

# OUI OU NON À LA CONSTITUTION EUROPÉENNE : L'ÉLOQUENCE DU FORUM<sup>1</sup>

Jessica MANGE, Pascal MARCHAND, André SALEM  
IUT Caen / LERASS, IUT Toulouse 3 / SYLED – CLA2T, Université Paris 3

## SOMMAIRE

1. Caractéristiques du corpus
  2. Comment évolue le vocabulaire des participants tout au long du débat ?
  3. Quels sont les arguments du débat ?
  4. La spécificité et la chronologie des arguments du *oui* et du *non*
- Bibliographie

**Résumé :** *Les analyses médiatiques dominantes sur le référendum pour le traité établissant une constitution pour l'Europe peuvent être résumées de la façon suivante (Voir Le débat, n°136 ; Les cahiers du CEVIPOF, n°42) :*

- *Il s'est agi d'un débat citoyen exceptionnel et passionné ;*
- *Le résultat a été marqué par des préoccupations plus « nationales » qu'européennes, pour deux raisons :*

*\*une campagne monothématique menée par les tenants du « non » et concentrée sur des questions sociales ;*

*\*un « cadrage » médiatique de la campagne plus national qu'européen.*

*Ces analyses dominantes présentent un certain nombre de paradoxes (les médias sont dans la position d'avoir favorisé la parole du « oui », tout en ayant développé une actualité sociale aboutissant au vote « non ») et de zones d'ombre (la plupart des analystes ne se donnent pour objet que de comprendre le vote « non »).*

*Notre position est de nous démarquer de ces analyses dominantes qui, soit en restent aux discours des élites censés refléter l'état « culturel » du débat à un moment donné, et dont on suppose l'application homothétique chez les citoyens, soit infèrent les opinions de ces mêmes citoyens de sondages d'opinion dont les questions – et leur formulation – sont largement guidées par les hypothèses posées a priori des enquêteurs ou des commanditaires.*

*Nous cherchons donc à vérifier, dans la conversation elle-même, la pertinence des analyses qui précèdent. Plusieurs méthodes relevant de l'analyse des données textuelles permettent de retracer la chronologie du débat, puis d'en dégager les arguments contradictoires et leur évolution.*

*Si l'on reproche parfois à l'approche informatique des textes un certain réductionnisme, force est de constater qu'elle permet ici, au contraire, d'aller au-delà des positionnements convenus et d'entrevoir des complexités qui dépassent les apparences commodes.*

## 1. Caractéristiques du corpus

On analyse un corpus rassemblant les contributions de participants à un forum de discussion ouvert sur un site web français (telerama.fr) durant les six semaines qui ont précédé le scrutin du 29 mai 2005. Deux mille deux messages ont été « publiés » sur le forum entre le 22/04/05 et le 22/06/05 (415 pseudonymes différents). Soixante-quinze messages ont également été envoyés pour être publiés dans les numéros précédant le 29 mai 2005.

Les messages recueillis sur le forum sont généralement assez courts, avec une moyenne de 74 mots. Rares sont ceux qui dépassent les 500 mots, mais les plus gros peuvent exceptionnellement atteindre 2000 mots. Le corpus total représente donc 326223 occurrences correspondant à 19092 formes lexicales différentes. Le style dominant, tout en étant une écriture de l'oral, respecte les règles principales de l'écrit canonique : les messages sont majoritairement écrits en minuscules accentuées, avec peu de problèmes de caractères non-reconnus.

L'analyse de ces discours a eu recours à des logiciels relevant de la statistique textuelle<sup>2</sup>. Si le but ultime de l'analyse du corpus est bien d'en produire une lecture interprétative, il s'agit ici de proposer une conduite de détour (voir Salem) dont chaque étape implique des opérations formelles qui, pour être automatisées, n'en impliquent pas moins différents niveaux interprétatifs.

<sup>1</sup> Cette communication présente une synthèse de travaux publiés dans les actes des *Huitièmes Journées Internationales d'Analyse des Données Textuelles* et à paraître dans la revue *Mots*.

<sup>2</sup> Voir, par exemple, Lebart & Salem (1994), Marchand (1998, 2002).

Nous avons proposé de définir deux lignes interprétatives, que nous avons nommées horizontale et verticale, et que résume le tableau ci-dessous (voir Marchand à paraître).

<b>Interprétations</b>	<b>Codage</b>	<b>Analyse</b>	<b>Interprétation</b>
<b>Horizontales</b>	Segmentation, réduction	Concordances, cooccurrences...	Sémantique des formes
<b>Verticales</b>	Partition	Analyse des distances, des correspondances, classifications...	Sémantique des profils (plans factoriels, arbres, classes)

Les interprétations au niveau du codage peuvent être ramenées aux décisions suivantes :

- Quelques abréviations courantes (pb ou pbm, qq ou qlq, bcp, qd, M., Mme) ont été remises à leur forme canonique.
- Les fautes de frappe ou d'orthographe ont été corrigées, et notamment les fautes d'accents (a/à, la/là, traite/traité, prive/privé).
- Les (rares) dysorthographies visiblement volontaires (*pôvre, ceusses, référendome, zommes politiques, môssieur*) ont été conservées.
- Les fautes de grammaire n'ont pas été corrigées. En revanche, pour une partie des résultats suivants, nous avons eu recours à une lemmatisation partielle ramenant à la forme infinitive la plupart des formes verbales. La majorité des fautes d'accord – genre, nombre, participes passés ou confusions entre les formes du futur et du conditionnel, par exemple – ont ainsi été neutralisées.

Sur cette base, trois questions se sont posées à nous, impliquant des partitions et des analyses spécifiques, mais se situant toutes dans une lecture « verticale » du corpus.

## **2. Comment évolue le vocabulaire des participants tout au long du débat ?**

Nous avons partitionné le corpus en six semaines (de la semaine 17 à la semaine 22 du calendrier 2005) :

- du 22 avril au 1<sup>er</sup> mai (161 messages)
- du 2 mai au 8 mai (268 messages)
- du 9 mai au 15 mai (512 messages)
- du 16 mai au 22 mai (464 messages)
- du 22 mai au 29 mai (560 messages)
- du 30 mai au 3 juin (36 messages auxquels s'est ajouté un message du 22 juin) : cette semaine suit la consultation électorale.

Il devient possible alors de construire un tableau croisant, en lignes les 2703 formes lexicales apparaissant plus de dix fois dans le forum, et en colonnes les six semaines analysées. L'analyse factorielle des correspondances (AFC) permet de rendre compte des distances entre les lignes, d'une part, et les colonnes, d'autre part.

Le premier axe de l'AFC restitue intégralement la chronologie du corpus. Cela traduit le fait, souvent constaté dans le cas de l'étude des séries textuelles chronologiques ou simplement de corpus longitudinaux, que le vocabulaire employé par les participants évolue progressivement dans le temps<sup>1</sup>.

Après identification du schéma général de l'évolution progressive du vocabulaire au fil des semaines, il devient possible de caractériser chacune des parties (ou groupe de parties) par le vocabulaire et les séquences qu'elle privilégie et par celui qu'elle évite. Le lexique spécifique de chaque semaine est ainsi mis en relation avec l'actualité politique et médiatique du débat.

Pour résumer les grandes lignes de l'évolution chronologique, il est possible de rechercher les formes lexicales et segments répétés non seulement les plus significatifs de chacune des six périodes, mais également ceux qui traduisent au mieux la chronologie de l'ensemble (Salem 1993). Nous considérerons ainsi la distribution des formes les plus significatives de la première et des deux dernières semaines.

Les unités textuelles les plus employées dans les deux premières semaines se révèlent être les formes qui réfèrent en principe à l'objet du débat : *constitution, traité, traité constitutionnel, l'union,*

<sup>1</sup> Pour des compléments sur les *séries textuelles chronologiques*, on consultera (Salem 1993).

*politique*. La première période s'organise donc sur la question générale de l'objet « *constitution* », de sa description (*notion, articles, règlement*) et de ses conséquences (*concurrence, SIEG et service public*).

Ces formes sont nettement moins utilisées dans les semaines qui correspondent à la fin du débat sur le forum. De même, la distribution des segments répétés les plus longs (supérieurs à 10 formes), qui renvoient à de longues citations, révèle que les internautes y ont souvent eu recours dans les premières semaines du débat mais qu'ils diminuent fortement dans les dernières parties :

... pour éviter que le fonctionnement du marché intérieur ne soit affecté ...  
... les états membres s'efforcent de procéder à la libéralisation des ...  
... qui correspondent au remboursement de certaines servitudes inhérentes à la notion ...  
... toute personne a le droit de travailler et d'exercer une ...  
... un état membre peut être appelé à prendre en cas de ...  
... des citoyens de l'union, au nombre d'un million au ...

Dans ces dernières périodes, on note au contraire une plus grande utilisation des termes : *tu, vous, dimanche, docteurs, patient, remède*, que l'on peut interpréter comme la marque d'un discours plus polémique, plus centré sur l'imminence de l'échéance électorale et dans lequel les enjeux principaux ont déjà été décrits. Le débat s'est déplacé vers l'affrontement de personnes (les « *tenants* », « *partisans* » ou « *défenseurs* » deviennent « *nonistes* » et « *ouistes* », et deviendront « *le camp* » après le vote), qui confinera même à l'insulte dans les quelques messages suivant immédiatement le scrutin : *mépris, mauvaise foi, tricheurs, imposteurs, manipulateurs*.

Ainsi peut-on confirmer la tonalité fortement passionnée du débat à propos de la Constitution européenne, tout en notant la montée en puissance progressive du vocabulaire exprimant cette passion.

### 3. Quels sont les arguments du débat ?

Nous avons cherché à définir statistiquement des classes d'arguments. La méthode implique ici une partition très différente et repose sur la classification hiérarchique descendante (CDH) d'un tableau lexical croisant le lexique (lemmatisé) avec des unités de contexte (Reinert 1990) : partant du corpus intégral, on cherche à définir, de façon itérative, des classes lexicales statistiquement indépendantes. Il s'agit ensuite de voir si ces classes d'arguments peuvent être corrélées avec le « oui » ou le « non ».

Un codage des messages a donc été effectué manuellement : seuls 71 messages (3,5%) n'ont pas pu être clairement attribués ; les autres se répartissent de façon équilibrée en 1000 « pro-oui » (48,1%) et 1006 « pro-non » (48,4%). Rappelons que ce codage n'intervient pas dans l'analyse du corpus elle-même mais en constitue une illustration possible.

Six classes lexicales sont issues de la CDH : deux sont corrélées avec le « oui », deux avec le « non » et deux ne sont corrélées ni avec l'un, ni avec l'autre. Nous donnons ci-dessous des termes significatifs de chacune des classes ainsi que des exemples de réponses significatives (également extraites de façon automatique).

Deux classes lexicales, issues de la même classe-mère, ne sont pas corrélées avec les codes *oui* et *non*. Il s'agit, d'une part de commentaires sur le forum de Télérama (avec les termes : *argument, Télérama, forum, je, lire, message, lecteur, merci, lire, éditorial, médias, débat, journal, bravo*) et d'autre part d'arguments techniques portant sur la procédure du référendum (*traité, question, poser, Nice, vote, renégociation, texte, négociation, approuver, point, référendum, emporter, ratifier, politique, revenir*).

Exemple de message de la classe 4 :

« ...un grand bravo à Télérama pour, d'une part, reconnaître dans ses colonnes que 65 % du **courrier reçu** viennent de **partisans** du non, alors que **Télérama** ... ».

Exemple de message de la classe 2 :

« on ne nous a pas **demandé** notre avis sur les **traités actuels** ? **Erreur /elatorn26/**, le **traité de Maastricht** a été **approuvé** par **référendum**. Quant aux **traités** ultérieurs, ils ont été **ratifiés** par un parlement **français** qui, il me semble, est élu et donc **responsable** devant ses électeurs ».

Ces deux classes, non liées aux opinions, ne présentent pas un intérêt majeur pour notre étude et nous n'en dirons pas davantage.

Deux classes lexicales sont corrélées avec le « non ».

Nous pourrions nommer la première « discours politique et citations du Traité ». Les termes significatifs de cette classe sont en effet : *service, public, concurrent, libre, fausser, entreprise, économie, article, général, intérêt, marché, travail, protection, emploi, privé* ainsi que la modalisation (*non, si, oui, pas, trop, ne, point, déjà, de-manière, pouvoir., donc, certes, cependant, falloir., vouloir., de-toute-facon, pour, surtout, mais, sur, parce-qu', à-travers, en-cas, d'-abord, aujourd'-hui, il-me-semble*).

Exemple de message de la classe 3 :

« ... **article** I.3.3. l'union oeuvre pour le **développement durable fondé** sur une **croissance économique équilibrée** et une stabilité des **prix**, une **économie sociale de marché hautement compétitive**. Les **services publics économiques** restent **soumis** aux **règles** de la **concurrence**, III.166, et à la limitation des **aides publiques**, III.167 ... »

« ... en-plus, l' **article** III.167 **interdit** toute **aide publique** aux **entreprises publiques**: sont incompatibles avec le **marché intérieur** les **aides accordées** par les états membres sous quelques **formes** que ce soit qui **fausse** ou qui **menacent** de **fausser** la **concurrence** en **favorisant** certaines **entreprises** ou certaines productions ».

La deuxième classe lexicale corrélée au « non » fait également largement appel au Traité lui-même, mais davantage en référence aux « cadres institutionnels de l'Europe » : *parlement, union, commission, membre, Etat, conseil, charte, compétent, droit, européen, initiative, OTAN* ainsi que des adjectifs et adverbes (*fondamental, législatif, exécutif, européen, national, coopératif, respectif, militaire, universel, étranger, antérieur, international, soumis, significatif, nécessaire, parlementaire, fédéral, progressif, défini, nouveau, précédent, direct*).

Exemple de message de la classe 6 :

« ...les **dispositions** de la **présente charte** qui **contiennent** des principes peuvent être **mises en oeuvre** par des **actes législatifs** et **exécutifs** pris par les **institutions, organes** et organismes de l'**union** et par des **actes** des **états** membres ... »

Ces deux classes lexicales, issues de la même classe-mère, ce qui confirme leur cohérence statistique, renvoient donc à un débat argumenté sur la base du contenu même du traité et de son analyse. Les internautes proposent ici une lecture interprétative des conséquences du texte.

Enfin, deux classes lexicales, encore une fois issues de la même classe-mère, sont corrélées avec le « oui ».

Nous avons proposé de nommer la première « discours philosophiques (concepts et valeurs) ». Les termes significatifs de cette classe sont en effet : *monde, nous, histoire, vivre, mondial, rêve, Europe, pays, guerre, solidarité, pauvre, petit, vie, chômage, modèle, jeune(s)* ainsi que des marqueurs de lieux et de pays.

Exemple de message de la classe 1 :

« ...nous n'avons jamais vraiment voulu directement **aider** les **pays** plus **pauvres** que nous à se développer. Où en serait (*sic*) l'**Espagne** et le **Portugal** aujourd'hui sans l'**Europe** ? L'**Europe** est le **seul exemple récent** de **réelle solidarité** entre les **pays** les plus **riches** et les plus pauvres... ».

La deuxième classe lexicale corrélée avec le « oui » est celle des acteurs politiques : *gauche, le Pen, droite, Fabius, Villiers, Chirac, extrême, parti, PS, Jospin, présidentielle, Besancenot, Buffet, Hollande, tour*. On y trouve également des adjectifs et adverbes, mais d'une tout autre tonalité que dans la classe 3 : *souverainiste, populaire, électif, trotskiste, populiste, xénophobe, frileux, électoral, partisan, prochain, généreux, majoritairement, second, majoritaire, quasiment*.

Exemple de message de la classe 5 :

« ...mais quand dans le **camp** du non on **retrouve** le **Pen**, les **trotskistes**, de **Villiers**, **Boutin**, **Fabius** et **Mégret** on se dit que la **démagogie** est plus importante dans le **camp** du non... ».

Si les deux classes corrélées avec le « non » présentaient une certaine cohérence, les deux classes corrélées au « oui » semblent relever de registres différents. Quoi de commun, en effet, entre l'idéal européen et les acteurs de la politique française ?

Le fait que les arguments du *Oui* et du *Non* soient distingués dans une classification automatique calculée indépendamment de ce codage nous autorise à approfondir leur spécificité.

#### 4. La spécificité et la chronologie des arguments du *oui* et du *non*

La recherche des formes et segments répétés spécifiques du *Oui* et du *Non* confirme la CDH :

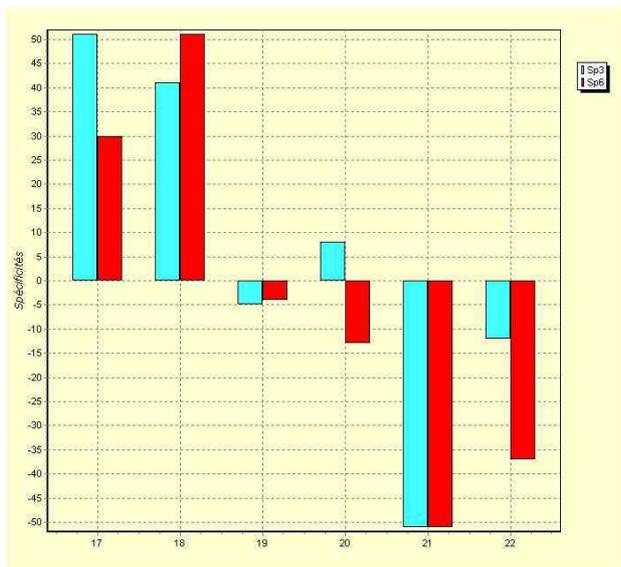
- du côté du *Oui* : *vous, gauche, Fabius, monde, tenants du non, extrême, Villiers, la France ...*

- du côté du *Non* : *constitution, partisans du oui, je voterai non, médias, euros, démocratie, article(s) ...*

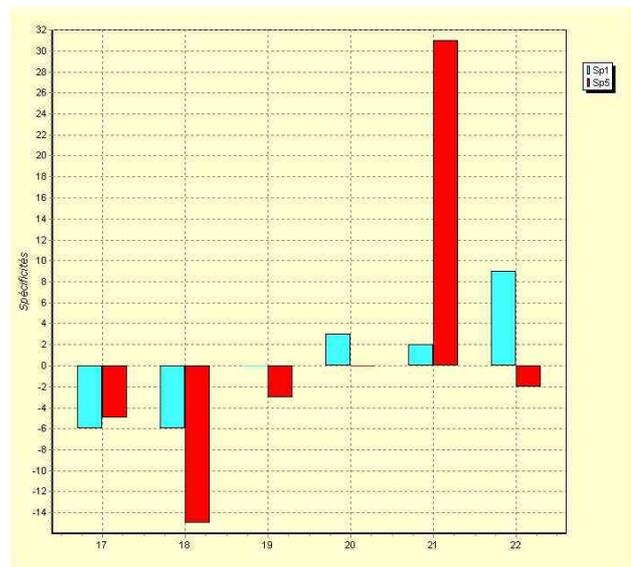
Nous avons voulu savoir si ces lexiques spécifiques apparaissent aux mêmes moments dans la chronologie du débat. À partir des vocabulaires dégagés comme caractéristiques pour les deux groupes de participants, nous avons construit deux unités textuelles que l'on appelle *types généralisés* (ou *tgen*) *spécifiques*<sup>1</sup>, dont nous avons calculé la ventilation dans les six semaines qui composent le corpus : le *tgen SP-oui* rassemble toutes les occurrences du corpus qui correspondent à une forme spécifique pour le groupe de participants qui se prononcent pour le *oui*. Le *tgen SP-non* rassemble celles des occurrences qui correspondent au contraire à des formes spécifiques pour le groupe des participants qui se prononcent pour le *non*<sup>2</sup>.

On observe alors que chacun des deux camps connaît un moment d'expression plus marquée des termes qui font sa spécificité (la semaine 18 pour le camp du *non*, la semaine 21 pour celui du *oui*). Autre fait remarquable, le camp du *non* délaisse dans cette même semaine 21 le vocabulaire spécifique mis en place lors de la semaine 18. Ces constats sont confirmés par une nouvelle partition du corpus en douze parties (6 semaines x 2 opinions), c'est-à-dire que chaque partie rassemble dorénavant des messages produits au cours d'une même semaine par des internautes d'opinions identiques.

Ces calculs reposent sur les spécificités du « oui » et du « non » prises globalement. Or, nous avons dégagé, plus haut, quatre classes distinctives du « oui » et du « non ». Nous avons donc cherché la distribution chronologique des *Tgens* de ces classes d'arguments, que rapportent les graphiques ci-dessous.



Distribution chronologique des formes spécifiques des classes 3 et 6 (« non »)



Distribution chronologique des formes spécifiques des classes 1 et 5 (« oui »)

Il apparaît nettement que les deux classes d'arguments corrélées avec le « non » suivent une évolution comparable et conforme à la distribution globale des arguments du « non ». C'est un peu différent pour le « oui » : la classe 1 (des valeurs européennes) se trouve bien spécifique du « oui », et bien en fin de débat, mais de façon moins marquée que la classe 6 (des acteurs de la politique française), qui connaît une spécificité considérable la semaine même du scrutin.

Nos analyses montrent que le débat à propos du TCE ne se réduit pas à des formules simplistes selon lesquelles le « oui » serait européen et le « non » serait national.

Ainsi, les arguments du « non » semblent s'organiser sur le traité et font volontiers référence au texte lui-même, qui fait l'objet d'interprétations idéologiques, soit sur l'organisation des institutions

<sup>1</sup> Le Tgen, « ensemble d'occurrences sélectionnées parmi les occurrences du texte », généralise la notion de type unité dont on peut recenser les occurrences dans le texte. Sur les types généralisés, on consultera (Lamalle & Salem 2002).

<sup>2</sup> Nous avons sélectionné, pour cette expérience les unités qui ont un indice de spécificité supérieur à 4 (i.e. celles pour lesquelles la méthode des spécificités renvoie à une probabilité inférieure à 1/10000<sup>e</sup>).

européennes, soit sur ses conséquences sociales et économiques (chômage, services publics, etc.). Par ailleurs, ces arguments s'installent très tôt dans le débat et ont tendance à faiblir par la suite.

Les arguments du « oui » reposent davantage sur une identification aux valeurs historiques fondatrices de l'union européenne et un recours aux symboles qui unissent les peuples et les citoyens. Le résultat le plus inattendu est certainement que c'est également du côté du « oui » que l'on trouve les références à la politique française et à ses acteurs. Mais on observe également que ces arguments apparaissent très tard dans le débat : les partisans du « Oui » peinent à installer leur argumentation en début de confrontation et ce n'est que dans la dernière semaine qu'elle prend toute sa dimension.

Le recours à la statistique distributionnelle nous a permis de « faire parler » un corpus textuel, dans une acception que nous voudrions sans doute rattacher à une rhétorique plus psychanalytique que policière... Il n'en demeure pas moins que le rapport des sciences humaines et sociales à la matérialité textuelle, focalisé sur la dimension « verticale », se distingue nettement de l'approche linguistique ou littéraire, plus préoccupée de considérations « horizontales ». S'agissant de discours persuasifs, si ce n'est propagandistes, il ne s'agit pas de « comprendre » un corpus dans ses dimensions intra voire même intertextuelles, de rechercher une structure, d'identifier une esthétique, d'attribuer un genre ou de questionner l'une ou l'autre de ces notions. Le corpus en sciences humaines et sociales n'est qu'un accès possible à son contexte sociohistorique de production et n'a de réelle valeur qu'en tant qu'il nous renseigne sur ce contexte. Quant à l'analyse automatique, elle n'a d'intérêt qu'en tant qu'elle peut mettre en évidence des effets de sens qui, pour des raisons multiples, n'émergent pas d'autres analyses.

## **BIBLIOGRAPHIE**

*Le débat*, n°136 : La France et le choc du 29 mai, septembre-octobre 2005.

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

*Les cahiers du CEVIPOF*, n°42 : Le référendum de ratification du Traité constitutionnel européen : comprendre le « Non » français, juillet 2005.

MANGE, J., MARCHAND, P. à paraître. Oui ou non à la Constitution européenne : l'éloquence du forum, *Mots*.

MANGE, J., MARCHAND, P., SALEM, A. 2006. Débats sur la toile, in *Actes des Huitièmes Journées Internationales d'Analyse des Données Textuelles* (Besançon, 19-21 avril), Presses Universitaires de Franche-Comté, coll. Les cahiers de la MSH Ledoux, série « archive, bases, corpus », pp. 667-676.

MARCHAND, P. 1998. *L'Analyse du Discours Assistée par Ordinateur. Concepts, méthodes, outils*, Paris, Armand Colin, coll. U, série psychologie.

MARCHAND, P. 2004. *Psychologie sociale des médias*, Rennes, Presses Universitaires de Rennes.

MARCHAND, P. 2005. Le grand oral de Dominique de Villepin, *Bulletin de Méthodologie Sociologique*, n°87, pp. 80-85.

MARCHAND, P. à paraître. Concepts, méthodes, outils, in C. Gauzente & D. Peyrat-Guillard (éds.), *Analyse et fouille de données textuelles : applications en gestion*, Editions Management et Société, coll. « questions de société ».

REINERT, M. 1990. « ALCESTE, une méthodologie d'analyse des données textuelles et une application : Aurélia de Gérard de Nerval », *Bulletin de Méthodologie Sociologique*, N°26.

SALEM, A. 1993. *Méthodes de la statistique textuelle*, Thèse, Université de la Sorbonne Nouvelle, Paris 3.