



## Mots. Les langages du politique

75 | 2004

Émotion dans les médias

---

Pascal Marchand, *L'analyse du discours assistée par ordinateur*

Jean-Marc Leblanc

---



### Édition électronique

URL : <http://journals.openedition.org/mots/3643>

ISSN : 1960-6001

### Éditeur

ENS Éditions

### Édition imprimée

Date de publication : 1 juillet 2004

Pagination : 141-143

ISBN : 2-84788-057-7

ISSN : 0243-6450

### Référence électronique

Jean-Marc Leblanc, « Pascal Marchand, *L'analyse du discours assistée par ordinateur* », *Mots. Les langages du politique* [En ligne], 75 | 2004, mis en ligne le 23 avril 2008, consulté le 19 avril 2019. URL : <http://journals.openedition.org/mots/3643>

---

destiné aux experts et non au destinataire universel, le discours des institutions internationales à vocation universelle ne peut être qu'un simulacre de discours constituant et feindre l'autorité des discours fondateurs.

Reste qu'à cette non-puissance les instances internationales ne peuvent substituer que les mots : une rhétorique qui constitue à la fois l'essentiel de leur pouvoir et la garantie de sa préservation.

*Fabienne Pierre*

Pascal Marchand,  
*L'analyse du discours assistée par ordinateur*,  
1998, Paris, Armand Colin, 222 pages.

Le titre de cet ouvrage se veut résolument généraliste. Pascal Marchand, du Groupe de Recherche sur la Parole de l'Université Paris 8-Saint-Denis, propose une recension des applications de l'outil informatique à l'analyse du discours et, plus largement, à l'ensemble des sciences humaines et sociales qui s'intéressent à l'étude de données verbales. Le volume présente d'abord les « concepts, méthodes et outils », qui sont illustrés, au travers d'un exemple d'application, dans une seconde partie.

Les cinq premiers chapitres sont autant d'axes pour l'analyse discursive auxquels l'outil informatique peut apporter des éléments de réponse. La part belle est faite à la statistique lexicale, le chapitre qui s'y rapporte représentant près de la moitié de cet ensemble. Le premier chapitre présente succinctement ce que l'auteur nomme les « analyses para-verbales ». Il y aborde certains aspects de l'analyse conversationnelle, en termes de verbalisation, de successions de tours de parole et de comportements verbaux, puis définit quelques problématiques relevant de la phonologie, citant des travaux menés sur l'intonation ou des recherches en psychologie sociale et en sociolinguistique, par exemple celles de Labov, en particulier son analyse de la prononciation dans les différentes couches socio-professionnelles new-yorkaises. Le deuxième chapitre, sur les analyses lexicales, s'appuie sur des travaux nombreux et divers, qui fournissent des repères conceptuels, abondamment nourris de citations de Muller, Benveniste, Bronckart, Bally, Saussure, Harris, etc. Il s'inspire largement du manuel bien connu de Ludovic Lebart et André Salem<sup>12</sup>, revient sur les différentes approches du lexique (psychologique, pragmatique, linguistique) puis définit les notions élémentaires d'unité lexicale, d'index, de forme graphique, avant d'aborder des aspects tels que la richesse du vocabulaire, la connexion et la distance lexicale, les concordances,

---

12. L. Lebart et A. Salem, 1994, *Statistique textuelle*, Paris, Dunod.

les segments répétés, les spécificités. Quelques exemples d'application à l'indexation documentaire sont aussi évoqués. Le sous-chapitre intitulé « Statistique textuelle » présente les deux familles de statistique multidimensionnelle que sont les analyses factorielles et les classifications automatiques, fournissant des aides à l'interprétation mais aussi un approfondissement des algorithmes utilisés. L'auteur clôt ce chapitre en introduisant la problématique des unités à considérer concernant les flexions d'une forme, les mots composés, les locutions et séquences figées. Il rappelle le débat entre « lemmatiseurs » et « formalistes » et les procédés de réduction de formes, notamment ceux utilisés par le logiciel *Alceste*. Il revient aussi sur les différentes conceptions du mot.

Consacré aux analyses morpho-syntaxiques, le troisième chapitre montre comment les traitements informatisés peuvent repérer les relations de syntaxe dans la phrase et les parties du discours. Les statuts de l'article et du pronom, des verbes et des adjectifs, des opérateurs et des connecteurs sont mis en relation avec des études relevant de plusieurs disciplines. Sont ensuite posés les principes de l'analyse sémantique : on distingue une approche consistant à synthétiser le contenu d'un texte pour en livrer une compréhension *a posteriori* par la classification des énoncés qui le composent, et une approche *a priori* consistant à projeter les catégories générales d'une langue sur plusieurs corpus. Les règles ontologiques, dictionnaires thématiques, constructions de catégories sont exposés par le biais de travaux couvrant le champ des sciences humaines. Une présentation du logiciel *Tropes* et de ses classes d'équivalence, illustrée d'un exemple d'application clôt le chapitre et l'auteur achève cette première partie en abordant la dimension pragmatique du discours et en soulignant la difficulté de coder une telle approche.

Embrassant les différents niveaux discursifs, il dresse ainsi un état des lieux de l'analyse automatisée du discours, par la description des caractéristiques et des fonctionnalités principales des logiciels les plus répandus en France dans les sciences humaines et par des exemples d'application et d'interprétation, replaçant chaque notion dans son contexte théorique. Sa terminologie pourrait, sur certains points, être discutée, si elle ne renvoyait à un vaste champ d'applications différentes. Les logiciels *Tropes*, *Hyperbase*, *Sphinx*, *Lexico*, *Intex*, *Spadt*, etc. sont présentés de façon pédagogique, ce qui concourt à conférer à l'ouvrage plusieurs niveaux de lecture. Le chercheur rompu aux techniques de la statistique lexicale y trouvera quelques approfondissements concernant les calculs mis en œuvre, tandis que le néophyte y découvrira un panorama méthodologique, mais aussi des liens bibliographiques vers les spécialistes de chaque approche.

Dans la seconde partie, la méthodologie laisse place à l'analyse, ce qui achève de donner à l'ouvrage sa dimension concrète. S'inspirant des approches évoquées dans la première partie, les auteurs (Pascal Marchand, Brigitte Lecat et Janine Larrue) étudient les interventions de 29 journalistes politiques interro-

geant les candidats aux élections européennes de 1994 dans le cadre d'émissions radiodiffusées, à partir de l'hypothèse qu'ils mettent en place des stratégies discursives différentes selon les invités qu'ils ont à interroger. L'analyse des temps de parole montre que les candidats des listes traditionnelles (PS, PC, Génération Écologie mais plus encore UDF/RPR et FN) investissent l'espace discursif de façon plus importante que les représentants des listes nouvelles. L'analyse lexicale, appliquée aux seules interventions des journalistes, confirme ces disparités. Les indices de longueur (en occurrences), de richesse et de banalité, obtenus au moyen de *Sphinx*, puis l'analyse des spécificités d'emploi, due à *Lexico*, mettent en évidence des comportements sensiblement différents des journalistes face à leurs invités. Une analyse factorielle des correspondances réalisée sur les formes graphiques révèle un système d'opposition entre les interventions tenues face aux listes nouvelles comme «Autre Europe» ou «Radical» et les listes plus traditionnelles (PS, PC, RPR/UDF, verts, FN). Cette hypothèse est confirmée par l'examen des emplois et distributions des pronoms, verbes et prédicats, modalisateurs et joncteurs, adjectifs qualificatifs, mené au moyen de *Tropes*. Une analyse factorielle réalisée sur les 50 catégories «logico-syntaxiques» recensées par le logiciel confirme les clivages mis en lumière par celle des formes graphiques. L'étude s'attache enfin aux grands thèmes abordés par les journalistes, tels qu'ils sont déterminés par *Tropes*. Cette recherche, dont certaines conclusions ont été présentées lors des Journées internationales d'analyse statistique des données textuelles (JADT, Nice, 1998), tend à démontrer que le traitement réservé aux candidats se fait moins en fonction de leur idéologie politique que de leur notoriété et de l'attente supposée dans l'opinion publique.

Le lecteur aura pris conscience de l'intérêt des interventions de procédures informatisées, au niveau de la verbalisation, du lexique, de la syntaxe, de la sémantique et, pourrions-nous ajouter, de l'énonciation. Démonstration est faite que l'opposition du quantitatif et du qualitatif n'est pas pertinente, tant il est vrai que ces deux aspects ne constituent, tout au plus, que deux moments de l'analyse. Il n'en reste pas moins que l'outil informatique ne saurait suppléer à une démarche méthodologique rigoureuse, à l'établissement d'hypothèses préalables et à la connaissance du matériau étudié. Il appartient au chercheur d'acquérir aussi une compétence minimale quant aux procédures de traitement, de constituer un corpus susceptible de permettre une interprétation fiable et d'éviter l'écueil qui consisterait à prêter aux résultats issus de l'ordinateur un caractère de vérité absolue. Car, au final, c'est à l'analyste qu'il revient de mener l'interprétation et de valider ou invalider ses hypothèses. Comme le souligne l'auteur, «on peut faire dire aux mots autant d'inepties qu'aux chiffres» (p. 138).

Jean-Marc Leblanc