

CORPUS

Corpus

4 | 2005

Les corpus politiques : objet, méthode et contenu

---

## Le sens du texte : entre opacité statistique et transparence énonciative

Dominique Desmarchelier

---



### Édition électronique

URL : <http://journals.openedition.org/corpus/372>  
ISSN : 1765-3126

### Éditeur

Bases ; corpus et langage - UMR 6039

### Édition imprimée

Date de publication : 1 décembre 2005  
ISSN : 1638-9808

### Référence électronique

Dominique Desmarchelier, « Le sens du texte : entre opacité statistique et transparence énonciative », *Corpus* [En ligne], 4 | 2005, mis en ligne le 01 septembre 2006, consulté le 20 avril 2019. URL : <http://journals.openedition.org/corpus/372>

---

Ce document a été généré automatiquement le 20 avril 2019.

© Tous droits réservés

---

# *Le sens du texte : entre opacité statistique et transparence énonciative*

Dominique Desmarchelier

---

- 1 L'analyse du discours, dans une perspective argumentative, lorsqu'elle s'applique à des corpus qu'on qualifiera de « lourds », place le linguiste face à des choix méthodologiques parfois délicats : comment concilier le désir de prendre en compte la globalité du texte (de 200 000 à plus d'un million d'occurrences), ce que permet effectivement un traitement lexicométrique, et le souci de mettre en évidence les orientations énonciatives et argumentatives, qui se fonderont sur des hypothèses sémantiques préexistant au corpus ? Comment passer d'un traitement statistique opaque par essence, puisqu'il est capable de saisir en aveugle, de classer, de trier toutes les unités présentes dans le corpus, à l'analyse fine de l'émergence d'arguments reposant non seulement sur le jeu de l'interaction, sur la présence de connecteurs ou d'opérateurs (*mais, parce que, justement, décidément, finalement ...*), de modalisations (*pouvoir, devoir, vouloir, savoir, croire...*), mais aussi sur les différentes formes d'implicites sémantiques ou pragmatiques. Comment, enfin, faire en sorte que les hypothèses sémantiques puissent orienter la collecte des données statistiques et que les résultats fournis par la lexicométrie éclairent le regard du sémanticien ?
- 2 J'illustrerai mes propos à partir d'une étude collective<sup>1</sup> portant sur l'« Argumentation dans l'espace public contemporain, le cas du débat sur l'immigration ». Les corpus recueillis à cette occasion portaient sur les débats parlementaires, les débats télévisés, les textes officiels relatifs à la scolarisation d'enfants issus de l'immigration, les articles de l'*Encyclopædia Universalis*, des sites Internet, des discussions entre étudiants et des conversations chez un marchand de journaux. Cette recherche exploite les compétences de deux équipes en analyse argumentative d'une part, et en analyse du discours et lexicométrie d'autre part.
- 3 On peut considérer qu'il existe au moins trois conceptions de l'argumentation :
  - Argumentation 1 : « Au moyen de... »
- 4 « Moyen de persuader un auditoire de la justesse d'une thèse ». On parle de « techniques discursives permettant d'accroître l'adhésion des esprits aux thèses qu'on présente à leur

assentiment » [Perelman & Olbrechts-Tyteca 1970]. Ce que certains ont pu appeler « idéal argumentatif » ; (situations dialogales / monologales, débat, conversation) ;

- Argumentation 2 : « A travers... »

5 « Mode de construction d'un discours visant à le rendre plus résistant à la contestation » ; discours, et contre-discours (perspective dialogique) [Ch. Plantin 1993] ;

- Argumentation 3

6 Adoptant une perspective ascriptiviste de la langue, Ducrot & Anscombre ([Anscombre & Ducrot 1983] puis [Anscombre 1995]) placent les orientations argumentatives à l'intérieur même de la langue. Dès lors, le recours à des corpus « authentiques » (fable de La Fontaine ou texte de Racine) ne peut avoir pour seule fin que de valider des hypothèses déjà étayées par des énoncés construits. Cette position les conduit à écarter de leurs recherches tout discours politique ou publicitaire.

- La description des interactions

7 Deux approches complémentaires seront ici en présence.

8 La première, lexicométrique, permet dans un débat de quantifier les prises de paroles, de les attribuer à différents locuteurs ou groupes, et d'opposer ainsi, dans les débats à l'Assemblée Nationale par exemple, prises de parole « autorisées » à « interruptions ». De même, dans un corpus long, il devient possible de repérer rapidement les spécificités lexicales (spécificité de l'apparition d'un événement textuel dans telle ou telle partie d'un corpus, plutôt que dans telle autre) et les segments répétés (ensemble syntagmatique de N éléments apparaissant plusieurs fois dans le texte [Lafon & Salem 1983]), en prenant en compte la dimension diachronique du corpus, afin de mettre en évidence certains glissements thématiques.

9 La seconde, s'appuyant sur des analyses argumentatives et conversationnelles, insiste davantage sur la nature des interventions et leur rôle dans la progression de l'échange. Le caractère complémentaire des méthodes utilisées apparaît clairement dès ce premier niveau que nous pourrions qualifier de descriptif.

- La catégorisation

10 Comme le montre Christian Plantin [1990], le travail de catégorisation (référentielle et non morphosyntaxique) constitue souvent la première étape vers l'élaboration d'une argumentation<sup>2</sup>. Nommer l'autre, c'est déjà le situer dans un cadre social, en l'opposant à ce qu'il n'est pas. *Omnis determinatio est negatio*, disait Spinoza. Cette remarque s'applique à la plupart des corpus analysés. Il n'est pas indifférent d'être classé « enfant d'immigrés », « enfant issu de l'immigration », ou encore « élève en difficulté ». De même, la dénomination « jeunes des cités » recouvre souvent un amalgame entre « enfants de la deuxième génération », « jeunes ayant quitté l'école », pour ne pas dire « sauvages ».

11 L'analyse lexicométrique, à travers les cooccurrences (ce qui revient à une contextualisation des termes), met également en évidence les qualifications attribuées à l'immigré : « clandestin » versus « en situation régulière », ou encore, sous forme polyphonique dans la bouche d'un député du PCF « ces étrangers qui nous dérangent ». Le recours à des formes génériques « ces gens-là », « eux », « tous ceux qui sont en France », permet également de ne pas spécifier avec précision les gens en question.

12 Nous prendrons également en compte :

- l'émergence et la circulation des arguments ;
- les marqueurs argumentatifs (connecteurs, modalisateurs) ;

- une typologie des arguments les plus récurrents : autorité, communauté, cadrage et analogie [Breton 2001] ;
  - les stratégies employées dans des lieux différents.
- 13 A la différence de l'analyse conversationnelle qui a le souci de matérialiser le plus fidèlement possible dans l'espace graphique les « enjeux » de l'interaction (interruptions, chevauchements, hésitations, gestes, attitudes posturales, éléments de proxémique), l'analyse lexicométrique ne peut que balayer l'intégralité d'un « texte », qui certes aura été enrichi (didascalies, identification des locuteurs...), mais qui présentera au final une linéarité souvent éloignée de la situation réelle de l'échange. Un travail de délinéarisation s'avèrera donc nécessaire pour redonner sa densité à l'interaction.
- 14 Un exemple de transformation d'un corpus brut en corpus enrichi peut-être donné :

#### 1 Corpus brut : page du JO (édition HTML)

ENTRÉE ET SÉJOUR DES ÉTRANGERS EN FRANCE ET DROIT D'ASILE  
 SÉANCE 1 jeudi 4 décembre 1997  
 Discussion, après déclaration d'urgence, d'un projet de loi  
 le président. L'ordre du jour appelle la discussion, après déclaration d'urgence, du projet de loi relatif à l'entrée et au séjour des étrangers en France et au droit d'asile (n<sup>os</sup> 327, 451, 483).  
 Avant que la discussion ne s'engage, je ferai deux observations.  
 Nous avons prévu un temps assez important pour ce débat puisque le vote final interviendra mardi en huit. Je souhaite que nous disposions du temps nécessaire pour examiner le texte et les amendements,...

[1] Jean-Louis Debré, RPR. Voulez-vous que nous déposions d'autres amendements ?  
 le président. ... sans qu'il y ait pour autant d'obstruction.  
 En second lieu, et surtout, comme il s'agit d'un sujet sensible, je souhaite que le débat se déroule dans le climat de dignité et de respect nécessaire.

[2] Rudy Salles, UDF. Ça dépend du Gouvernement !  
 le président. La parole est à M. le ministre de l'intérieur.

[3] Jean-Pierre Chevènement, MDC, *ministre de l'intérieur*. Monsieur le président, mesdames, messieurs les députés, nous allons débattre de l'entrée et du séjour des étrangers et du droit d'asile. Mais parler des étrangers, c'est aussi une autre manière de parler de la France. Sur tous ces bancs, nous nous en faisons une certaine idée qui transcende souvent, et c'est heureux, les appartenances politiques.  
 Le peuple français, parce qu'il a mêlé historiquement des peuples appartenant au Nord et au Sud de l'Europe, des Flamands, des Catalans, des Bretons et des Corses, des Alémaniques et des Languedociens, ne s'est jamais défini par la souche, mais toujours par le bonheur d'être ensemble pour donner sens à son histoire.(...)

#### 2 Le même texte enrichi selon [Heiden 1999]

```
<?NSL DDB /optb/corpus/lxp/reseda/reseda.lml.ddb 0>
<DIV N=1 LEVEL=2 TITLE="ENTRÉE ET SÉJOUR DES ÉTRANGERS EN FRANCE ET
DROIT D'ASILE">
<TITLE>Discussion, après déclaration d'urgence, d'un projet de loi</TITLE>
<SP COM=1 COM=1 WHO='le président' PARTI=NA><P> L'ordre du jour appelle la
discussion, après déclaration d'urgence, du projet de loi relatif à l'entrée et au
séjour des étrangers en France et au droit d'asile </P><REM>nos 327, 451, 483</
REM>.
<P>Avant que la discussion ne s'engage, je ferai deux observations.</P><P><C
V=Nous>nous</C> avons prévu un temps assez important pour ce débat puisque le
vote final interviendra mardi en huit. <C V=Je>je</C> souhaite que nous disposions
```

du temps nécessaire pour examiner le texte et les amendements,...</P></SP>

<SP COM=2 WHO='Jean-Louis Debré' PARTI=RPR><P> Voulez-vous que nous déposions d'autres amendements ?</P></SP>

<SP COM=1 WHO='le président' PARTI=NA><P> ... sans qu'il y ait pour autant d'obstruction.</P><P>En second lieu, et surtout, comme il s'agit d'un sujet sensible, je souhaite que le débat se déroule dans le climat de dignité et de respect nécessaire.</P></SP>

<SP COM=2 WHO='Rudy Salles' PARTI=UDF><P> Ça dépend du Gouvernement !</P></SP>

<SP COM=1 WHO='le président' PARTI=NA><P> <C V=La>la</C> parole est à <ABBR EXPAN=M.>Monsieur</ABBR> le ministre de l'intérieur.</P></SP>

<SP COM=1 WHO='Jean-Pierre Chevènement' PARTI=MDC TITLE="ministre de l'intérieur"><P> Monsieur le président, mesdames, messieurs les députés, nous allons débattre de l'entrée et du séjour des étrangers et du droit d'asile. <C V=Mais>mais</C> parler des étrangers, c'est aussi une autre manière de parler de la France. Sur tous ces bancs, nous nous en faisons une certaine idée qui transcende souvent, et c'est heureux, les appartenances politiques.</P>

<P>Le peuple français, parce qu'il a mêlé historiquement des peuples appartenant au Nord et au Sud de l'Europe, des Flamands, des Catalans, des Bretons et des Corses, des Alémaniques et des Languedociens, ne s'est jamais défini par la souche, mais toujours par le bonheur d'être ensemble pour donner sens à son histoire.</P>

- 15 Le traitement lexicométrique facilite l'exploitation documentaire du corpus et permet son traitement statistique.
- Dimension du corpus et de ses partitions.
- 16 Sur l'ensemble du corpus, tous partis confondus, la répartition des interruptions et des interventions autorisées des orateurs est la suivante :

Tableau 1 : Dimensions du Corpus RESEDA

Corpus	Occurrences	Formes	Phrases	Prises de Parole
<b>Orateurs</b>	<b>188 027</b>	12 605	9776	2096
<b>Interrupteurs</b>	13 689	2423	2603	<b>2349</b>
<b>Total</b>	<b>201 716</b>	<b>15 028</b>	<b>12 379</b>	<b>4445</b>
<b>% interruptions</b>	<b>6,8%</b>	<b>16,10%</b>	<b>21%</b>	<b>52,85%</b>

- 17 On notera sans surprise que les interruptions se réduisent le plus souvent à une phrase (2 603 phrases pour 2 349 interruptions) alors que les orateurs autorisés emploient en moyenne 4,6 phrases par interventions (9 776 phrases pour 2 096 interventions).
- 18 Soit au total 4 445 interventions contenant 12 379 phrases soit 2,8 phrases par prise de parole.
- 19 Ici, la phrase doit être comprise comme un ensemble d'éléments graphiques isolé par des signes de ponctuation forte (il n'y a pas d'étiquetage morphosyntaxique pour cette recherche).
- Concordances : présentation du contexte immédiat pour une forme donnée. L'exemple de « courage » (extraits) :

- 20 Il y a 2 occurrences de « courage » dans le corpus reseda/sp com=1-sp-parti-COM

<a href="#">3 (1627)</a> , <a href="#">Andre Serin, COM</a>	une politique de gauche qui réussisse . c' est cela le	<b>courage</b>	aujourd' lui Monsieur le président monsieur le
<a href="#">5, 7999 (3579)</a> , <a href="#">Andre Serin, COM</a>	qui réussisse . c' est de ce côté qu' il faut chercher le	<b>courage</b>	je voudrais appeler plus particulièrement l' attention

- 21 Il y a 7 occurrences de « courage » dans le corpus reseda/sp com=1-sp-parti-SOC

1	<a href="#">3, 6996 (2108)</a> <a href="#">l, Yann Galet, SOC</a>	, sociale et culturelle . nous devons avoir le	<b>courage</b>	de dire qu' il n' y a pas de lien direct entre immigration et
2	<a href="#">4, 6996 (2816)</a> <a href="#">l, Jean-Marc Ayrault, SOC</a>	mes chers collègues . le Gouvernement a eu le	<b>courage</b>	de présenter un texte équilibré , un texte amendé ,
3	<a href="#">5, 6996 (3329)</a> <a href="#">l, Gerard Bouzes, SOC</a>	vos électeurs dans le mauvais sens du poil ! Ayez le	<b>courage</b>	de résister . Allez voir vos électeurs et expliquez
4	<a href="#">7, 6996 (3926)</a> <a href="#">l, Christophe Carasche, SOC</a>	un hommage particulier . il fallait avoir le	<b>courage</b>	et la volonté de dire la vérité aux Français , c' est-à-dire que
5	<a href="#">8, 6996 (4041)</a> <a href="#">l, Gerard Bouzes, SOC</a>	quelle autre politique alors qu' ils n' ont pas eu le	<b>courage</b>	de s' exprimer ici-même , se bornant à l' expliquer devant
6	<a href="#">8, 6996 (4045)</a> <a href="#">l, Gerard Bouzes, SOC</a>	de l' intérieur . qu' il vous faut beaucoup de	<b>courage</b>	pour résister aux critiques de ceux qui ne proposent finalement
7	<a href="#">8, 6996 (4445)</a> <a href="#">l, Jean-Marc Ayrault, SOC</a>	les devoirs des étrangers en France . il fallait du	<b>courage</b>	et de la franchise , monsieur le ministre . vous n' en

weblex, copyright © Serge Heiden UMR8503,  
[slh@ens-lsh.fr](mailto:slh@ens-lsh.fr)

- Les cooccurrences

- 22 A l'intérieur d'un espace de rencontre préalablement défini (phrase, paragraphe), mise en évidence d'unités, à une distance n, n+1, n+x d'une unité choisie.

## Cooccurrences du corpus reseda-sp-com-1-sp-parti-SOC

Seuils :  $f_3$ ,  $cf_3$ ,  $p$  5.0E-2,  $d_m$  1000.0

A	B	$f_A$	$f_B$	$cf$	$p$	$d_m$
chers collègues		64	71	63	7e-108	0.0
flux migratoires		39	33	33	6e-65	0.0
monsieur ministre		78	89	48	2e-50	1.0
regroupement familial		19	17	17	3e-39	0.0
projet loi		70	112	43	3e-39	1.0
droit asile		102	72	40	1e-35	1.6
Patrick Weil		14	16	14	3e-33	0.0
droits homme		46	38	24	2e-32	2.2
maîtrise migratoires		15	33	15	2e-28	2.1
maîtrise flux		15	39	15	4e-27	1.1
lois Pasqua-Debré		64	19	18	6e-27	0.6
Monsieur ministre		118	89	38	1e-25	2.5
asile territorial		72	17	17	1e-25	0.1
affaires étrangères		10	12	10	7e-25	0.0
vie familiale		24	16	13	4e-24	1.4
certificat hébergement		10	14	10	1e-23	1.0
Monsieur président		118	29	22	3e-22	1.1
motivation refus		14	18	11	7e-22	2.4
privée familiale		11	16	10	9e-22	1.2
carte résident		37	15	13	9e-22	1.0
faveur liberté		14	28	12	1e-21	2.2
Front national		9	14	9	4e-21	0.0
reconduite frontière		9	14	9	4e-21	2.0
titre séjour		22	75	17	6e-21	1.1

48543 couples (4500 paires), 1373 couples sélectionnés  
(Weblex – Serge Heiden UMR8503)

- Les segments répétés

- 23 Ensembles syntagmatiques de  $n$  éléments contigus, les plus fréquents. Exemple dans le sous-corpus du parti socialiste :

## Segments Répétés des SOC (sur l'ensemble des 8 séances) de fréquence 5, 4 et 3

- “ les autorités de la République ont toujours le droit de donner asile à tout étranger persécuté en raison de son action en faveur de la liberté ou qui sollicite la protection de la France pour un autre motif ” (4 occ.)
- “ Tout homme persécuté en raison de son action en faveur de la liberté a droit d'asile sur les territoires de la République ” (3 occ.)
- “ la dépénalisation du défaut de déclaration d'entrée sur le territoire ” (4 occ.)
- “ le relèvement d'un arrêté de reconduite à la frontière depuis ” (3 occ.)
- “ l'article 8 de la Convention européenne des droits de l'homme ” (3 occ.)
- “ loi sur l'entrée et le séjour des étrangers en France ” (5 occ.)
- “ d'une commission du titre de séjour ” (3occ.)
- “ des cas d'attribution de plein droit ” (3 occ.)
- “ l'institution d'une carte de séjour ” (3 occ.)
- “ là où il n'y avait que ” (3 occ.)
- “ la prolongation de vingt-quatre heures du délai ” (3 occ.)
- “ Le préambule de la constitution de 1946 ” (3 occ.)
- “ responsable de la montée du Front national ” (3 occ.)

## Segments répétés avec retour au contexte. Exemple :

JO : séances AN 3, 6 et 8. Argumentaire préétabli par les députés communistes

**André Gerin**, COM. (séance 3 : 4/12/97) On prétendait travailler au rayonnement de la pensée française en commençant par fermer portes et fenêtres. Si nous pouvions nous dégager des passions du moment, **nous comprendrions peut-être que ces étrangers qui nous dérangent peuvent nous rendre un grand**

**service** : nous inciter à nous interroger sur notre société, notre identité et notre culture.

**André Gerin**, COM. (séance 6 : 25/2/98) Comment la droite pouvait-elle prétendre travailler au rayonnement de la pensée française en commençant par fermer portes et fenêtres ?

Si nous pouvions nous dégager des passions du moment, **nous comprendrions peut-être que ces étrangers qui nous dérangent peuvent nous rendre un grand service** en nous incitant à nous interroger sur notre société, notre identité et notre propre culture.

- Spécificités

24 Il est possible de comparer la fréquence absolue d'une unité (nb. d'occurrences) dans les différentes parties du corpus, ou, plus pertinemment, d'étudier la fréquence relative des unités (fréquence absolue dans les sous-corpus/taille des sous-corpus). Mais l'outil statistique le plus performant reste les spécificités. Exemple :

Spécificités décroissantes de la partie 1 du corpus reseda-sp-com-1-sp-parti-COM

Liste limitée aux 25 premières lignes webdex, copyright © Serge Heiden UMIR8503, [silvacoens-icth.fr](http://silvacoens-icth.fr)

construire	15	9	8+	1	-1	1	1+	3	-1	0	-1	0	-2	1	-1	0	-1
politique	436	58	7+	103	2+	19	3+	123	2+	4	-1	66	-8	59	-4	4	1+
voulons	62	17	7+	9	-1	4	2+	10	-1	1	1+	7	-3	14	1+	0	-1
culture	36	13	7+	6	-1	0	-1	8	-1	0	-1	2	-3	3	-2	4	5+
avec	542	62	6+	109	1+	17	2+	117	-1	6	-1	117	-2	112	1+	2	-1
députés	78	18	6+	9	-2	4	2+	32	4+	1	-1	7	-4	7	-3	0	-1
reste	61	15	6+	10	-1	0	-1	13	-1	1	1+	11	-1	11	-1	0	-1
humaine	35	11	6+	4	-1	0	-1	12	1+	3	2+	3	-2	2	-2	0	-1
combattre	23	9	6+	2	-1	1	1+	6	1+	0	-1	1	-3	4	-1	0	-1
suspicion	22	9	6+	6	1+	0	-1	6	1+	0	-1	1	-2	0	-3	0	-1
progressiste	13	7	6+	1	-1	0	-1	5	1+	0	-1	0	-2	0	-2	0	-1
fermer	8	6	6+	0	-1	0	-1	0	-1	0	-1	1	-1	1	-1	0	-1
fraternité	8	6	6+	1	-1	0	-1	0	-1	0	-1	1	-1	0	-1	0	-1

Partie	1-COM	2-SOC	3-Verts	4-MDC	5-RS	6-RPR	7-UDF	8-FN									
T	187856	11959	37377	3965	44414	2483	48196	38463	992								
f	sp	f	sp	f	sp	f	sp	f	sp								
communistes	29	22	21+	0	-3	0	-1	2	-2	0	-1	2	-2	3	-1	0	-1
droite	159	45	18+	20	-2	11	4+	52	3+	6	2+	11	-9	10	-7	4	2+
nous	1825	210	16+	376	1+	60	4+	270	-21	27	1+	469	1+	40	2+	9	-1
défiance	14	13	15+	1	-1	0	-1	0	-2	0	-1	0	-2	0	-2	0	-1
populisme	12	11	13+	1	-1	0	-1	0	-2	0	-1	0	-2	0	-2	0	-1
etc	3434	323	12+	747	3+	47	-4	770	-2	51	1+	820	-3	65	-2	22	1+
idées	26	14	11+	2	-2	3	2+	3	-1	1	1+	2	-2	1	-2	0	-1
lois	251	45	10+	64	2+	8	1+	33	-5	13	5+	53	-2	33	-3	2	1+
gauche	121	29	10+	17	-2	7	2+	30	1+	1	-1	17	-3	17	-2	3	2+
immigrés	103	25	9+	29	2+	3	1+	15	-2	0	-1	13	-4	16	-1	2	1+
ensemble	75	21	9+	10	-2	2	1+	24	2+	0	-1	10	-3	8	-2	0	-1
travailler	35	15	9+	7	-1	0	-1	4	-2	0	-1	5	-2	4	-1	0	-1

- 25 Il permet, en premier lieu de quantifier les prises de paroles dans un débat. Une partition du corpus affinera cette description par orateur, par parti, par statut institutionnel. On distinguera ainsi, orateur autorisé / interrupteur (à l'Assemblée nationale), journaliste / homme politique lors d'un débat télévisé ou radiophonique.
- 26 Ensuite, le traitement informatique, en s'appliquant à plusieurs corpus, mettra en évidence l'emploi de formes récurrentes à partir de leur fréquence relative.
- 27 Il attirera l'attention du linguiste sur certaines structures syntaxiques propres à un type de locuteurs : les interrupteurs à l'A.N ont d'avantage recours aux négations ou aux interrogations rhétoriques.
- 28 Il fait apparaître dans des corpus « lourds » des ensembles syntagmatiques récurrents (les *segments répétés*) ; l'introduction d'une perspective diachronique dans l'analyse d'un corpus peut révéler l'existence de véritables argumentaires (ex : 8 séances à l'Assemblée nationale).
- 29 Enfin, l'établissement d'« attirances lexicales » (ex : *immigration* + [*clandestine, maîtrisée, galopante...*]) propres à certains locuteurs (ou groupes) aboutit à une typologie des catégorisations.
- 30 Le croisement de ces données mettra en évidence par exemple que les interrupteurs à l'Assemblée Nationale font des phrases courtes, souvent interrogatives (réelles ou rhétoriques), sans prédicat verbal, dans le cas d'insultes ou d'invectives. On pourra ainsi éditer le « texte des interruptions ».
- 31 Si l'on fait l'hypothèse que l'émergence des arguments repose autant sur le jeu de l'interaction que sur les phénomènes de polyphonie et d'intertextualité, il sera possible d'interroger certains indices susceptibles de provoquer cette émergence.
- 32 Les outils de la *deixis* et les instances du discours [Benveniste 1974] je / nous / vous, aujourd'hui / hier / demain, ici / là-bas, etc. constituent les premiers indices de l'orientation énonciative du texte. Viennent ensuite les *modalisations* du discours (en liaison avec le temps et l'aspect). L'opposition partisane droite / gauche se traduira fréquemment par des échanges : « *nous ne pouvons plus.../ vous n'avez pas su...* ». Les diverses manifestations de la *négativité discursive* (négation, dénégation et connotation négative) sont également à prendre en compte. Dans un débat, ce que le camp A « fera » correspond souvent à ce que le camp B « n'a pas voulu faire », ou a été incapable de réaliser. D'où l'importance, de pouvoir orienter la recherche automatique vers les diverses marques formelles de négation, autres que la simple locution adverbiale « ne pas » (préfixes nominaux ou adjectivaux, variantes orales, etc.).
- 33 Nous l'avons indiqué dans le début de l'article, en faisant référence aux travaux de Ducrot, la présence de *connecteurs* spécifiques (*mais, donc, pourtant, parce que* ) , bien qu'elle ne soit pas obligatoire dans un échange, est souvent l'indice de prises de positions argumentatives.
- 34 Enfin, le repérage des traces de *discours rapporté* conduit l'analyste à les interpréter comme autant d'appels au témoignage ou à l'argument d'autorité (pour confirmer ou réfuter une position).
- 35 Comme nous l'avons signalé précédemment, l'apparition de modalisations dans un texte constitue pour nous l'indice (mais non la preuve) d'un début d'argumentation. Nous avons soumis au programme Weblex un certain nombre de termes (verbes modaux mais aussi noms) appartenant à ce champ. Les résultats obtenus sur les *concordances* du mot

« courage » en constituent une bonne illustration. Signalons que le même traitement a été appliqué à d'autres unités à valeur modale : « volonté », « intention », « désir ».

- 36 On remarque que, si pour les composantes de la majorité au pouvoir (Communistes, Socialistes MDC, Verts) le courage relève du champ déontique (*devoir, falloir*), il s'inscrit au contraire sous forme de négations (ou de questions rhétoriques) chez les membres de l'opposition (RPR et UDF), associées à l'aspect accompli ou prospectif (« *vous n'avez pas eu, aurez-vous ?* »). De même, en observant les spécificités lexicales des principaux partis, nous avons remarqué que les premiers verbes conjugués à la deuxième personne du pluriel (*Vous...ez*) sont respectivement « *critiquez* » pour les socialistes, « *videz* » pour le RPR et « *assouplissez* » pour l'UDF. A noter que ces verbes n'apparaissent dans chaque cas que dans un seul parti. Le retour au contexte met en évidence la stratégie de chaque camp. Pour la majorité, il s'agit de soutenir le texte du ministre de l'intérieur, en reprochant aux opposants de critiquer le texte, alors que pour le RPR et l'UDF, il convient de reprocher au présent texte de vider de son contenu le précédent texte (loi Pasqua-Debré), ou d'assouplir les dispositions, les rendant ainsi inapplicables.
- 37 Rappelons l'objectif que nous nous sommes fixé. Il s'agit de réaliser un va-et-vient entre analyse lexicométrique et linguistique de l'énonciation, dans une perspective argumentative. Bien que cette démarche présente quelques difficultés méthodologiques, elle nous semble non seulement possible, mais bien plus, nécessaire face à des corpus de taille importante.
- 38 Qu'il s'agisse de débats (forme dialogale) où les opposants sont face à face (donc identifiables), de discours (monologiques) au sein desquels les arguments de « l'autre » sont actualisés sous forme d'allusions ou de discours rapporté, le recours à la quantification semble indispensable, en amont comme en aval, des analyses qualitatives.
- 39 En amont, parce que la lexicométrie fournit une vision « globale » du corpus, en apportant des informations non prévisibles par une simple analyse énonciative. C'est le cas notamment des segments répétés. En aval, parce qu'elle permet d'aller interroger la totalité du corpus à partir d'hypothèses argumentatives, habituellement limitées à de simples échanges locaux.

---

## BIBLIOGRAPHIE

- Anscombe J.-Cl. (1995). *Théorie des Topoi*. Paris : Kimé.
- Anscombe J.-Cl. & Ducrot O. (1983). *L'argumentation dans la langue*. Liège : Mardaga.
- Benveniste É. (1974). « L'appareil formel de l'énonciation », dans *Problèmes de linguistique générale*. Paris : Gallimard. Tome 2 : 79-88.
- Bonnafous S. & Desmarchelier D. (1999). « Quand les députés coupent le RESEDA », *Mots* 60 : 93-109.
- Breton Ph. (2001<sup>2</sup>). *L'argumentation dans la communication*. Paris : La découverte.

- Desmarchelier D. & Doury M. (coordinateurs) (2001). *L'argumentation dans l'espace public contemporain : le cas du débat sur l'immigration*. Rapport pour l'ARASH, mars 2001. GRIC (CNRS et Lyon II) et ANACOLUT (ENS Fontenay/Saint-Cloud).
- Desmarchelier D. (2005). « La parole est à M. le député... », *Parole & Pouvoir* 2 : 119-134.
- Desmarchelier D. (2005). « La rhétorique dans les débats parlementaires », in P. Marillaud & R. Gauthier *Rhétorique des discours politiques*, 25ème colloque d'Albi. Toulouse : CALS, pp. 283-296.
- Ducrot O. (1993). « Les Topoi dans la 'Théorie de l'argumentation dans la langue' », in C. Plantin (éd.) *Lieux communs, stéréotypes, clichés*. Paris : Kimé, pp. 233-248.
- Habert B. et al. (1997). *Les linguistiques de corpus*. Paris : Colin.
- Heiden S. (1999). « Encodage uniforme et normalisé de corpus. Application à l'étude d'un débat parlementaire », *Mots* 60 : 113-134.
- Lafon P. & Salem A. (1983). « L'inventaire des segments répétés d'un texte », *Mots* 6 : 161-177.
- Perelman Ch. & Olbrechts-Tyteca L. (1970). *Traité de l'argumentation*. Bruxelles : Editions de l'Université de Bruxelles.
- Plantin C. (éd.) (1990). *Lieux communs, topoi, stéréotypes, clichés*. Paris : Kimé.

## NOTES

- 1.. Desmarchelier Dominique et Doury Marianne (Coordinateurs), mars 2001 : *L'argumentation dans l'espace public contemporain le cas du débat sur l'immigration*. GRIC (CNRS et Lyon II) et ANACOLUT (ENS Fontenay/Saint-Cloud), Lyon, 413 pages.
- 2.. L'opération de catégorisation peut introduire une orientation argumentative, mais pas forcément. C'est un point essentiel. Des expressions comme *ces gens, ces personnes...* n'introduisent pas d'orientation, elles sont susceptibles d'entrer dans des prédications et des argumentations cohérentes d'orientations opposées ; elles seront dites neutres. (Rapport ADI).

## RÉSUMÉS

Concilier les approches quantitatives, issues de la lexicométrie, et les analyses sémantiques et argumentatives propres à la linguistique de l'énonciation, telle est l'ambition de cet article. Plus que des résultats, on trouvera ici des pistes de dialogue entre spécialistes des « corpus lourds » et descripteurs des mouvements argumentatifs dans les textes.

From statistical opacity to enunciative transparency.

This article aims at reconciling quantitative approaches derived from lexicometry, with the semantic and argumentative analyses that are specific to enunciative linguistics. Beyond results, suggestions will be found here for a debate between specialists of “heavy corpora” and describers of argumentative movements within texts.

## INDEX

**Mots-clés** : sémantique, argumentation, lexicométrie, catégorisation, concordances, segments répétés

## AUTEUR

**DOMINIQUE DESMARCHELIER**

Université Paris V, CEDITEC EA 3119