets de l'archive : l'analyse de discours du côté de l'histoire », *Langages*, n° 81, pp. 43-57; Guilhaumou, Maldidier, Robin (1994) : *Discours et archive. Expérimentations en analyse de discours*.

14. Voir au départ : C. Haroche *et alii* (1971) « Le sémantique et la coupure saussurienne : langue, langage, discours », *Langages*, n° 24, pp. 93-106. Cf. aujourd'hui un renouvellement de la question : (Guilhaumou 2004) ; D. Mayaffre (2004) « Formation(s) discursive(s) et discours politique : l'exemplarité des discours communistes *versus* bourgeois dans l'entre-deux-guerres », *Texte !* ; (Rastier 2004).

Fac-similé

effet imprévu, par une forme de ruse de l'histoire – l'abandon d'un concept comme « formation discursive » participe à l'abandon de l'arrière-plan marxiste de l'AD, dont le principal avantage, pour le propos qui nous intéresse, était d'insister sur la matérialité de la langue ou plus étroitement du texte à traiter, lorsque l'étude des genres ouvre la voie, peut-être malgré elle, à des analyses plus littéraires que linguistiques dans lesquelles les figures discursives secondaires prennent le pas sur les faits de langue essentiels.

Le problème de la méthode est plus grave peut-être. Ce qui frappe en AD aujourd'hui, c'est le flou ou la diversité des méthodes utilisées, lorsqu'une discipline réclame une règle de conduite commune aux membres d'un corps.

Pourtant, originellement *discourse analysis* est né explicitement d'une méthode : la méthode distributionnelle (Harris, 1952, traduction 1969). Malgré les travaux initiaux de (Maldidier, 1971) ou (Pêcheux, 1969) et encore aujourd'hui de (Bourigault, 2002), celle-ci fut finalement peu pratiquée en France, mais l'AD politique reposa dans l'hexagone sur une méthodologie bien trempée : l'approche informatisée du texte et la statistique lexicale. C'est la lexicométrie en effet qui a donné, sur la durée, une stabilité, une trame de fond, une efficacité, une crédibilité à l'AD. Ainsi retrouvons-nous la lexicométrie en bonne place dès Robin, puis Maingueneau ou Prost<sup>15</sup>. La lexicométrie fut évidemment centrale durant 30 ans à Saint-Cloud. Aujourd'hui encore lorsque la Linguistique de corpus essaye de s'affirmer, ses théoriciens accordent à la lexicométrie une place de choix (Habert *et alii* 1997 : 183-211) et ses praticiens la célèbrent tous les deux ans au cours de Journées internationales d'Analyse de données textuelles (*JADT*, 2004, 2002, 2000, 1998...).

Pourtant l'AD nouvelle formule n'a plus la lexicométrie comme point d'ancrage méthodologique, ni l'ordinateur comme outil principal de traitement. Et à toucher à toutes les méthodes, l'AD semble ne plus en avoir aucune.

De manière générale, en l'absence de protocole méthodologique normalisé, les analyses de discours risquent de devenir des analyses

---

15. Respectivement : (Robin 1973 : 124-139) ; (Maingueneau 1976 : 22-45 et 1991 : 48-70) ; (Prost 1988 : 255-275).

intuitives, flottantes, sauvages, au mieux philologiques, qui empruntent indifféremment à la littérature, à la communication, à la sémiologie, etc. Même les travaux les plus remarquables n'échappent pas à un déficit de méthode linguistique, entendons notamment une méthode qui s'arrêterait d'abord aux faits linguistiques avant de s'attacher aux figures de style. Le retour à la rhétorique et à l'argumentation par exemple, très en vogue ces dernières années (Amossy, 2000 ; Buffon, 2002 ; Bonnaïfous *et alii*, 2003, Colloque d'Albi, 2004), serait profitable s'il n'était trop souvent synonyme d'un retour à une approche purement littéraire du texte à peine saupoudrée d'analyse énonciative ou lexicale, plus guère d'approche distributionnelle, rarement de logométrie, exceptionnellement de sémantique textuelle.

Instabilité de l'objet, évanescence de la méthode : c'est sans doute conscients de ces deux objections que les architectes du *Dictionnaire d'analyse du discours* ont entrepris leur œuvre. À défaut d'objet clairement défini et de protocole méthodologique linguistique sûr, au moins peut-on s'entendre sur un corps de concepts opérants ; et même si l'on doit remarquer, comme l'atteste la déclinaison en plusieurs sens différents de chacune des entrées du dictionnaire, que ces concepts sont hétérogènes pour la pragmatique, la praxématique, la stylistique, l'argumentation... gageons qu'il s'agit d'un point de départ nécessaire et d'une fondation d'avenir.

En ce qui nous concerne, la réponse aux objections contre une AD centralement au milieu de nulle part, « *qui tout à la fois dilue les notions de la linguistique et limite l'appréhension de l'historicité des textes* » (Guilhaumou 2004), prend une autre forme, moins ambitieuse, mais plus concrète : le traitement logométrique de grands corpus textuels.

#### 4. La logométrie comme garde-fou

La rencontre des chercheurs avec la logométrie ne bénéficie pas aujourd'hui de l'effervescence initiale des années 1960-1970. Elle est le plus souvent une rencontre avec des outils déjà aboutis tels les logiciels *Hyperbase* ou *Lexico*<sup>16</sup>. Dans ces conditions, il convient de rappeler avec

---

16. *Hyperbase* est conçu par É. Brunet et produit par l'UMR 6039, Bases, Corpus et Langage (Nice) ; *Lexico 3* est conçu par A. Salem et produit par SYLED-CLA2T (Paris 3).

Bachelard que les instruments sont toujours (et seulement) des « théories matérialisées »<sup>17</sup> et que derrière un logiciel comme *Hyperbase* se cachent des options épistémologiques longuement établies dont il n'est finalement que la mise en forme. En d'autres termes, la logométrie, par ses exigences propres, prévient un certain nombre de critiques évoquées plus haut quant à l'objet d'étude et quant à la méthode d'analyse.

### *Le corpus comme objet*

Le traitement informatisé implique d'abord la définition stricte et impérieuse de l'objet étudié : le corpus textuel. Non seulement l'ordinateur exige la saisie matérielle des données qui constituent ainsi un corps d'étude bien délimité, mais la statistique linguistique fonctionne nécessairement sur un corpus clos selon les impératifs d'une norme dite « endogène »<sup>18</sup>. La linguistique *de* corpus, la linguistique *sur* corpus n'est donc pas seulement une option ou un luxe, elle devient nécessité absolue ; elle est la condition technique du traitement. Aussi notre objet est-il moins le « discours » avec ses contours indéfinis que le « corpus textuel » dans l'évidence de son existence physique.

Concomitamment, la logométrie suppose que les unités textuelles à traiter soient spécifiées ou formalisées pour devenir, *via* l'ordinateur, les seules entrées du chercheur dans l'analyse. En logométrie tout nous ramène avec l'entêtement du système binaire (oui/non) aux unités linguistiques que la machine est invitée à rechercher ; dans un premier temps rien d'autre n'entre en ligne de compte.

On le sait, dans cette logique de corps à corps exclusif avec le texte, le traitement des formes graphiques a été longtemps privilégié. S'il y avait une certaine hypocrisie à vouloir faire absolument de nécessité une loi – pourquoi ne pas reconnaître que l'absence de lemmatiseur performant empêchait d'aller plus loin ? –, retenons tout de même cette exigence initiale de s'en tenir à la matérialité graphique du texte.

17. G. Bachelard, *Le nouvel esprit scientifique*, 1934, Première partie, introduction.

18. Impossible de revenir ici sur le b-a-ba de la lexicométrie : il n'existe pas de fréquence en Langue ; la fréquence d'un mot est toujours relative à un corpus ; ce corpus, composé nécessairement de sous-corpus, constitue donc la norme par rapport à laquelle les sous-fréquences constatées dans les sous-corpus prennent du sens.

Fac-similé

Car c'est cette exigence absolue de considérer exclusivement la matérialité du texte qui est la théorie sous-jacente de la lexicométrie. C'est elle qui aujourd'hui encore garantit contre les faux pas des analyses qui s'évadant du matériau textuel cessent rapidement d'être linguistiques.

Dans cette perspective, notons seulement qu'aujourd'hui *Hyperbase* s'articule sans difficulté avec le lemmatiseur *Cordial* pour traiter non seulement les formes graphiques – que l'on ne perd jamais de vue – mais les lemmes, les codes grammaticaux et les enchaînements syntaxiques (cf. des exemples *infra.*). On assiste ainsi depuis quelques années à une linguistisation de la pratique statistique d'origine dont le caractère fruste put un moment rebuter les linguistes. Et passons ainsi de la lexicométrie à la logométrie (Mayaffre, 2005).

Dans tous les cas, le corpus textuel avec toutes ses unités linguistiques (de la lettre aux combinaisons syntaxiques en passant par le lemme) et toutes ses dimensions (du caractère au texte en passant par la phrase et le paragraphe) n'est pas seulement central mais unique au moment du traitement : comme rappelé plus bas, l'Histoire intervient, en deçà, au moment de la constitution du corpus puis, au-delà, au moment de l'interprétation ou de l'inférence politiques, mais pas au moment du traitement scientifique de l'objet textuel.

De plus, ajoutons, si nécessaire, que ce traitement est selon un principe théorisé dès le départ, *systématique* et *exhaustif*<sup>19</sup> : c'est ce qui change beaucoup de choses par rapport à une lecture oculaire aléatoire et discrétionnaire. Ainsi le corpus ne saurait être *l'alibi* d'une analyse dans lequel le chercheur s'en irait trouver ici, puiser là, un exemple adéquat. Il est l'alpha et l'oméga, jusqu'à son épuisement. C'est évidemment dans cette volonté d'épuisement du corpus – surtout lorsque celui-ci est grand – que l'ordinateur et l'approche quantitative deviennent indispensables.

Enfin, il ne faut pas éluder l'articulation – d'apparence difficile, en réalité maîtrisée –, entre Linguistique et Histoire dans le traitement logométrique en soulignant combien, tout en étant central, le corpus

19. « Pour la lexicométrie, l'objectivité commence quand les projections tendent à faire place à des formalisations. On ne fixe d'avance ni mots-témoins, ni grilles, ni schèmes, ni hiérarchies. On prend tout (exhaustivité des relevés), de la même manière (uniformité du dépouillement) et selon un critère unique... » (Collectif (rééd 1978), *Des tracts en mai 1968*, Paris, Champ Libre, p. 20).

se trouve borné en amont (construction de l'hypothèse de travail) comme en aval (inférence sociopolitique) par l'extralinguistique.

Le corpus est un objet construit : c'est d'ailleurs cela qui en fait un objet de recherche<sup>20</sup>. C'est un objet heuristique qui n'a de sens qu'au regard des interrogations qu'on va lui poser. Or, pour nous, le langage politique participe à et de la construction sociale : c'est en ce sens qu'il est politique. Les corpus de discours politiques ont donc vocation à témoigner de la société (plus que de la Langue), et en conséquence ce sont bien les compétences historiques ou sociales qui président à la sélection des textes politiques puis à leur interprétation ; c'est dire si elles sont fondamentales. Mais répétons-le, ces compétences n'interviennent pas au moment du traitement de l'objet linguistique.

La pluridisciplinarité pour être crédible ne doit pas être de l'indisciplinarité. Dans l'analyse logométrique, trois phases sont à distinguer qui réclament chacune des compétences spécifiques : la constitution du corpus sur la base d'une problématique historico-sociale / le traitement de l'objet textuel qui doit s'en tenir rigoureusement à une approche linguistique / l'inférence sociopolitique qui mobilise l'ensemble de notre culture. Pour favoriser l'étanchéité (c'est-à-dire le contrôle disciplinaire) de ces phases, et particulièrement du caisson central que doit constituer le moment de l'analyse du texte, il a été indiqué ailleurs que les corpus gagnaient toujours à être grands et *réflexifs* (Mayaffre 2002 b) : un corpus idéal étant celui qui par sa taille comporte en lui même ses propres ressources interprétatives c'est-à-dire, autant que possible, son co-texte ou son intertexte.

#### *L'analyse de données textuelles comme épistémè<sup>21</sup>*

Les chercheurs qui maîtrisent mal l'outil logométrique insistent avant tout sur sa dimension documentaire. Inestimable, cette dimension première n'est pas fondamentale. Inestimable, elle l'est en effet car l'ordinateur permet de rechercher systématiquement les occurrences linguistiques dans de grands corpus textuels, là où la mémoire

20. Pour divers développements sur ce point voir le n° 1 de *Corpus* (2002), « Corpus et recherches linguistiques ».

21. Les enjeux épistémologiques de la logométrie ont été esquissés ailleurs, avec d'autres mots, sous une autre lumière : (Mayaffre 2002 a).

humaine touche à ses limites. Il s'agit alors d'attester de la présence, de l'absence, de la récurrence d'un terme ou d'une unité linguistique dans l'œuvre de Robespierre, Mendès France ou Pompidou. Cette fonction documentaire est d'autant plus utile que la convocation de tous les contextes d'utilisation d'une forme permet d'en contrôler le sens selon la vieille affirmation de Pierre Guiraud<sup>22</sup>. Plus loin, la compilation de tous les paragraphes contenant un mot-pôle permet d'en étudier statistiquement l'univers lexical ou grammatical.

Pourtant, fondamentalement, le corpus est examiné ici plus comme une ressource que comme un objet à épuiser. On retrouve le risque d'un corpus alibi que l'on convoquera par intermittence et que l'on exploitera partiellement par le prisme d'une thématique définie *a priori*. Il s'agit en effet d'illustrer une démonstration, de confirmer une hypothèse le plus souvent localement dans le corps de textes. De la même manière que ce sont les hypothèses historiques qui ont présidé au rassemblement des textes en corpus, ce sont encore elles qui risquent d'orienter le traitement (que l'on voudrait linguistique) du corpus. Nous sommes dans le cadre d'une démarche hypothético-déductive dans laquelle le texte (c'est-à-dire l'objet linguistique) se trouve être la caution à la fois d'une hypothèse et d'une déduction historiques. De Gaulle a-t-il bien dit « je vous ai compris ». Si oui, où et quand ? Dans l'océan des discours contemporains, l'ordinateur – certes, ce n'est pas rien ! – retrouvera la phrase pour pointer le discours. Patriote, ce même de Gaulle a-t-il utilisé beaucoup *patrie* et l'a-t-il associé à d'autres mots sans équivoque comme *destin*, *grandeur* ou *puissance* ? L'ordinateur fera un relevé exhaustif de *patrie* dans le discours de de Gaulle, et comparera ce relevé à celui, par exemple, des discours de Mitterrand ou Chirac ; plus loin, il étudiera le réseau des cooccurrences du mot pour en discerner le sens emphatique et patriotique chez le général.

La description est sûre, documentée, comparée mais elle est partielle et seulement confirmatoire (parfois infirmatoire) pour le chercheur en Histoire. Le véritable intérêt de la logométrie est ailleurs : c'est d'être exploratoire c'est-à-dire que sa valeur est avant tout heuristique.

---

22. Le sens d'un mot « se définit finalement par la somme de ses emplois » (P. Guiraud (1960), *Problèmes et méthodes de la statistique linguistique*, Larousse, p. 19).

Une chose est d'interroger localement un corpus, une autre est de se laisser interroger par lui dans toute son étendue. Une chose est de contrôler la fréquence ou l'utilisation d'un mot-élu, une autre est de laisser l'ordinateur traiter tous les mots pour nous interpeller sur ceux qui sont objectivement les plus sous-employés par tel locuteur, sur-employés par tel autre, sur ceux qui ont une distribution chronologique étale et ceux qui connaissent de grandes variations diachroniques, etc.

Fondamentalement la démarche devient inductive, cela signifie, pour ce qui concerne ce débat, que lors du traitement *nous allons du relevé textuel vers l'interprétation sociale, du fait linguistique vers l'inférence politique* ; et non l'inverse. Avec la logométrie, l'analyste part toujours de faits linguistiques matériels massivement attestés (y compris dans leur absence ou leur rareté) pour commencer à réfléchir et le cas échéant inférer un positionnement idéologique ou une attitude politique. Que l'on se comprenne bien, il ne s'agit pas de minimiser le rôle de l'Histoire à qui il revient la belle partie terminale. Mais la démarche logométrique est rigoureuse et peut satisfaire les linguistes car elle s'appuie *toujours et seulement* sur le repérage systématique de *toutes* les unités linguistiques saillantes du corpus dont on souligne la régularité ou la récurrence ; l'interprétation sociolinguistique, les spéculations politiques inférentielles viennent après.

##### **5. Étude de cas : l'ethos chiraquien**

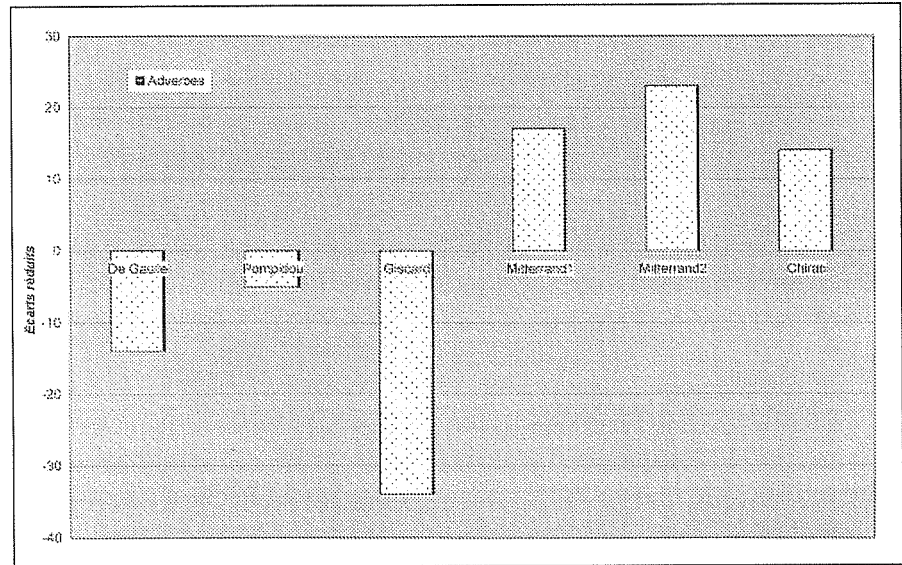
Sans doute convient-il d'illustrer ces propos. On le fera sur le corpus de discours présidentiel de la V<sup>e</sup> République (1958-2002). On cherchera moins ici à développer l'une des principales conclusions de *Paroles de président* (Mayaffre, 2004) qu'à respecter le procès de la démonstration pour donner à voir la dynamique de la démarche logométrique.

Un vaste corpus de plus de 500 discours, composé de l'ensemble des interventions télévisées de de Gaulle, Pompidou, Giscard, Mitterrand et Chirac a été constitué. Dès lors que ces textes ont été rassemblés et que le corpus a été segmenté en 6 parties pour faire contraster les différents présidents ou les différents mandats successifs (de Gaulle (1958-1969)/ Pompidou (1969-1974)/ Giscard (1974-1981)/ ...), nous avons laissé *Hyperbase* décrire le corpus en



toute objectivité pour nous interroger. L'idée essentielle étant, on l'a compris, de ne pas projeter nos *a priori* d'historien ou de citoyen sur les textes.

Après un traitement systématique et exhaustif, le logiciel d'Étienne Brunet relève alors les unités linguistiques statistiquement caractéristiques de de Gaulle ou Pompidou, Mitterrand ou Chirac. Ainsi, par exemple il souligne combien Chirac, après Mitterrand, sur-utilise la catégorie adverbiale à hauteur d'un écart réduit de +13 (graph. 1).

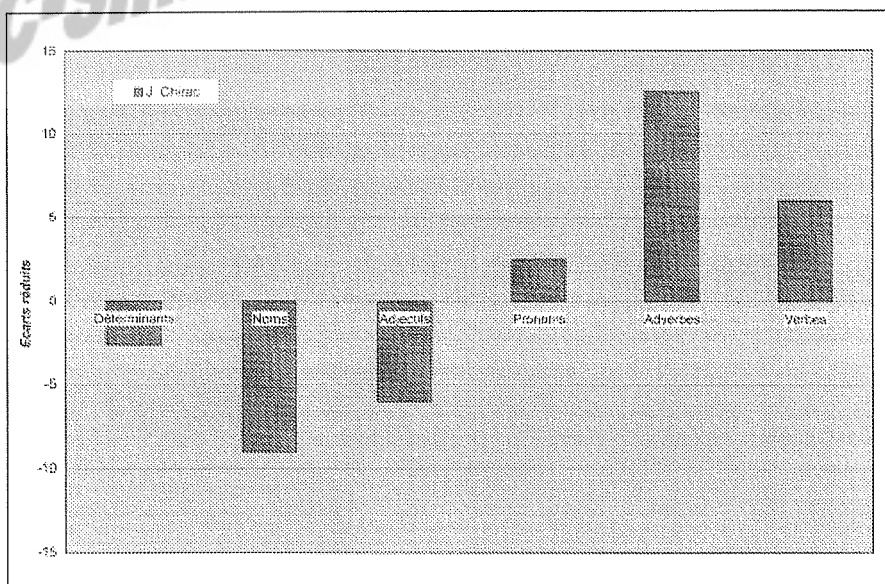


Graphique 1:

Répartition des adverbes dans le corpus présidentiel (1958-2002)

L'interpellation statistique est forte et inattendue. Mieux, reconnaissons-le, il n'était pas prévu, initialement, d'étudier l'usage des adverbes chez Chirac : véritablement, c'est l'ordinateur, après examen systématique de la répartition des catégories grammaticales dans le corpus, qui force l'analyse. Plus généralement, la description rhétorico-grammaticale qu'*Hyperbase* donne du discours chiraquien (par rapport au discours présidentiel moyen) est celle représentée par le graphique 2 (voir page suivante)

En un mot, le discours de Chirac est, d'un point de vue linguistique, un discours assez médian dont les caractéristiques grammaticales



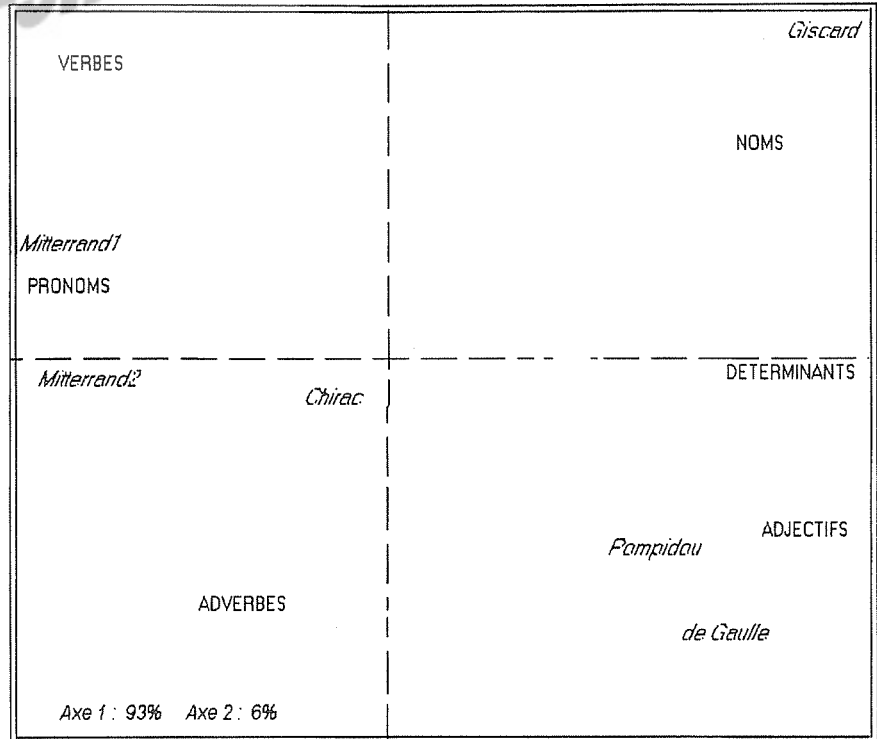
Graphique 2 :  
Spécificités grammaticales de Chirac dans le corpus (1958-2002)

(peu de déterminants, de noms et d'adjectifs, beaucoup de pronoms et de verbes) sont atténuées (écarts réduits négatifs ou positifs relativement faibles, inférieurs à 10). Sa principale originalité – la seule qui soit vraiment importante (écart réduit fort) d'un point de vue statistique – est le suremploi massif de l'adverbe.

Avec d'autres procédés mathématiques (AFC) et dans une vision plus générale du corpus des présidents, l'instrument logométrique indique aussi les proximités représentées sur le graphique 3 (ci-après)

Grossièrement, le discours politique contemporain connaît deux temps. De de Gaulle à Giscard (partie droite du graphique), le discours est *nominal*. Ensuite il devient *verbal* (partie gauche du graphique). Plus précisément, ces deux grands types de discours qu'Étienne Brunet a mis à jour dans le discours littéraire<sup>23</sup> se déclinent comme suit : le discours dit *verbal* de Mitterrand est surtout pronominal, celui de Chirac, assez médian, on l'a dit, est surtout adverbial. Le discours dit *nomi-*

23. É. Brunet (1981), *Le vocabulaire français de 1789 à nos jours*, Paris, Champion. Voir aussi, M. Kastberg Sjöblom (2005), *L'écriture de J.M.G. Le Clézio*, Paris, Champion.



Graphique 3 : Analyse factorielle des correspondances (AFC) des catégories grammaticales du discours présidentiel (1958-2002)

nal de de Gaulle ou Pompidou combine noms et adjectifs, celui de Giscard, très didactique voire austère, s'en tient aux seuls noms, etc.

Bref, à ce stade de la démonstration, le lecteur se trouve amené au point de départ heuristique de la recherche. Sous réserve d'être initié aux traitements statistiques élémentaires répandus dans l'ensemble des sciences humaines, la description systématique du corpus prend la forme d'une vaste interrogation à laquelle on ne peut se soustraire : pourquoi de Gaulle sur-emploie-t-il les noms ? Pourquoi Mitterrand privilégie-t-il les pronoms ? Pourquoi Chirac utilise-t-il anormalement les adverbes ? Pourquoi plus généralement, sous la V<sup>e</sup> République, le discours évolue-t-il d'une tonalité nominale vers une tonalité verbale ? Pourquoi... ?

Le lecteur trouvera une réponse complète dans *Paroles de Président*. Tout juste, est-il possible ici de résumer la conclusion :

Chirac développe une rhétorique de l'*ethos* lorsque Mitterrand développe le *pathos* et les autres présidents le *logos*. En effet, là où les trois premiers présidents utilisaient des noms pour développer avec rigueur une démonstration qui s'appuyait sur un contenu et des concepts politiques, là où Mitterrand cultivait le *pathos* en vidant le discours de son contenu objectif mais en mettant en scène, de manière dramatique, une tension énonciative (pronominale) entre le « moi / je » du président et le « vous » des citoyens, Chirac, lui, invente l'*ethos* présidentiel.

À l'analyse, la sur-utilisation massive des adverbes chez Chirac traduit une volonté éperdue de modaliser, d'actualiser le discours quitte à ce que le prédicat prenne le pas sur le thème, le dire sur le dit, l'illocutoire sur le locutoire. Cette modalisation adverbiale renvoie en effet toujours à l'instance d'énonciation du discours, c'est-à-dire, en l'occurrence, à la personnalité ou à l'autorité du locuteur que le discours est, précisément, en charge d'établir. L'exemple le plus flagrant étant sans aucun doute la sur-utilisation du mot qui entre tous les mots a la faveur de Chirac : « naturellement » (écart réduit de +23 par rapport aux autres présidents). Cet adverbe constamment répété est la clef de voûte de la rhétorique chiraquienne : il a pour fonction de présenter comme évidents, comme « naturels », des propos qui ne le sont pas. Or, après examen, cette (fausse) évidence puise sa force uniquement dans une légitimité absolue (relevant de la Nature) de la personne du président sous la Ve République gaullienne ; personne du président que le discours présidentiel a pour vocation de rappeler ou de (re)construire.

### Conclusion

On le voit, l'analyse débouche ici sur un traitement rhétorique du discours présidentiel, bénéficiant notamment de la remise au goût du jour du triangle aristotélicien par exemple dans les travaux de Ruth Amossy (Amossy, 1999 et 2000). Plus loin, l'étude aide à la compréhension du fonctionnement institutionnel de la Ve République présidentielle ou du rôle iconique de la télévision – média privilégié de la parole présidentielle – dans la démocratie. Cependant, le propos repose sur un traitement linguistique rigoureux qui ne cautionne pas mais préside à cette démonstration comme aux autres dans notre travail. En ce sens, l'analyse du discours politique sur laquelle débouche

la linguistique de corpus, l'analyse des données textuelles et particulièrement la logométrie, apparaît crédible car elle se fonde toujours et seulement sur le corpus textuel qu'elle entend épuiser dans ses composantes linguistiques avant de s'évader vers des interprétations littéraires, historiques ou citoyennes d'autant plus tranchées qu'elles s'enracinent dans des faits linguistiques massifs avérés.

### RÉFÉRENCES BIBLIOGRAPHIQUES

- ADAM J.-M. (1999), *Linguistique textuelle. Des genres de discours aux textes*, Paris, Nathan.
- AMOSSY R. (éd.) (1999), *Images de soi dans le discours. La construction de l'éthos*, Lausanne, Delachaux & Niestlé.
- AMOSSY R. (2000), *L'argumentation dans le discours*, Paris, Nathan.
- BONNAFOUS S. et alii (s. dir.) (2003), *Argumentation et discours politique*, Rennes, PUR.
- BOURIGAULT D. et LAME G. (2002), « Analyse distributionnelle et structuration de terminologie. Application à la construction d'une ontologie documentaire du Droit », *Traitement automatique des langues*, n° 47 : 1.
- BRUNET É. (2000), « Qui lemmatise, dilemme attise », *Lexicometrica*, n° 2, (<http://www.cavi.univ-paris3.fr/lexicometrica/article/numero2/brunet2000.html>)
- (2002) « Le lemme comme on l'aime », in *JADT 2002*, Saint Malo, IRISA-INRIA, pp. 221-230.
- BUFFON B. (2002), *La parole persuasive : théorie et pratique de l'argumentation rhétorique*, Paris, PUF.
- CHARAUDEAU P. & MAINGUENEAU D. (dir.) (2002), *Dictionnaire d'analyse du discours*, Paris, Seuil.
- CORPUS, n° 1, 2002, « Corpus et recherche linguistiques ».
- GUILHAUMOU J. (2004), « Où va l'analyse de discours ? Autour de la notion de formation discursive », *Texte !* ([http://www.revue-texte.net/Inedits/Guilhaumou\\_AD.html](http://www.revue-texte.net/Inedits/Guilhaumou_AD.html)).

- GUILHAUMOU J., MALDIDIER D., ROBIN R. (1994), *Discours et archive. Expérimentations en analyse de discours*, Liège, Mardaga.
- HABERT B. et alii (1997), *Les linguistiques de corpus*, Paris, Colin, 1997.
- LABBÉ D. & MONIERE D. (2003), *Le discours gouvernemental. Canada, Québec, France (1945-2000)*, Paris, Champion.
- MAINGUENEAU D. (1976), *Initiation aux méthodes de l'analyse du discours*, Paris, Hachette.
- (1987), *Nouvelles tendances en analyse du discours*, Paris, Hachette.
- (1991), *L'Analyse du Discours. Introduction aux lectures de l'archive*, Paris, Hachette.
- MAINGUENEAU D. (dir.) (1995), *Langages*, n° 117, « Les analyses du discours en France ».
- MALDIDIER D. (1971), « Le discours politique de la guerre d'Algérie : approche synchronique et diachronique », *Langages*, n° 23, pp. 57-86.
- MALRIEU D. & RASTIER, F. (2001), « Genres et variations morphosyntaxiques », *Traitements automatiques du langage*, n° 42-2, pp. 547-577.
- MAYAFFRE D. (2002 a), « L'herméneutique numérique », *L'Astrolabe*, (<http://www.uottawa.ca/academic/arts/astrolabe>).
- (2002 b), « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, n° 1, pp. 51-71.
- (2004), *Paroles de président. Jacques Chirac et le discours présidentiel sous la v<sup>e</sup> République*, Paris, Champion.
- (2005), « De la lexicométrie à la logométrie », *Astrolabe* (<http://www.uottawa.ca/academic/arts/astrolabe>).
- PECHEUX M. (1969), *Analyse automatique du discours*, Paris, Dunod.
- PROST A (1988), « Les mots » in R. Rémond (dir.), *Pour une histoire politique*, Paris, Seuil, pp. 255-285.
- RASTIER F. (2001), *Arts et sciences du texte*, Paris, PUF.
- (2004), « Doxa et lexique en corpus — pour une sémantique des "idéologies" », Actes des Journées Scientifiques en linguistique 2002-2003, J. Pauchard et F. Canon-Roger (dir.), PUR, n° 22.

- (2005), « Enjeux épistémologiques de la linguistique de corpus », conférence aux 2<sup>e</sup> Journées de Linguistique de Corpus, Lorient, septembre 2002 (à paraître dans la publication des actes sous la direction de Geoffrey Williams (dir.) consultable sur *Texto !* ([http://www.revue-texto.net/Inedits/Rastier\\_Enjeux/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier_Enjeux/Rastier_Enjeux.html)))

RASTIER F. et MALRIEU D. (2003), « Analyse du discours et épistémologie », *Texto !*, 2003 (<http://www.revue-texto.net/Dialogues/Rastier-Malrieu.html>).

ROBIN R. (1973), *Histoire et Linguistique*, Paris, Colin.

- (1986), « Postface. L'Analyse de discours entre la linguistique et les sciences humaines : l'éternel malentendu », *Langages*, n° 81, pp. 121-128.

Fac-similé