

Vocabulaire de la gestion : usage et sémantique

Laure Pairet

CRIISEA - Amiens- 80027 Cedex - France

Résumé

Cette étude sur la terminologie d'acteurs en gestion des entreprises met en évidence des dissemblances sémantiques en fonction du rôle respectif de ceux-ci. Le corpus étudié ici est composé de trois sources : entretiens semi directifs, revue scientifique, communiqués de presse. Les méthodes de statistique textuelle, basées sur l'extraction des formes signifiantes, leurs contextes et les spécificités ont été utilisées pour les comparaisons. Les outils informatiques permettent de révéler ces variations et variétés d'usages.

Abstract

This study on management actors' terminology brings to light semantic disparities according to their respective role. The corpus studied here consists of three sources: semi directive interviews, scientific journal, press releases. The statistical textual methods of analysis, based on the extraction of the forms, their contexts, their specificities and factor analysis were used for the comparisons. The computing tools allow to see these variations and varieties of the language usage.

Keywords: statistical analysis of textual data, actors' terminology, management science

1. Introduction

La gestion, au sens académique, est une discipline récente en France, issue de la loi sur la formation professionnelle de 1968, « l'introuvable conciliation de la science et de la pratique » (Godelier, 2004).

Ce travail tend à mettre en exergue des usages du langage d'acteurs de la gestion des entreprises et organisations. Des mots-clés communs existent-ils? Sont-ils polysémiques? Trois sources jouent un rôle différent : des entreprises, organisations opérationnelles ; la revue scientifique RFG ¹, un des vecteurs académiques ; le MEDEF, une institution.

Ces trois catégories d'acteurs participent à la création du vocabulaire de la gestion : dirigeants d'entreprise à l'échelon local ; auteurs d'articles dans la discipline, à un niveau d'inter – médiation ; rédacteurs des communiqués de presse, au plan national.

Trois vues sont proposées, ici : la première d'ordre macroscopique, où les trois discours forment le corpus d'ensemble ; la deuxième, longitudinale porte sur la comparaison des deux dont le genre est proche (communiqués et résumés) ; la troisième, les corpus sont perçus séparément (Morin, 1994).

¹ Revue Française de Gestion.

2. Type de corpus en Sciences de Gestion

Divers travaux ont montré la pertinence d'une typologie basée sur une variable structurelle, quelque soit le genre du corpus en Sciences humaines ou de la Société (Erlos, 2006).

La variable structurelle est la source du discours dans cette étude.

En Sciences de Gestion, où l'acteur se situe dans l'organisation, la dimension est individuelle et collective. Les spécificités des sous-corpus sont souvent liées à la fonction de l'acteur, à la taille de l'entreprise.

2.1. Les corpus étudiés

Les corpus étudiés ici sont :

- les résumés des articles de la Revue Française de Gestion, recueillis sur le site via CAIRN, une quarantaine de numéros, environ 350 KO,
- les communiqués de presse du MEDEF, accessibles sur le site éponyme, une centaine d'enregistrements, environ 300 KO,

Sur la période 2002-2007

- les réponses à questions ouvertes de responsables d'entreprises régionales, enregistrées selon un protocole défini dans le cadre d'une recherche régionale sur les réseaux de coopération en 2002-2003 ; une vingtaine de questionnaires, environ 1,2 MO.

Les noms sont porteurs de sens dans le discours, ici les communiqués. Les verbes d'actions, les pronoms sont révélateurs de comportement dans les interviews. Dans les textes scientifiques, les substantifs sont les supports sémantiques (Bergeron and Labbé, 2004 ; Burner, 2004).

Le sens de termes signifiants diffère du singulier au pluriel : relation/relations, client/clients, partenariat/partenariats. Ces constats d'ordre général se retrouvent ici.

Nous avons combiné plusieurs méthodes pour extraire, à partir des usages, des vocables représentatifs du domaine : sélectionner les signifiants, s'intéresser à leurs relations sémantiques, leurs coefficients d'implication réciproque, aux spécificités, synthétiser par carte factorielle.

2.2. Méthodologie

La méthodologie s'appuie sur la comparaison de corpus, typiques du domaine associée à l'exploration des concordances.

La sélection de termes signifiants a été effectuée sur la base d'un seuil de fréquence, de cooccurrences sur un contexte délimité, significatifs statistiquement, puis une lemmatisation partielle pour tenter de définir des catégories. Enfin, une représentation visuelle à l'aide d'une analyse de données textuelles (ADT), sur le corpus de base puis sur ce corpus réduit montrant une relative stabilité (Lebart and Salem, 1994).

Nous avons donc utilisé l'idée de contexte minimal ou unité minimale de cooccurrences de Mayaffre (2008) afin de limiter l'ambiguïté et le paramétrage quelque peu arbitraire.

L'étude porte sur le corpus global puis longitudinal, enfin local par source, réduit aux formes signifiantes de fréquence supérieure à 5 puis aux 400 premières, soit 0,2% du corpus initial, afin d'extraire le vocabulaire signifiant à l'aide du contexte et des spécificités.

Le cheminement a été le suivant : dictionnaire du corpus d'origine, mise en évidence de certaines formes signifiantes, concordance de ces formes signifiantes, mesure d'association de cooccurrences, spécificités, ADT, enfin la reprise du processus sur un corpus réduit (longitudinal et local).

2.2.1. Vocabulaire

La préparation des bases requiert un formatage pour l'utilisation avec les logiciels adaptés ².

Dans un premier temps, le corpus n'a pas été lemmatisé, il est traité à l'état brut.

L'indice lexicométrique classe les formes modales au début. Elles sont porteuses des thèmes principaux (Lebart and Salem, 1994). Elles permettront d'identifier les appartenances au domaine.

Les formes caractéristiques d'une partie, celles dont la fréquence dans la partition est supérieure à la fréquence modale théorique de la loi hypergéométrique, révèlent les thématiques principales de la partie. Elles sont obtenues par le calcul des spécificités.

L'ambiguïté sémantique éventuelle des formes sera restreinte par l'ajout des segments répétés.

Les formes présentant un nombre d'occurrences faible ne sont pas toujours à négliger. Elles indiquent les sujets émergents. Elles peuvent être associées à une logique de veille concurrentielle ou commerciale, inter-entreprises, ou intra-entreprise voire une innovation organisationnelle. La concordance des formes les précisera.

2.2.2. Sémantique

Les cooccurrences indiquent les associations multiples de formes à proximité. Une mesure d'attraction lexicale, développée par W. Martinez, donne le nombre de liens, la spécificité, le nombre d'unités de contexte (Martinez, 2005).

Divers indicateurs d'association existent, par exemple le coefficient d'implication réciproque.

Ces cooccurrences dans le corpus forment des catégories homogènes de mots. Or, si des mots identiques sont proches quelques fois dans le texte, il est fort probable qu'ils décrivent un contexte proche. Si leurs fréquences sont élevées, elles représentent les thèmes centraux.

L'apport d'information sur le contenu découle des cooccurrences significantes. Quels critères fixer afin de limiter les cooccurrences significantes? Arbitraires?

Une méthode ³ proposée par J.P. Courtial permet une caractérisation de ces catégories, en fonction de leur centralité et leur densité. La première est somme des liens des mots les composant avec les autres mots externes ; la seconde, la valeur moyenne des liens entre mots du thème. Après avoir obtenu les descripteurs, issus de l'index lexicométrique, identifié les associations fortes (contextes), par classement des valeurs d'association ⁴, il construit un diagramme « stratégique ». L'origine est la valeur médiane des axes, en abscisse la centralité, en ordonnée la densité. Nous avons repris cette méthode pour illustrer les thèmes du MEDEF.

2.2.3. Typologie

L'analyse factorielle lexicale conduit à établir un tableau de contingence entre le vocabulaire et une variable signalétique afin de visualiser des spécificités éventuelles dans un plan.

² Lexico 3, COOCS, Wordmapper 7.01 et Spad 5.6.

³ Utilisée par Wordmapper.

⁴ Association entre deux mots-clés: produit des probabilités d'avoir un mot-clé quand on a l'autre ; 0 ces mots jamais ensemble ; 1 ces 2 mots toujours ensemble ; 0,25 implication symétrique ; association dissymétrique 0,125. La probabilité d'erreur sur ces relations est mesurée par des tests du khi deux.

Cette vision est complémentaire de la précédente. Une seconde analyse sert à confirmer, sur un corpus réduit aux formes modales les plus significantes statistiquement d'une part, avec un objectif de catégorisation d'autre part.

Une analyse factorielle met en évidence deux plans portés par des optiques différentes, qui pourraient être des attributs de l'activité principale. La stabilité des résultats est à vérifier, par tirages aléatoires des parties du corpus et itération de la procédure, ici avec une réduction du vocabulaire sur d'autres critères (Lebart, 2009 ; Leroux, 2009).

3. Vue d'ensemble

Dans les corpus de genre discours, communiqués de presse, et articles, le pourcentage de noms communs dans les signifiants est une caractéristique. Ici (Tab. 1), plus de 85% des signifiants sont des noms.

Ici, en moyenne 0,4% de la taille et 4% du vocabulaire ont été conservés. Les quatre cents premières formes obtenues dans le dictionnaire (sans les mots outils), sachant que la forme modale signifiante « entreprise » arrive cent dixième dans l'index brut, ont été retenues.

	<i>ENTREPRISE</i>	<i>MEDEF</i>	<i>RFG</i>	<i>ENSEMBLE</i>
Taille	221416	35725	54027	311168
Vocabulaire	12346	5776	8237	26359
Richesse %	5,6	16,2	15,3	8,5
Hapax %	2,6	7,9	8,9	3,1
Fréquence max. de signifiants	395	287	168	400

Tableau 1 : Corpus, caractéristiques

3.1. Dictionnaire

L'index lexicométrique global, dont seuls les substantifs ont été reportés, donne les formes et fréquences suivantes (sans les mots-outils articles, auxiliaires, pronoms) :

Entreprises 612, *auteur* 473, *client* 417, *temps* 413, *titre* 411, *clients* 413, *entreprise* 359, *gens* 358, *MEDEF* 323, *travail* 315, *produit* 293, *prix* 288, *qualité* 270, *problème* 270, *marché* 255, *commercial* 238, *France* 234, *article* 233, *gestion* 230, *produits* 225, *confiance* 153, *management* 151, *relation* 148, *coopération* 137, *compétences* 135, *sociétés* 121, *organisations* 110, *pouvoir* 109.

Des formes comme *auteur*, *titre*, *MEDEF*, *gens*, *article* sont génératrices de bruit, ici ; de plus comme la partie issue directement des entreprises est prépondérante, elle se retrouvera au début de l'index global mieux représentée que les autres.

Parmi les termes de fréquence modale utilisés dans les trois sous corpus, les thèmes principaux, après élimination des mots outils sont :

Gestion avec 168 occurrences, puis *Entreprise(s)* avec une sur représentation dans la RFG ; *Client(s)*, forme la plus fréquente et spécificité de la modalité ENTREPRISE puis *temps*, *produit*, *prix* ; *Entreprises*, dont le nombre d'occurrences est le plus élevé, 287 alors que *Travail* est une caractéristique du MEDEF, comme *France* et *marché*.

Les préoccupations sont diverses bien que le sujet initial, l'entreprise soit identique (Fig.1).

Ensuite, une lemmatisation partielle a été effectuée, ramenant au masculin singulier les féminins et pluriels. Les formes de fréquence supérieure ou égale à 6, sans les mots-outils ont été conservées comme signifiants : ENTREPRISE : 3735, MEDEF : 644, RFG : 291.

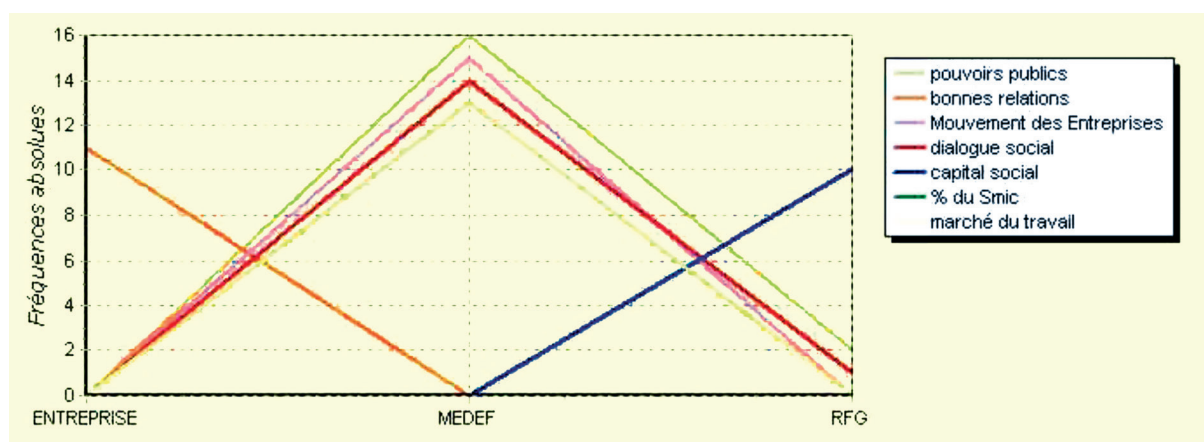


Figure 1 : Exemples de segments répétés, fréquences par source (logiciel Lexico)

Ci-dessous le début du dictionnaire (Tab. 2) :

ENTREPRISE	MEDEF	RFG
Client	Entreprise	Entreprise
Temps	Emploi	Gestion
Entreprise	Salarié	Management
Produit	Négociation	Stratégie
Commercial	Politique	Organisation

Tableau 2 : Lemmatisation partielle

Cette lemmatisation partielle complète la vision précédente (Tab. 2). Les entreprises sont pragmatiques, elles gèrent au quotidien ; la RFG, approfondit le sujet ; le MEDEF pilote avec les partenaires sociaux.

Ces formes modales communes sont des catégories du domaine. Les segments répétés délimitent le sens des formes. Tab. 3 propose les segments de fréquence modale. Il n'expose pas les mêmes formes, mais indique des sous-thèmes.

ENTREPRISE	MEDEF	RFG
cahier des charges	partenaires sociaux	création de valeur
chiffre d'affaire	entreprises françaises	sciences de gestion
sous-traitance (t(s))	mouvement des entreprises	dans les organisations
donneurs d'ordre	marchés publics	management stratégique
bonnes relations	pouvoirs publics	relations inter organisationnelles
matière première	formation professionnelle	ressources humaines

Tableau 3 : Segments de fréquence modale

Alors que les groupes de mots les plus spécifiques par parties, reflètent les particularités.

Ce repérage puis classement, par ordre décroissant de fréquence et de spécificité pour chacune des parties, des segments répétés confirment les différences (Fig. 2).

La terminologie des entreprises, de terrain, relève de la sphère opérationnelle ; celle de la RFG, évoque un aspect analytique, informationnel ; celle du MEDEF, national, montre le niveau décisionnel.

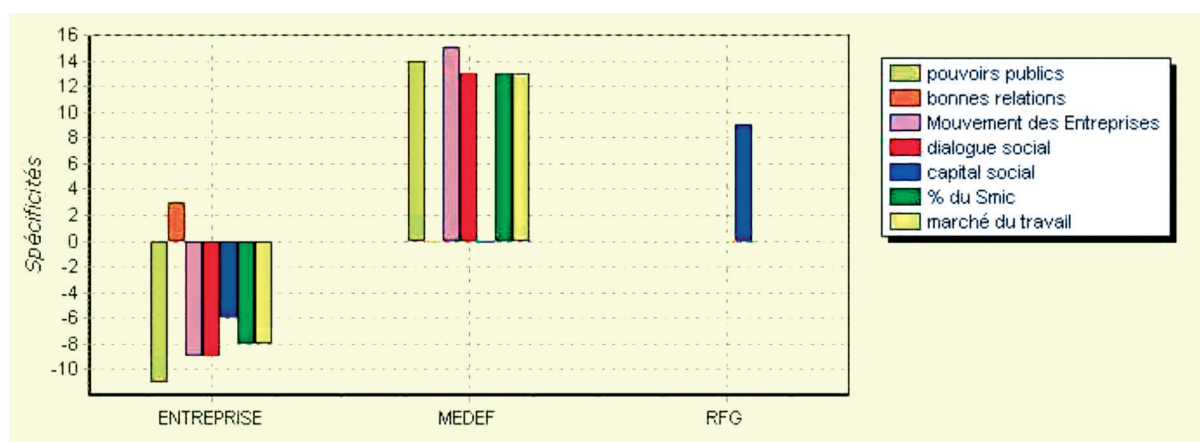


Figure 2 : Segments spécifiques par source (logiciel Lexico)

3.2. Les cooccurrences

Les poly-cooccurrences, ont été extraites parmi les 10 formes avant et après la forme pôle. Les cooccurrences par parties (Tab. 4), 4 formes avant et après chaque forme pôle ont été regardées.

FORME	spécificités	liens niveau 1 niveau 2
entreprises	9	consommateurs, communauté, développer, fixé, création, esprit
	5	organisationnel, propice
négociations	9	MEDEF, Présidente, européennes,
	5	commerciales
organisation	9	Management
	5	formes, réseau, virtuelles
Travail	14	Coût
	8	augmenter, emploi

Tableau 4 : Cooccurrences

Ce classement est effectué par ordre décroissant de spécificité d'association puis par niveau.

Les cooccurrences par partie révèlent les classes homogènes (Tab. 5 : Coocs, seuil 5%).

Les contextes limités à quatre formes exposent un usage par partie dissemblable. En effet, le *marché* n'apparaît pas, par exemple dans le discours des dirigeants de terrain, à proximité des entreprises, le *client* semble le remplacer. La perspective de développement des entreprises apparaît pour le MEDEF sans coopération, alors que pour la RFG *développement*, *relation*, *marché* sont explicitement cités avec les entreprises.

La gestion disparaît des entreprises pour le terrain, alors que ce concept est primordial dans la RFG, apparent pour le MEDEF qui s'intéresserait plus à la gestion des compétences. Les relations sont avec les clients pour les entreprises, de confiance pour la revue.

Le calcul d'un coefficient d'implication réciproque restitue ces divergences sémantiques.

Cette différence pointe la séparation entre le pouvoir national, décisionnel, du MEDEF et le pouvoir local, opérationnel des entreprises. Les formes communes relèvent du domaine conceptuel, du projet. Le vecteur informationnel est joué par la revue scientifique.

FORME POLE	ASSOCIEES	ENTREP.	RFG	MEDEF
entreprises	Temps	1	0	2
	Client	2	1	2
	Marché	0	2	2
	Relation	1	2	0
	Développement	0	2	7
	Coopération	5	1	0
gestion	Entreprises	0	4	2
	Recherche	0	5	1
	Management	0	2	1
	Compétences	0	2	4
relations	Confiance	1	5	1
	Client	1	3	0
	Clients	3	0	1
	Entreprises	1	2	0
confiance	Client	3	1	0
	Relations	1	5	2

Tableau 5 : Cooccurrences par partie

3.3. Dissemblance

Une analyse factorielle permet de visualiser cette dissemblance.

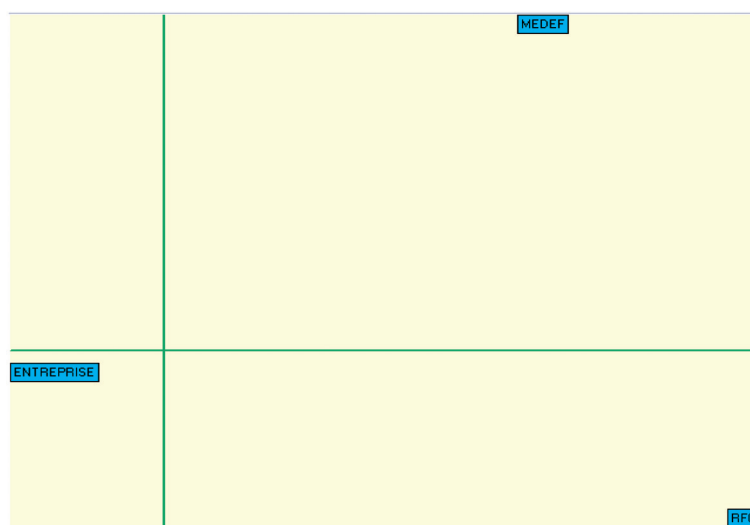


Figure 3 : Analyse factorielle des correspondances (logiciel Lexico)

Les demi-plans séparés par l'axe horizontal mettent en évidence un lien ENTREPRISE-RFG isolant MEDEF. Les demi-plans séparés par l'axe vertical mettent en évidence un lien RFG-MEDEF isolant ENTREPRISE. La RFG apparaît comme un vecteur du vocabulaire (Fig. 3).

L'hétérogénéité des cooccurrences, les plus fréquentes, des formes modales significantes communes renvoient à la variété des usages des catégories.

4. Vue longitudinale

Elle concerne les résumés de la RFG et les communiqués de presse du MEDEF sur la période 2002-2007.

L'index lexicométrique réduit aux substantifs signifiants, donne comme formes modales principales :

152 développement, 139 travail, 137 management, 132 recherche, 112 France, 110 organisations, 94 marché, 92 processus, 91 compétences, 86 salariés, 85 PME, 84 relations, 83 négociation, 82 organisation, 82 stratégie, 79 étude, 78 formation, 77 stratégies, 75 économique. (En enlevant les formes : MEDEF, article).

Les cooccurrences par partie des termes signifiants communs sont assez différentes entre RFG et MEDEF. Ces termes sont rarement employés dans le même contexte, comme nous l'avons vu ci-dessus (3.2). Le calcul d'un coefficient d'implication réciproque, rapport du nombre de cooccurrences au produit de leur fréquence, permet de classer les cooccurrences simultanées de deux ou plusieurs formes. Il a été réalisé suivi d'un classement par partie qui confirme un usage souvent polysémique. Exemples : le MEDEF « marchés publics » « PME », la RFG « marchés financiers » pas de cooccurrences communes aux parties. La « compétence » y partage un cooccurent « mobilisation » le coefficient est 0,61% pour RFG, 0,50% pour MEDEF. Mais pour le MEDEF, le plus haut coefficient 0,89% est « gestion de » suivi à 0,73% de « reconnaissance », alors que la RFG exprime des compétences organisationnelles avec un coefficient d'implication réciproque à 0,64%.

La dichotomie thématique est manifeste sur le premier plan factoriel (Fig. 4).

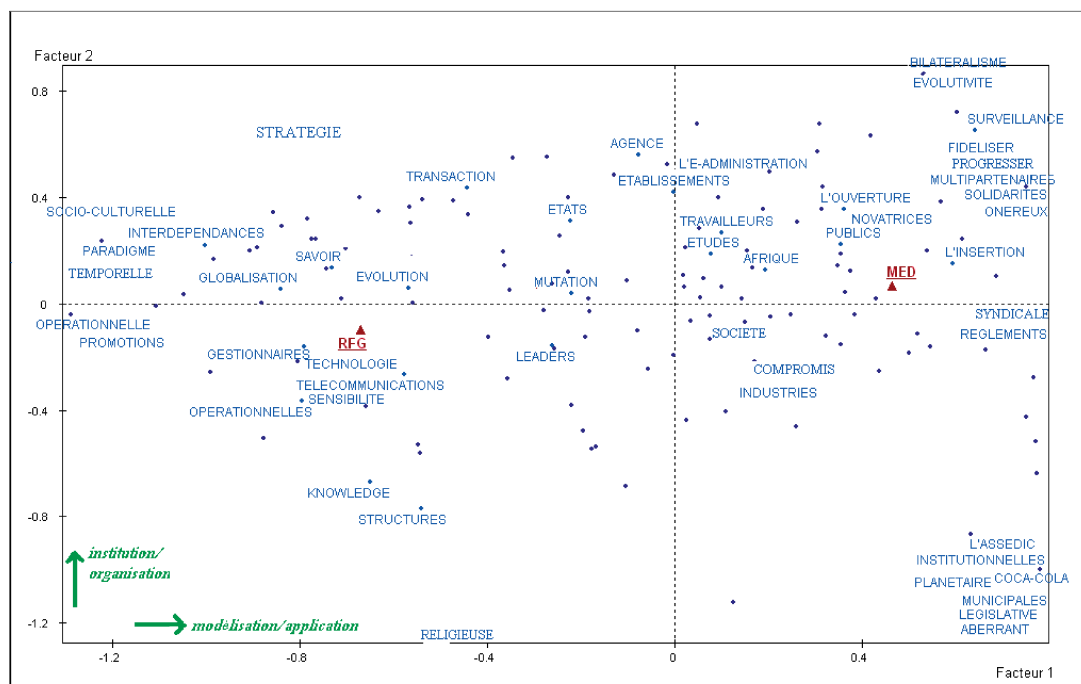


Figure 4 : Théorie/Pratique (logiciel Spad)

S'opposent : sur l'axe horizontal, les concepts aux faits ; sur le vertical, les normes au pilotage. L'axe trois sépare les partenaires extérieurs. La comparaison est approfondie en effectuant la différence des matrices de cooccurrences, un des deux protagonistes, le MEDEF, est présenté. D'après la méthode de JP Courtial, les clusters à forte densité et forte centralité sont centraux.

Le pilotage vers une croissance durable apparaît en haut à droite. Des thèmes émergent sur l'engagement, la volonté individuelle, les liens avec les institutions en bas à gauche (Fig. 5)

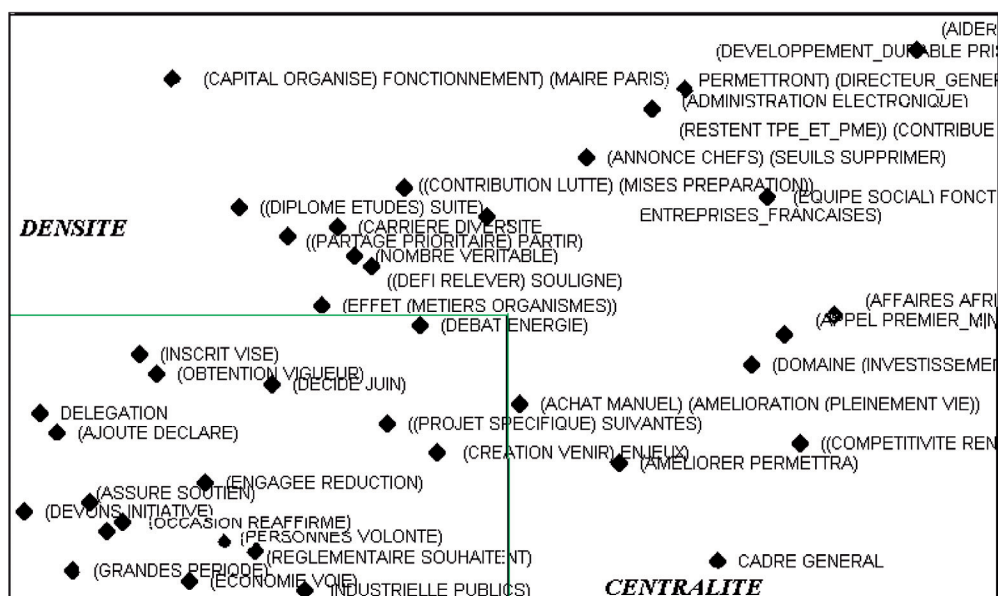


Figure 5 : Thèmes du MEDEF (logiciel Wordmapper)

Sur le moyen terme, persistent les variations d'usages entre RFG et MEDEF.

5. Vues locales

Les trois vues intra-source qui suivent dégagent un profil particulier à chacune.

5.1. Centres d'intérêt associés pour le MEDEF

La carte factorielle ci-dessous (Fig. 6) a été obtenue en croisant le sujet du communiqué avec les termes du corpus. Certains thèmes sont relativement liés, d'autres s'opposent :

finance et recherche ; social et emploi ; politique et droit, pour les plus présents. Les *PME* et *l'international* ; la *recherche* et la *finance*. Les catégories atypiques et opposées : *jeunes*, *protection sociale*.

L'entrepreneuriat face à la politique sociale (Fig. 6 bas/haut). L'axe horizontal est porté par les entreprises, l'axe vertical par le social. La carte longitudinale (non représentée ici) montre que les thèmes n'évoluent pas de manière linéaire dans le temps.

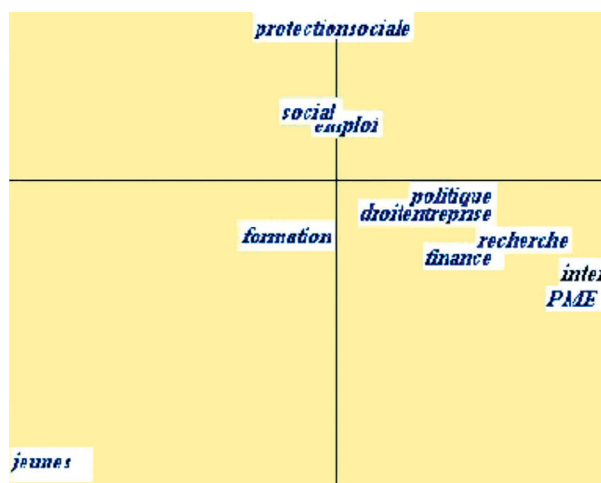


Figure 6 : Carte factorielle MEDEF-thèmes (logiciel Lexico)

5.2. Évolution de la RFG

Chaque numéro, variable structurelle représentée ici, propose une thématique (Fig. 7)

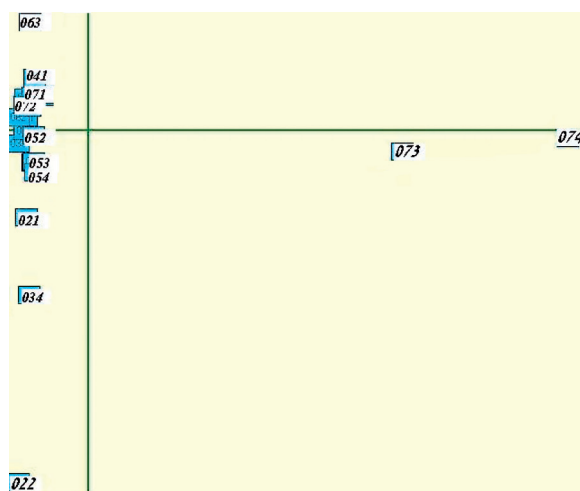


Figure 7 : Carte factorielle RFG (logiciel Lexico)

Le premier numéro de la période 022 concernait un auteur, J. March, ce qui justifie sa position qui le différencie des autres (superposés proches). En fin de période, une thématique atypique surgit, relative aux TIC : valeurs d'usage des technologies de la communication (073), communautés de pratiques et changement stratégique (074).

5.3. Le contexte régional

Le contexte est celui d'une recherche pluridisciplinaire ⁵, dans le cadre d'un travail sur les réseaux d'entreprises en Picardie, plus spécialement la coopération. L'analyse factorielle binaire croisement du discours avec la variable « taille » atteste d'un lien entre la taille de l'entreprise et les termes des dirigeants.

La coopération, n'est pas perçue de manière identique par les dirigeants. Le terme est polysémique et des mots comme partenariat, *confiance*, *relation*, *collaboration* sont ici synonymes. La proximité de termes : *Contrat-partenariat* ; *confiance-relation* ; *contact-parlé* dans le plan factoriel montre leur lien respectif.

Les contextes ont permis de voir que la gestion se situe dans un contexte plutôt interne, la confiance semble associée au regroupement, beaucoup plus relationnelle dans un contexte non formalisé. La collaboration est associée avec les prix.

La taille apparaît comme un facteur discriminant dans le discours face à la coopération.

- petite : moyens de stimuler son activité, en direction du donneur d'ordre
- moyenne : respect des normes du métier, en direction de la concurrence
- grande : image diffusée à la clientèle, en direction de la clientèle.

Le premier axe horizontal pourrait être le marché : du partenariat informel à la convention contractuelle, le deuxième vertical la production : de la reconnaissance de la marque au savoir-faire du métier (Fig. 8).

⁵ Réseaux d'acteurs, connaissances et capacités en Picardie ; CRIISEA, LARIA, CCI, DRIRE.

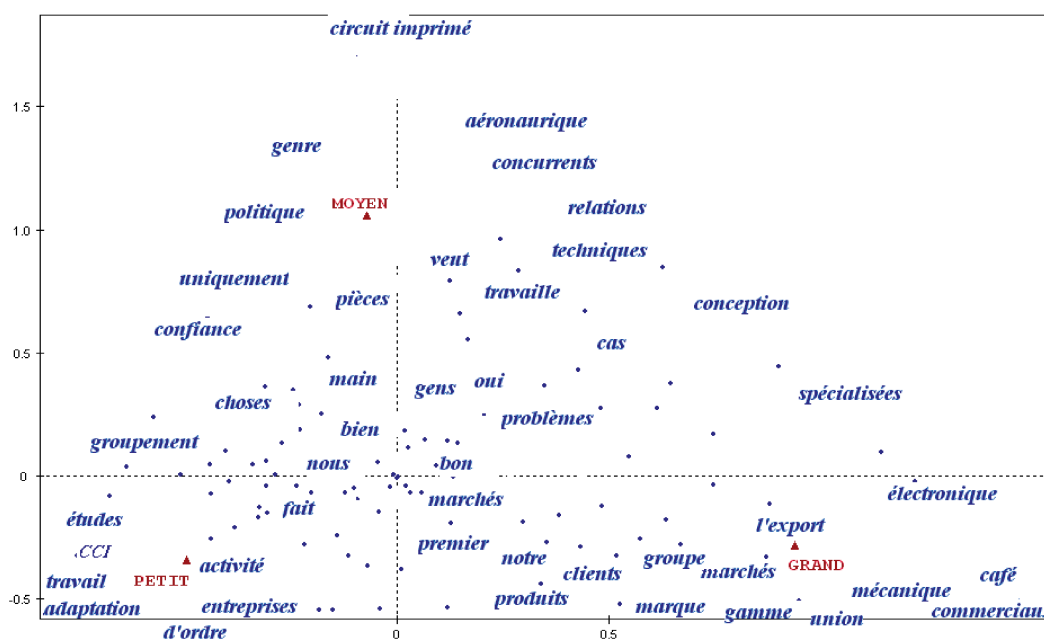


Figure 8 : Carte factorielle ENTREPRISE-effectif (Logiciel SPAD)

Une confirmation de cette analyse a été effectuée par tirages aléatoires du vocabulaire. La structure est conservée, le vocabulaire prédominant de manière relative pour chaque modalité de la taille de l'entreprise est synonyme. La terminologie du dirigeant est fonction de la dimension de l'entreprise. La typologie implique deux catégories : partenariat et marché.

Le vocabulaire retenu comme catégorie induit des attributs, ses cooccurrences. En partant d'une production de vocabulaire diversifiée, des catégories communes ont été mises en évidence par filtrage contextuel puis complétées par des vues locales afin de trouver des attributs complémentaires communs mais la variété de ceux-ci exprime la variation d'usage.

6. Conclusion

La complémentarité de différentes méthodes de statistiques textuelles nous a offert la possibilité d'extraire des connaissances terminologiques spécifiques au domaine de la gestion. Les critères, autres que de significativité, restent relativement arbitraires pour restreindre le vocabulaire signifiant. Les dissemblances sémantiques des termes par partie impliquent la difficulté à élaborer un thésaurus du domaine, voire à concilier pratique, théorie et pouvoir.

De ces trois types d'acteurs en gestion, vecteurs de vocabulaire, qui détient le plus grand pouvoir de légitimité?

La forme dans son contexte se révèle un nouveau concept multidimensionnel issu du web. Le type d'acteur du domaine permettrait-il de définir d'autres descripteurs pour les mots des Sciences de Gestion?

Références

- Bergeron J.G. and Labbé D. (2004). Analyser les entretiens sociologiques. In *Actes JADT*.
 Bouroche J.M. and Saporta G. (1992). *L'analyse des Données*. Que sais-je? Paris : PUF, 5^{ème} édition.
 Burner S. (2004). Le rapport à l'autre dans la psychose bi-polaire. In *Actes JADT*.

- Calberg-Challot M., Candel D. and Roche C. (2007). De la variation des usages au consensus terminologique, vers un dictionnaire de l'ingénierie nucléaire. In *Actes Toth*.
- Courtial J.P. (2002). Analysis of Social Representations in Action Based on Words Associated by Scientific Articles. *European Review of Applied Psychology*, 52 (3-4) : 221-230.
- Erlos F. (2006). Entités nommées et domaines d'activité dans les discours de communication institutionnelle : Profils types, variations diachroniques et topographiques. In *Actes JADT*.
- Godelier É. (2004). Le changement de l'entreprise vu par les sciences de gestion ou l'introuvable conciliation de la science et de la pratique. *Entreprises et Histoire*, 35 : 31-44.
- Gouadain D. (2001). Les mots de la gestion. *Gérer et comprendre*, 66 : 58-80.
- Grange D. and Lebart L. (1993). *Traitement statistiques des enquêtes*. Paris : Dunod.
- Habert B. and Salem A. (1995). L'utilisation de catégorisations multiples pour l'analyse quantitative de données textuelles. *TAL*, 36 : 1-2.
- Lancini A. (2002). Réalité de la gestion des connaissances dans les entreprises. In *Sciences de gestion et Pratiques Managériales, réseau des IAE*, Paris.
- Lebart L. (2009). Visualization of multivariate data : What is a valid pattern? Application to textual data. In *Symposium on learning and data science*.
- Lebart L. and Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Lebart L., Marchand P., Yvon F., Sourty-Leguëllec M.J., Zimina M. and Salem A. (2001). Analyse de données textuelles. *SFDS, Enquêtes, modèles et applications*, Paris.
- Lebaron F. (2009). Geometrical data analysis in a sociological research program. In *Symposium on learning and data science*.
- Lemoigne J.L. (1977). *Les Systèmes d'Information dans les Organisations*. Paris : PUF.
- Leroux B. (2009). Homage to P.Bourdieu and H.Rouanet. In *Symposium on learning and data science*.
- Malrieu D. Genres et variations morphosyntaxiques. Quelles variables pertinentes? UMR 2329 MoDyCo Paris X - *Équipe Sémantique des textes*, CNRS. 22/05/2001.
- Martinez W. (2005). COOCS - Outils lexicométriques pour l'analyse des cooccurrences – *Manuel d'utilisation*. SYLED-CLA2T.
- Mayaffre D. (2008). Étude de cas et réflexion sur la cooccurrence. In *Actes JADT*.
- Morin E. (1994). *La complexité humaine*. Paris : Flammarion l'essentiel.
- Pincemin B. (2004). Lexicométrie sur corpus étiqueté. In *Actes JADT*.
- Reinert M. (2001). Approche statistique et problème du sens dans une enquête ouverte. *Journal de la Société Française de Statistique*, 142, 4 : 59-71.

Sites

www.medef.fr communiqués de presse.

www.cairn.info/revue-francaise-de-gestion.htm, www.cavi.univ-paris3.fr/lexicometrica Actes des JADT, 2004, 2006, 2008.