



HAL
open science

Les thèmes du discours. Du concept à la méthode

Mahé Ben Hamed, Damon Mayaffre

► **To cite this version:**

Mahé Ben Hamed, Damon Mayaffre. Les thèmes du discours. Du concept à la méthode. Mots: les langages du politique, 2015, Les thématiques du discours politique ; exploration méthodologique, 108, pp.5-13. hal-02466786

HAL Id: hal-02466786

<https://hal.science/hal-02466786>

Submitted on 4 Feb 2020

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.



Mots. Les langages du politique

108 | 2015

Thèmes et thématiques dans le discours politique

Les thèmes du discours. Du concept à la méthode

Mahé Ben Hamed et Damon Mayaffre



Édition électronique

URL : <http://journals.openedition.org/mots/21975>

ISSN : 1960-6001

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 6 octobre 2015

Pagination : 5-13

ISBN : 978-2-84788-727-3

ISSN : 0243-6450

Référence électronique

Mahé Ben Hamed et Damon Mayaffre, « Les thèmes du discours. Du concept à la méthode », *Mots. Les langages du politique* [En ligne], 108 | 2015, mis en ligne le 06 octobre 2015, consulté le 19 avril 2019. URL : <http://journals.openedition.org/mots/21975>

Les thèmes du discours. Du concept à la méthode

Le repérage, la description ou l'établissement des thèmes (*topics*) des discours sont un objectif majeur pour les analystes qui étudient le langage politique. Debout dans la cité, le discours politique véhicule de manière explicite des messages, traite ou sous-entend des sujets, trahit des positions qu'il convient dès lors de révéler ou de déconstruire. Les thèmes (étymologiquement *ce qui est posé dans/par le discours*) sont non seulement un élément de comparaison utile entre les locuteurs dans le cadre de l'affrontement discursif constitutif du politique (de quoi parle chaque protagoniste ?), mais aussi la condition de l'appréciation de prises de parole dont la vocation pragmatique est évidente : dire le monde pour le changer ou pour le conserver, prendre position sur les sujets de société et contre l'adversaire, thématiser et anathémiser.

Description du contenu des programmes électoraux des partis, par exemple, dans le fonctionnement démocratique de nos sociétés contemporaines, établissement de l'agenda setting des organes de presse dans la médiacratie actuelle, mise à jour de la cohérence sémantique des discours de grands leaders historiques, analyse des topoï, de la doxa ou de l'idéologie que véhiculent les textes : l'approche thématique nous paraît à la fois l'enjeu le plus évident de l'analyse du discours et son enjeu, au-delà des examens naïfs, le plus complexe.

D'un point de vue méthodologique en effet, le repérage scientifique des thèmes d'un corpus d'étude, ici politique, reste un problème non résolu ; et cela en dépit d'une littérature importante sur le sujet et plus encore de l'usage surabondant, mais sans autres précisions, du mot « thème », dans une infinité d'études déclinées : « le thème de la guerre chez Hobbes et Rousseau », « le thème de l'amour dans l'œuvre de Shakespeare », « le thème de la mort chez Roger Martin du Gard », etc.

Les études littéraires y ont certes apporté des solutions érudités et artistiques mais ce dossier ne pourra s'en revendiquer ; tout juste s'en inspirer. La « critique thématique », que figure par excellence Jean-Pierre Richard,

représente en effet un univers riche et profond hors de la modestie de notre horizon. Le thème n'y est plus défini seulement comme l'à-propos des discours mais comme l'objet « immanent et transcendant » du texte (Brémond, 1985, p. 422), qui « exprime la relation affective d'un sujet au monde sensible » (Collot, 1988, p. 81) ; et l'approche poétique du texte le dispute souvent à l'approche psychanalytique du locuteur (Jean-Pierre Richard, 1961 et 1974). La froide efficacité des discours politiques que l'on entend ci-devant étudier offre sans doute un objet trop prosaïque pour la finesse de ce type d'analyses qui embrassent l'esthétique générale des discours bien au-delà de leur simple sens.

Les études linguistiques balisent quant à elles plus concrètement le programme mais apparaissent dans l'état de l'art insatisfaisantes pour ce présent numéro, en réduisant l'empan d'observation du discours à l'énoncé et en accordant à l'organisation syntaxique, qui détermine la polarité du message, une place déterminante. De fait, en linguistique, le thème n'est presque jamais perçu en lui-même pour lui-même mais toujours dans son opposition, certes intéressante mais non nécessaire à notre propos, au *rhème* (ou *prédicat*). Il en va bizarrement jusqu'aux deux dictionnaires français de référence pour l'analyse du discours (Détrie, Siblot et Vérine, 2001, p. 356-358 ; Charaudeau et Maingueneau, 2002, p. 572-573) dans lesquels, de manière extraordinaire, ne figure pas d'entrée indépendante pour le thème, et où la paire thème/rhème, dans les deux cas, nous empêche d'accéder au singleton¹. Cette position des linguistes, y compris donc ceux engagés en analyse de discours, nous paraît une rançon excessive au fonctionnalisme et à une linguistique phrastique dont l'analyse du discours se proposerait à contre-emploi d'étendre le principe : une vaine grammaire du texte comme il existe une grammaire effective de la phrase. Pour le monde anglo-saxon, nous renvoyons ainsi particulièrement au travail fondateur de Van Dijk dans lequel les *discourse topics* apparaissent finalement comme une extension des *sentence topics* (Van Dijk, 1977) et aux auteurs de la *Critical Discourse Analysis* (CDA) actuelle tels Chilton et Schäffner (2002).

Entre autres conséquences, cette position fonctionnaliste qui s'origine à l'échelle internationale dans l'École pragoise fait la part belle au *processus propositionnel*, plus qu'à la chose dans son ensemble et, par là, au linéaire, au syntagmatique ou à la progression, là où notre approche du thème, comme

1. Nous ne saurions trop insister sur cette étrange réalité. Sur les quelque 1 000 entrées confondues des deux dictionnaires, « thème/rhème » est une des très rares entrées doubles (nous ne parlons pas du système de renvois habituel dans les dictionnaires), comme si la notion de thème, fort utilisée à la fois dans l'usage scientifique et dans l'usage courant, ne méritait pas une entrée indépendante et ne pouvait pas être définie en elle-même. Du reste, les premières phrases de l'entrée double « thème/rhème » dans le dictionnaire de Charaudeau et Maingueneau ne laisse aucun doute sur l'objectif des auteurs : traiter du couple dans sa dualité oppositionnelle et non traiter du thème en tant que tel : « Thème/Rhème : Cette distinction apparaît dans les travaux... » (Charaudeau, Maingueneau, 2002, p. 572).

nous le verrons, renvoie aussi au paradigmatique. Cette position enfin adopte le plus souvent une focale micro ou locale, au cœur serré des transformations (passage du thème 1 au rhème 1 ; établissement dès lors du rhème 1 en thème 2 puis passage au rhème 2 ; établissement à son tour du rhème 2 en thème 3 puis passage au rhème 3, etc.), là où l'on ne renonce pas à avoir aussi une vision macro ou globale des corpus, dans leur ensemble, soumis à l'expertise².

Ce dossier semble donc avoir des préoccupations différentes, et s'il ne saurait prétendre apporter de réponse définitive, il entend contribuer simplement au débat méthodologique qu'engagent ces questions générales et élémentaires : de quoi parlent nos discours ? *What's the discourse about?* Que nous disent-ils ? Et, corollairement, car le fond est dans la forme comme la fin dans les moyens : de quelles manières ?

Si l'on souhaite articuler le thème à la *matérialité discursive* du corpus, convenons minimalement qu'un thème se construit – donc se repère – par la saillance et l'agrégation d'unités linguistiques (classiquement des mots) qui s'organisent sémantiquement. Le thème, c'est une idée suivie et développée dans le corpus qui prend corps par la chair linguistique que les locuteurs/interlocuteurs mobilisent et structurent à dessein dans le discours.

Par-là, deux impératifs se dessinent, que l'on aurait tort de considérer comme deux étapes successives et indépendantes : le repérage du matériau linguistique, particulièrement lexical, agent ou acteur des thèmes ; le repérage de l'organisation sémantique de ce matériau brut ou primaire. Car, précisons-le en préalable, le thème a un besoin *sine qua non* de mots mais ne s'y réduit pas, comme le laisse penser à tort la mode des nuages (*word clouds*) qui fleurissent partout dans le ciel de l'Internet. Le thème dans nos corpus est une composition linguistique et il est fait matériellement de mots, mais mots et thèmes ont une relation qui n'est ni automatique ni bijective : tous les lexèmes ne sont évidemment pas des thèmes, et chaque thème pourra être lexicalisé de manière variée, plurielle, *ad hoc*. En d'autres termes, l'analyse du discours, nécessairement assistée par ordinateur aujourd'hui, a toujours affiché une autre prétention que l'analyse de contenu, et cette haute prétention, c'est-à-dire au fond cette conscience linguistique, s'énonce sans doute le plus évidemment dans l'idée de thème et de thématique du discours.

Un thème, c'est un « contenu » graphique ou lexical, certes, mais qui s'anime (au sens fort) sémantiquement ou prend sens à la faveur de *parcours de lecture* ; l'expression est ici directement reprise à la théorie de François Rastier

2. Ainsi, dans cette littérature, semble se dégager un distinguo terminologique que nous ne reprendrons pas : le « thème » serait *ce qui est posé localement dans la phrase* alors que la « thématique » serait *ce qui est posé globalement dans le texte*. Nous voulons ici tout au contraire réfléchir le thème au niveau du texte et même du corpus, et utilisons le mot thématique dans son acception habituelle par exemple du Robert pour désigner ce qui est « relatif au thème ».

dont il faut souligner l'épaisseur tant épistémologique que pratique (Rastier, 1987, 2001, 2011) et auquel nous renvoyons pour ce dossier dans sa livraison de 1995 : *L'analyse thématique des données textuelles* (Rastier éd., 1995).

Chevillés à la matière textuelle, les thèmes, pas plus que le sens, ne sont *déjà-là* ; déjà constitués dans le texte. Ils ne nous sont pas livrés ou révélés, sans autres subtiles précisions sémantiques : ils s'élaborent au sein de parcours interprétatifs complexes dans lesquels l'analyste, à la suite du locuteur, articule les briques lexicales attestées en corpus pour architecturer un sens toujours à composer, discuter, interpréter. Cette posture herméneutique fondamentale que Jacques Guilhaumou a depuis longtemps réclamée pour l'analyse du discours (Guilhaumou, 1993, 2004) ou que François Rastier théorise pour la linguistique du texte nous semble nécessaire à l'appréhension de la notion de thème dans toute son acuité : « Comme toutes les unités sémantiques, un thème est un construit, non un donné : en cela, la thématique dépend de conditions herméneutiques. » (Rastier éd., 1995, p. 232)

Dès lors, notre questionnement se précise tout en devenant vertigineux : quelles méthodes pour encadrer, au sein du corpus, les parcours interprétatifs qui performant ou construisent les thèmes du discours ? Même ramenée à l'intérieur du corpus – car nous écartons à ce stade les considérations extra-textuelles –, l'objectivation de parcours interprétatifs ne saurait évidemment faire l'unanimité, pas plus que ce dossier ne peut ici prétendre épuiser tous les points de vue.

Adressé à tous, l'appel à contribution a trouvé un écho majoritaire dans la communauté de lexicométrie/textométrie/logométrie, comme si les méthodes d'objectivation thématique aujourd'hui ne pouvaient plus faire l'économie de l'outillage informatique et, au sein du numérique, l'économie de l'approche quantitative. Le repérage systématique des occurrences et des récurrences que permet l'ordinateur, la recension systématique des présences/absences et des itérations lexicales qu'autorisent les logiciels semblent ainsi désormais scientifiquement bien établis dans le mouvement élargi des humanités numériques, de la linguistique de corpus et de l'analyse de discours. Plus loin – nous y reviendrons – les rapports statistiques et distributifs entre deux ou plusieurs mots dans le corpus, autrement appelés, de manière générique, *calcul des cooccurrences*, semblent pour la plupart des contributeurs une bonne préfiguration de la relation sémantique entre ces mots, c'est-à-dire un bon indice de l'agrégation ou cimentation des mots en thèmes.

Il est vrai que la lexicométrie politique dont la revue *Mots* fut pionnière s'est largement attachée à la question depuis les origines. Au-delà des dictionnaires de fréquences et du calcul du vocabulaire spécifique des textes qui permettent de rendre compte de manière brute (dictionnaire) ou subtile (spécificités) du contenu lexical des corpus, il s'est immédiatement agi de calculer et

de représenter les *networks* de termes, les échos sémantiques entre mots, les cooccurrences : en un mot inavoué alors, d'étudier ces constellations de lexies qui, ensemble, finissent par faire sens, finissent par faire *thème*. Il convient en effet ici de relire l'éditorial fondateur de la revue où, en peu de lignes, Maurice Tournier énonce quelques règles de la lexicométrie politique naissante : « La seconde règle de prudence consiste à n'en pas rester au vocable isolé. Le mot : cet acteur de sens que seuls d'autres mots peuvent actionner dans un sens. Tout, dans l'énonciation, est séquence, réseau, co-occurrences, équilibre ou fuites entre formes, scansion d'emplois affrontées. » (Tournier, 1980, p. 6) Et l'on renvoie, pour la littérature saint-clousienne, à quelques articles fondamentaux sur la cooccurrence et les lexicogrammes de cooccurrences (Lafon, 1981 ; Tournier, 1987 ; Heiden, Lafon, 1998, etc.)

Sur ces principes, la lexicométrie littéraire a aussi développé, explicitement cette fois-ci, la question thématique, et l'ouvrage déjà cité (Rastier éd., 1995), par exemple, met en œuvre plusieurs approches thématiques appuyées sur la cooccurrence pour en pointer le véritable enjeu : passer de la coprésence statistique et matérielle (au niveau du signifiant) à la corrélation sémantique ou thématique (au niveau du signifié) ; en d'autres termes, pour reprendre le titre de l'ouvrage plus récent de Margareta Kastberg sur l'œuvre de Jean-Marie Le Clézio, pour passer « des mots aux thèmes » (Kastberg, 2006). Ainsi, de manière significative, une des plus fameuses fonctions lexicométriques d'un des logiciels libres sur le marché scientifique, Hyperbase, s'intitule depuis les années quatre-vingt-dix, THEME : elle calcule et met en forme visuellement les constellations lexicales construites autour d'un mot-pôle pour identifier des mondes sémantiques ou univers thématiques (Brunet, à paraître). Et dans le domaine de la stylogométrie, le travail de Jean-Marie Viprey sur les cooccurrences généralisées et la dynamique du vocabulaire dans *Les Fleurs du mal* reste une référence (Viprey, 1997).

De fait, si, comme l'affirment Oswald Ducrot et Jean-Marie Schaeffer, le repérage des thèmes est d'abord « une perspective paradigmatique » sur le texte qui consiste à rechercher des systèmes « de ressemblance (et donc d'opposition) entre les unités souvent très distantes » dans le corpus (Ducrot, Schaeffer, 1995, p. 639), alors l'approche réticulaire (c'est-à-dire en réseaux) que permettent l'analyse de données textuelles et l'hypertextualité nous paraît bien indiquée. De plus, si le thème (à l'image de l'*isotopie* qui apparaît comme une notion très proche) est, comme semblent l'indiquer les définitions les plus anciennes et les mieux établies, « itération » (Roland Barthes, cité par Collot, 1988, p. 79), « répétition » (Jean-Pierre Richard, cité par Collot, 1988, p. 79), « récurrence » (Rastier éd., 1995, p. 224), c'est-à-dire redondance au sein du corpus d'un même trait sémantique³, alors l'approche systématique, sérielle,

3. Ainsi François Rastier pose comme premier jalon : « Nous nommerons *thème* une structure stable de traits sémantiques (ou *sèmes*), récurrente dans un corpus, et susceptible de lexicalisations diverses. » (Rastier éd., 1995, p. 224).

quantitative, statistique n'est pas la moins légitime pour assurer la démarche. Du reste, dans l'univers technique de la linguistique computationnelle internationale ou du *Natural Language Processing* (TALN), les algorithmes aussi puissants que nombreux mis en place pour la détection automatique des thèmes (*topic detection* ou *topic tracking*) dans de très grands ensembles s'appuient le plus souvent sur des régularités quantifiables (Dascălu, 2014 ; Charu, ChengXiang, 2012 ; Wartena, Brussee, 2008, Allan, 2002, etc.).

Quoi qu'il en soit, Magali Guaresi propose ici dans son article, « Les thèmes dans le discours électoral des candidat-e-s à la députation sous la Cinquième République. Perspective de genre (1958-2007) », un protocole méthodologique ambitieux fondé sur le traitement cooccurentiel du corpus. Partant du postulat fort que la structure sémantique fondamentale d'un texte peut être modélisée par une matrice mathématique puissante et élémentaire croisant entre eux les 300 substantifs les plus utilisés, elle traite cette matrice cooccurentielle 300 mots x 300 mots grâce à trois méthodes successives et trois logiciels : l'analyse factorielle des correspondances (Hyperbase), la classification hiérarchique descendante (Iramuteq) et les *Netwoks* (Gephi). De ces traitements riches et complémentaires émergent, de manière endogène, les thèmes d'un corpus politique d'envergure consacré, dans la lignée des travaux d'Antoine Prost (1974), aux professions de foi des député(e)s français(es) durant cinquante ans (1958-2007) ; si la perspective est genrée (thèmes politiques de femmes *versus* thèmes politiques d'hommes), l'ambition est plus générale dans la description historique des débats électoraux sous la Cinquième République. L'intérêt de l'approche est d'être émergentiste ou *bottom up* (on pourrait dire aussi sémasiologique), là où les thèmes sont habituellement pressentis a priori comme si l'idée pouvait précéder le verbe. C'est bien par le repérage systématique des signes et plus concrètement de leur signifiant, et par la mise au jour de leur distribution statistique en co(n)texte (c'est-à-dire au sein du paragraphe dans lequel on regarde systématiquement leur fréquentation cooccurentielle) que Magali Guaresi construit ce que l'on est en droit, dès lors, d'appeler les thèmes du discours.

C'est ce même postulat cooccurentiel qui préside au logiciel Iramuteq et à la méthode Alceste (Reinert, 1983) dont Pascal Marchand et Pierre Ratinaud, d'une part, et Emmanuel Marty, d'autre part, font ici une utilisation experte.

Emmanuel Marty, dans son article « Les élections municipales au prisme de la presse quotidienne régionale. Des cadres médiatiques aux thématiques politiques », montre comment la méthode permet de découper le corpus en segments de texte (c'est-à-dire unité cotextuelle ou *fenêtre* paramétrable ici de 40 mots) pour y observer les relations cooccurentielles qui matérialisent des relations sémantiques. Se dégagent alors du corpus, grâce à la classification hiérarchique descendante, ce que Max Reinert nomme « mondes lexicaux » (Reinert, 1983), ce que nous pourrions appeler « thème », ce que l'auteur qualifie, sous

certaines conditions, d'« espaces conceptuels ». Emmanuel Marty confronte dès lors les espaces conceptuels du corpus aux *cadres médiatiques* du discours de presse durant la campagne des municipales de mars 2014 et conclut que la presse quotidienne régionale est bien sûr assujettie aux pouvoirs locaux, mais est aussi le lieu d'une expression citoyenne marquée par la proximité.

L'article de Pierre Ratinaud et Pascal Marchand, dans « À la hache, à la machette et au scalpel. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014) », nous emmène vers d'autres horizons puisqu'il traite de manière exemplaire un corpus sans beaucoup d'égal en lexicométrie : 140 millions de mots soit l'ensemble des débats de l'Assemblée nationale française de 1998 à 2014 ; et ce bond quantitatif au niveau du corpus implique des réajustements méthodologiques dont le lecteur pourra prendre connaissance mais qui ne touchent pas à l'essentiel de la méthode Alceste. Pierre Ratinaud et Pascal Marchand définissent en introduction les thématiques du discours « comme un ensemble de formes pleines cotextuelles liées entre elles par leur objet et leur contexte ». Par là, ils réfutent ou renversent une démarche onomasiologique qui irait de l'idée aux mots. C'est donc encore de l'endogénéité des thèmes en corpus, à base lexicale, dont il est question dans une démarche dont le mérite est d'être non supervisée. Si les thèmes sont des construits, cette construction est *bottom up* : du matériel lexical du corpus vers l'interprétation sémantique, via l'outil informatique et la classification hiérarchique descendante qu'implémente aujourd'hui le logiciel libre Iramuteq.

À côté de ces trois articles qui recourent aux traitements les plus sophistiqués de la statistique textuelle, l'article de Grégory Piet, Régis Dandoy et Jeroen Joly, « Comprendre les programmes électoraux. Comparaison des méthodes d'encodage manuel et automatique », permettra au lecteur, dans un utile contrepoint, de réfléchir sur les avantages d'un encodage manuel *versus* un encodage automatique des thématiques. Dans le premier cas, ce sont directement les analystes qui encodent les thèmes phrase par phrase, attribuant à celle-ci ou à celle-là une des 249 thématiques consignées dans un *code book* nord-américain de référence ; entreprise périlleuse et considérable qui demanda dix-huit mois de travail. Dans le second cas, c'est le logiciel Prospero qui a la charge de catégoriser les thèmes du discours sur la base d'ontologies prédéfinies, d'indices lexicaux, voire d'un typage (ici aussi manuel) de l'utilisateur. Dans les deux cas, le parcours interprétatif est donc fortement contraint et contrôlé par l'intervention du chercheur, soit directement, soit via un logiciel d'essence qualitative.

Enfin, l'article de Fabienne Baidier, « Évaluation de la *présidentiabilité*. Sémantique lexicale à l'interface de la linguistique de corpus et de l'analyse du discours », revient sur l'approche cooccurrence dans une étude focalisée. Pour tester le thème de la *présidentiabilité* dans les débats et articles de presse autour de la primaire socialiste en 2011 et de la campagne présidentielle

de François Hollande en 2012, Fabienne Baider convoque les « collocats » du mot-pôle (*présidentiable* ou *présidentiability*) pour en définir les contours polémiques selon que l'on traite de François Hollande ou de Martine Aubry. Au cours de l'analyse, l'approche quantitative cède le pas à une approche plus qualitative par l'intermédiaire de concordances et de nombreux extraits, comme pour rappeler que la recherche des thèmes, si elle ne saurait aujourd'hui plus faire l'économie de l'outil statistique et du traitement informatique, ne saurait non plus négliger le retour au texte et l'acte de lecture.

Références

- ALLAN James, 2002, *Introduction to Topic Detection and Tracking*, New York, Springer US.
- BRUNET Étienne, à paraître (2015), *Au bout du compte. Questions linguistiques*, B. Pinchemin éd., préface de F. Rastier, Paris, Champion.
- BRÉMOND Claude, 1985, « Concept et thème », *Poétique*, n° 64, p. 415-424.
- CHARAUDEAU Patrick, MAINGUENEAU Dominique éd., 2002, *Dictionnaire d'analyse du discours*, Paris, Le Seuil.
- CHARU Aggarwal, CHENGXIAN Zhai éd., 2012, *Mining Text Data*, Londres, Springer.
- CHILTON Paul, SCHÄFFNER Christina, 2002, « Introduction: themes and principles in the analysis of political discourse », *Politics as Text and Talk. Analytical Approaches to Political Discourse*, P. Chilton, C. Schäffner éd., Amsterdam, Benjamins, p. 1-41.
- COLLOT Michel, 1988, « Le thème selon la critique littéraire », *Communications*, n° 47, p. 79-91.
- CUSIMANO Christophe, 2012, *La sémantique contemporaine. Du sème au thème*, Paris, PUPS.
- DASCĂLU Mihai, 2014, *Analyzing Discourse and Text Complexity for Learning and Collaborating. A Cognitive Approach Based on Natural Language Processing*, Springer International Publishing.
- DÉTRIE Catherine, SIBLOT Paul, VÉRINE Bertrand éd., 2001, *Termes et concepts pour l'analyse du discours. Une approche praxématique*, Paris, Champion.
- DUKROT Oswald, SCHAEFFER Jean-Marie, 1995 [1972], *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Le Seuil.
- GUILHAUMOU Jacques, 1993, « À propos de l'analyse de discours. Les historiens et le "tournant linguistique" », *Langage et société*, n° 65, p. 5-38.
- 2004, « Où va l'analyse de discours? Autour de la notion de formation discursive », *Texto!*, en ligne : http://www.revue-texto.net/Inedits/Guilhaumou_AD.html (consulté le 2 février 2015).
- HEIDEN Serge, Lafon Pierre, 1998, « Cooccurrences. La CFDT de 1973 à 1992 », *Des mots en liberté*, Paris, ENS Éditions, p. 65-83.
- KASTBERG Margareta, 2006, *L'écriture de J.M.G. Le Clézio. Des mots aux thèmes*, Paris, Champion.
- LAFON Pierre, 1981, « Analyse de lexicométrie et recherche de cooccurrences », *Mots*, n° 3, p. 95-148.

- MAYAFFRE Damon, 2014, « Plaidoyer en faveur de l'analyse de données co(n)textuelles. Parcours cooccurrentiels dans le discours présidentiel français (1958-2014) », *JADT 2014, Proceedings of the 12th International Conference on Textual Data Statistical Analysis*, E. Née, M. Valette, J.-M. Daube, S. Fleury éd., Paris, Inalco - Sorbonne nouvelle, p. 15-32, en ligne : <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/01-JADT2014.pdf>.
- PROST Antoine, 1974, *Vocabulaire des proclamations électorales de 1881, 1885, 1889*, Paris, PUF.
- RASTIER François, 1987, *Sémantique interprétative*, Paris, PUF.
- 2001, *Arts et sciences du texte*, Paris, PUF.
- 2011, *La mesure et le grain. Sémantique de corpus*, Paris, Champion.
- RASTIER François éd., 1995, *L'analyse thématique des données textuelles*, Paris, Didier.
- REINERT Max, 1983, « Une méthode de classification descendante hiérarchique. Application à l'analyse lexicale par contexte », *Les Cahiers de l'analyse des données*, vol. VIII (2), p. 187-198.
- RICHARD Jean-Pierre, 1961, *L'univers imaginaire de Mallarmé*, Paris, Le Seuil
- 1974, *Proust et le monde sensible*, Paris, Le Seuil.
- TOURNIER Maurice, 1980, « En souvenir de Lagado », *Mots*, n° 1, p. 5-9.
- 1987, « Cooccurrences autour de "travail" », *Mots*, n° 14, p. 89-123.
- VAN DIJK Teun, 1977, *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*, Londres, Longman Linguistic Library.
- VIPREY Jean-Marie, 1997, *Dynamique du vocabulaire des Fleurs du mal*, Paris, Champion.
- WARTENA Christian, BRUSSEE Rogier, 2008, « Topic detection by clustering keywords », *Database and Expert Systems Application, 2008. DEXA '08. 19th International Workshop on*, p. 54-58.