



**Bulletin
de méthodologie sociologique**
Bulletin of sociological methodology

**100 | 2008
October**

Comment ont évolué les thématiques des 99 premiers numéros de BMS ?

Analyse avec un logiciel de fouille de texte

Mathilde de Saint Léger



Édition électronique

URL : <http://bms.revues.org/3153>
ISSN : 2070-2779

Éditeur

Association internationale de
méthodologie sociologique

Édition imprimée

Date de publication : 1 octobre 2008
Pagination : 16-33
ISSN : 0759-1063

Référence électronique

Mathilde de Saint Léger, « Comment ont évolué les thématiques des 99 premiers numéros de *BMS* ? », *Bulletin de méthodologie sociologique* [En ligne], 100 | 2008, mis en ligne le 01 octobre 2011, consulté le 04 octobre 2016. URL : <http://bms.revues.org/3153>

Ce document a été généré automatiquement le 4 octobre 2016.

© BMS

Comment ont évolué les thématiques des 99 premiers numéros de BMS ?

Analyse avec un logiciel de fouille de texte

Mathilde de Saint Léger

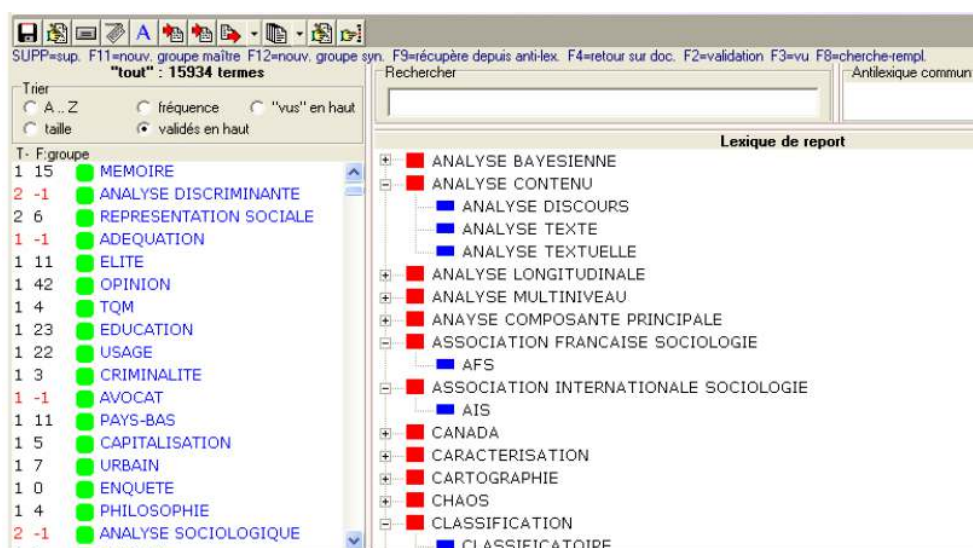
Introduction

- 1 A l'occasion de la parution du centième numéro de *BMS*, nous avons voulu faire une analyse des thématiques des articles déjà publiés à l'aide de logiciels de fouille de texte. Calliope, l'un d'eux, a déjà été présenté dans des numéros de *BMS* (de Saint Léger et van Meter, 2005 ; Demazière, 2005 ; van Meter et Saint Léger, 2008). Un corpus d'articles est ainsi analysé et les principales thématiques sont positionnées sur un diagramme appelé « diagramme stratégique ». Chacune est représentée par un réseau lexical, ou « cluster », dont un des termes est central, au sens où il a un lien de cooccurrence avec tous les autres termes de son réseau. Le diagramme stratégique permet de positionner les thématiques sur un graphique bidimensionnel en fonction de leur centralité (ou mesure des liens avec d'autres thématiques) et densité (ou mesure de la cohérence interne d'une thématique). Un deuxième module de Calliope permet d'établir les liens entre deux diagrammes stratégiques représentant les thématiques de deux corpus différents. Le logiciel calcule pour chaque terme du corpus, un poids quantifié en fonction de sa participation à la construction des clusters et du diagramme stratégique (de Saint Léger, 1997). Ainsi, à partir de deux corpus cartographiés, des termes sont émergents (ceux dont le poids augmente), d'autres sont stables (poids équivalent) et d'autres encore sont déclinants (termes dont le poids diminue). Signalons que ce n'est pas la première fois que Calliope est utilisé dans l'analyse des articles et rapport de recherche publiés dans le *BMS* (van Meter, Cibois et de Saint Léger, 2004).

Méthodologie

- 2 L'analyse des articles du *BMS* est faite sur leur titre, leur résumé et les mots clés. Pour analyser la dynamique du contenu de ces 315 articles, on les a divisés en trois corpus :
 - Période 1983 à 1992, 85 articles
 - Période 1993 à 2000, 119 articles
 - Période 2001 à 2008, 111 articles
- 3 La première période est plus longue que les deux autres car Calliope est un outil d'analyse statistique, pour cela le nombre de documents analysés doit être sensiblement équivalent d'une période à l'autre.
- 4 La première étape consiste à faire une extraction automatique de la terminologie des 315 articles puis à valider et homogénéiser le vocabulaire extrait. A la fin de cette étape, chaque article du corpus est représenté par des termes validés qui sont alors traités par le deuxième module de Calliope. Ainsi, les termes « analyse discours », « analyse texte » et « analyse textuelle » sont validés mais reportés sur le terme « analyse de contenu ». De même, les termes « donnée textuelle », « validité donnée » ou « scientométrie » sont validés, alors que les termes « analyse » ou « étude » ne le sont pas, car ils sont trop génériques et n'apportent aucune information spécifique au contenu des articles analysés (voir Figure 1).
- 5 Dans la deuxième étape, Calliope construit le diagramme stratégique de chaque corpus qui est une cartographie thématique du contenu suivant la méthode des mots associés. Il est fondé sur la méthode des mots associés (Callon, Courtial et Turner, 1991). Enfin la troisième étape consiste à calculer le poids des termes et à les trier suivant la variation de leur poids d'une période à l'autre, en termes émergents, stables ou déclinants (de Saint Léger, 1997).

Figure 1 : Construction de la terminologie et validation des termes pertinents



Diagrammes stratégiques

- 6 Un diagramme stratégique est construit avec comme axe des abscisses la médiane de la centralité et comme axe des ordonnées la médiane de la densité. La distribution des clusters dans les différents quadrants donne alors une synthèse de la dynamique du domaine qu'ils représentent.
- 7 Pour mieux appréhender la connexion des thématiques d'une période à l'autre, les tableaux ci-dessous reprennent les valeurs minimale, médiane et maximale de la centralité et de la densité (Tableaux 1 et 2):

Tableau 1 : Evolution de la centralité

Centralité	Min.	Médiane	Max.
1983 à 1992	258	743	948
1993 à 2000	375	1011	1155
2001 à 2008	570	947	1183

- 8 Les valeurs extrêmes de la centralité augmentent au cours du temps. On peut faire l'hypothèse alors que les thématiques des articles qui portent principalement sur les méthodologies d'analyse, s'ouvrent et s'interconnectent (voir Figures 2, 3 et 4). Parallèlement, la distribution des clusters autour de la médiane est plus équilibrée dans la troisième période (Figure 4). Le domaine atteint alors un rythme de croisière dynamique, il n'y a pas de clusters isolés très homogènes et peu connectés (quadrant 2). En revanche, il y a autant de thématiques peu connectées et peu homogènes, donc émergentes (quadrant 4) que de thématiques très connectées et très homogènes, donc fédératrices (quadrant 1).
- 9 D'autre part, d'après le tableau des densités ci-dessous (Tableau 2), le maximum des valeurs de densité diminue : les thématiques perdent relativement en cohérence interne, cela peut s'expliquer par l'augmentation de leur interconnectivité. On peut faire l'hypothèse qu'au début des années 1980, dans le BMS, on traitait des problématiques d'analyse statistiques en sociologie, puis, avec les années et notamment avec l'usage grandissant de l'informatique, toutes ces problématiques se sont interconnectées, depuis le traitement des données jusqu'à l'assistance par ordinateur pour des enquêtes. Ainsi la thématique autour de la « Qualité donnée » (Figure 5) dans la période 1983-1992 peut être considéré comme s'insérant dans un cadre plus global dans la période 2001-2008 qui est la « Statistique » (Figure 18).

Tableau 2 : évolution de la densité

Densité	Min.	Médiane	Max.
1983 à 1992	67	224	744
1993 à 2000	67	150	444
2001 à 2008	54	190	195

Figure 2 : Diagramme stratégique du BMS 1983-1992

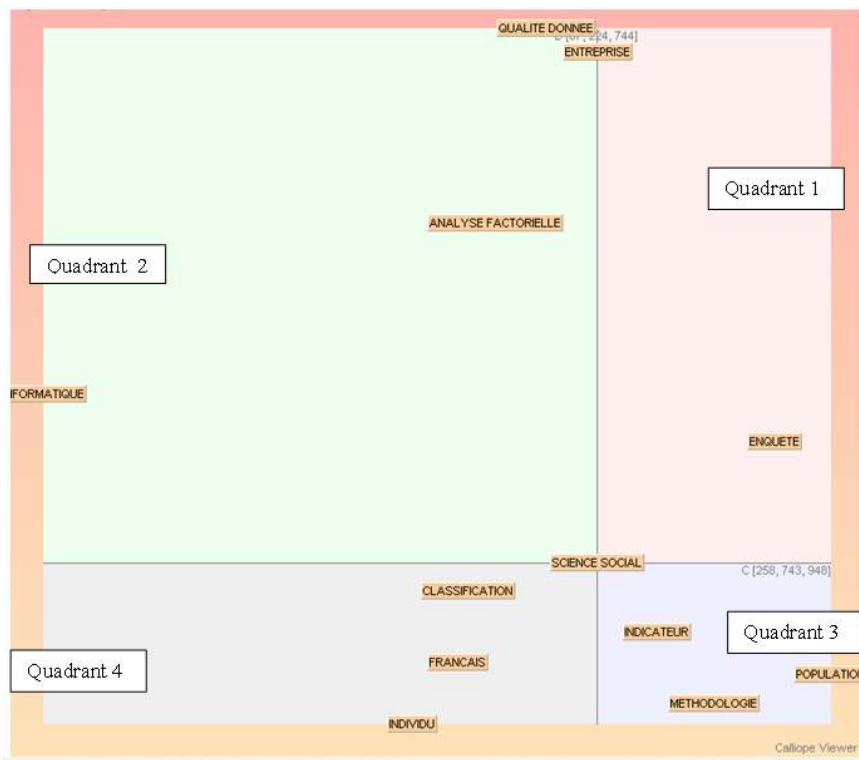


Figure 3 : Diagramme stratégique du BMS 1993-2000

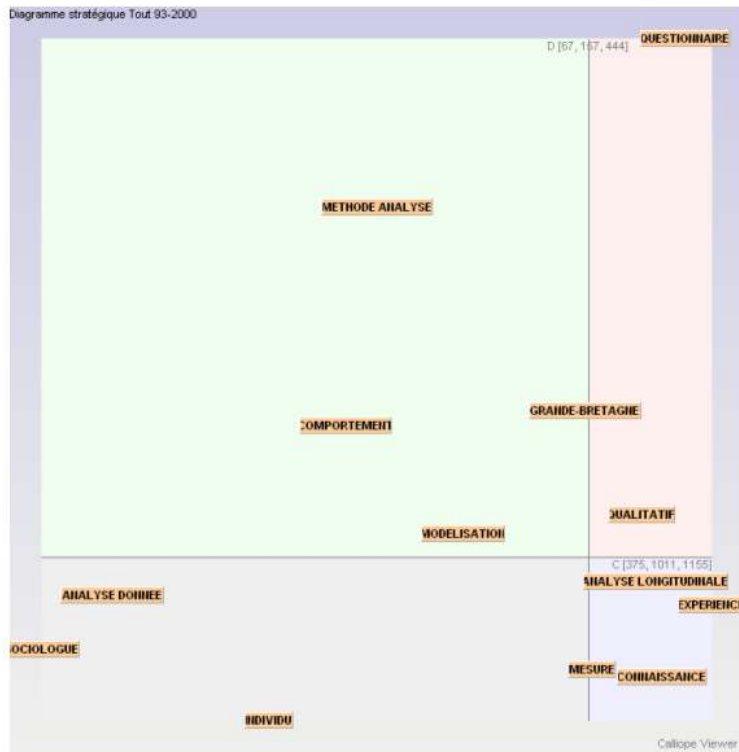
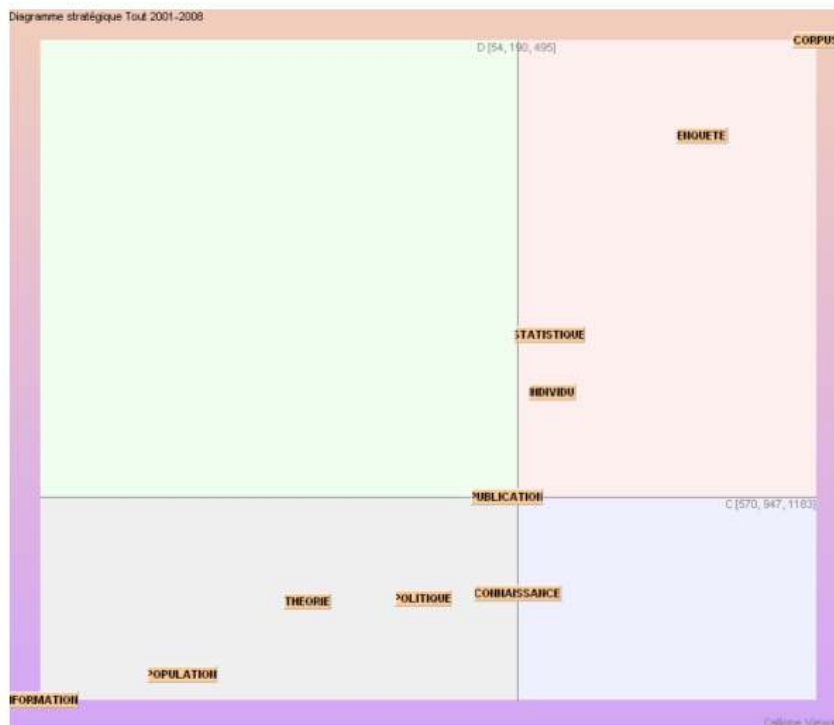


Figure 4 : Diagramme stratégique du BMS 2001-2008



Analyse des clusters

- 10 Au-delà d'un diagramme stratégique, cartographie donnant une vision globale du corpus analysé, il est nécessaire de zoomer sur le contenu de chaque cluster pour avoir une vision détaillée et donc une analyse plus fine des thématiques du corpus.

Période 1983-1992

- 11 Quatre clusters du diagramme stratégique de cette période (cf. Figure 2) arrivent en tête sur le diagramme stratégique, il s'agit des clusters « Qualité donnée », « Entreprise », « Analyse factorielle » et « Enquête ». Ils sont les plus centraux et les plus denses.
- 12 C'est le premier cluster qui a été construit par Calliope pour cette période, il regroupe 13 articles sur 83¹, soit 15%. Un retour sur les dates (voir graphique après la Figure 5) montre que cela concerne principalement des articles parus en 1991 et porte sur le mode d'administration d'une enquête (ou entretien ou interview), d'une part, et sur la qualité des données recueillies, d'autre part.

Figure 5 : Composition du cluster « Qualité donnée »

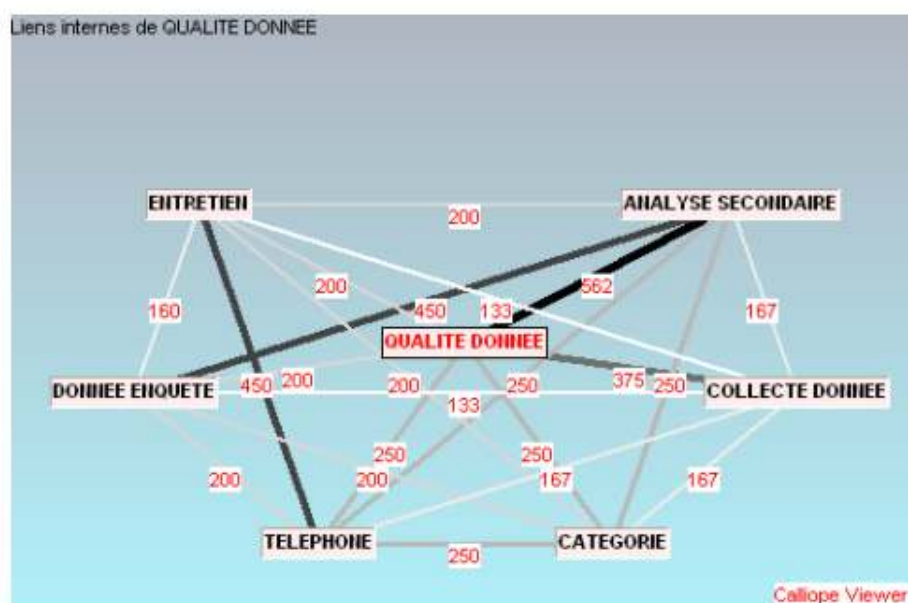


Figure 6 : Termes liés au cluster « Qualité donnée »

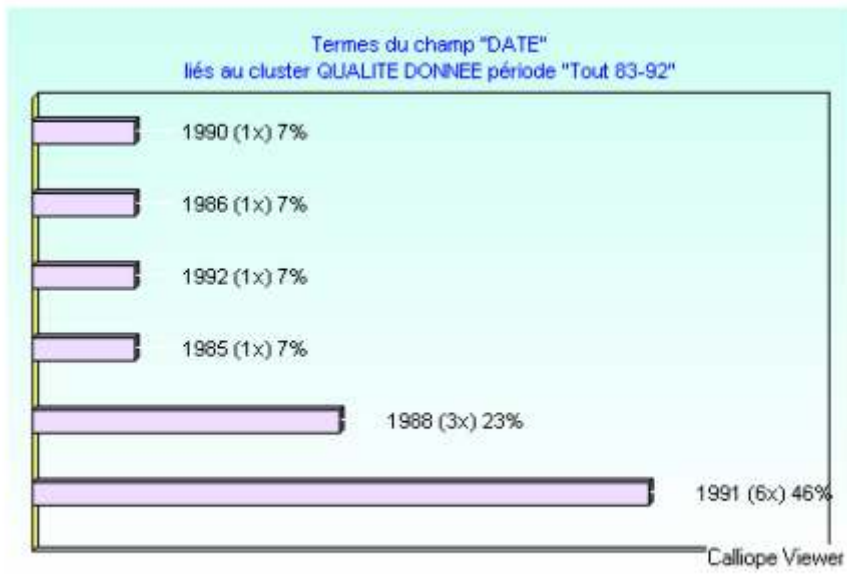
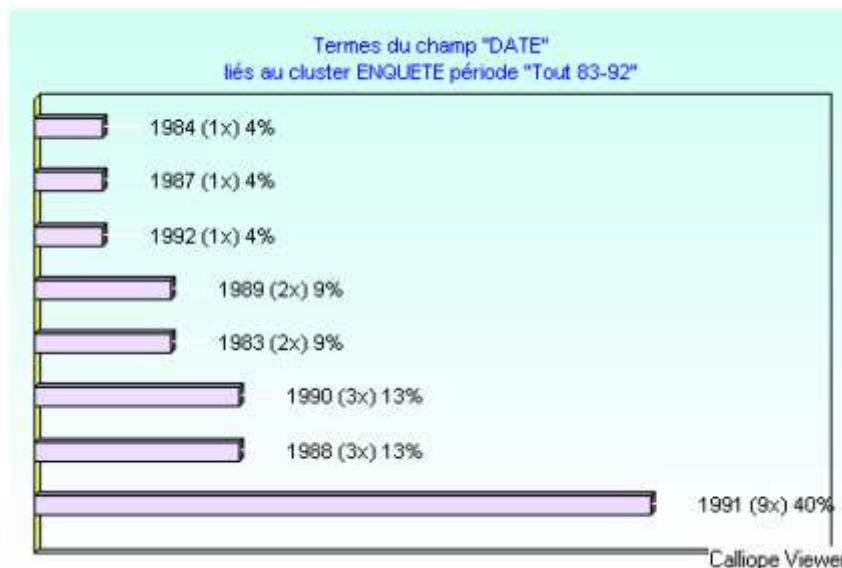
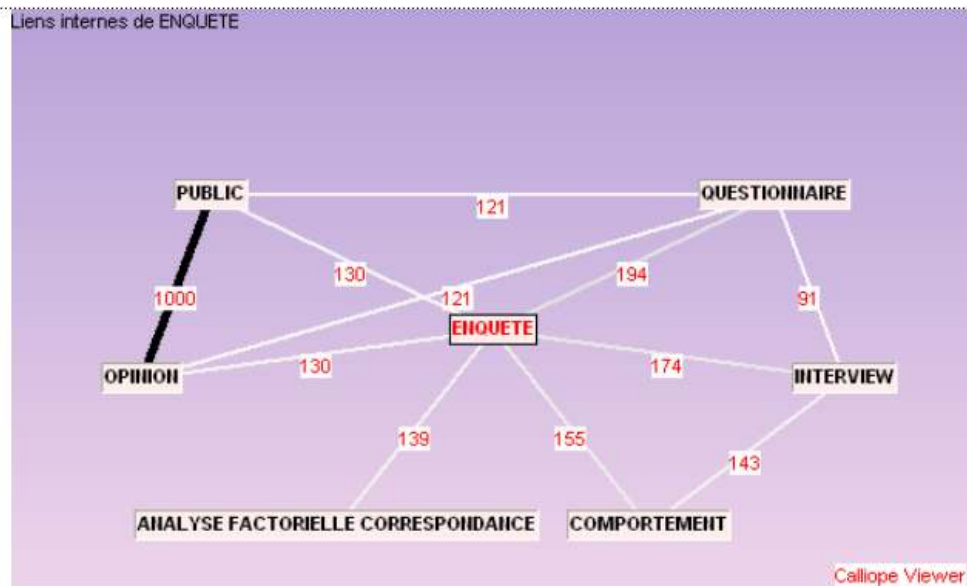


Figure 7 : Composition du cluster « Enquête » et ses liens internes



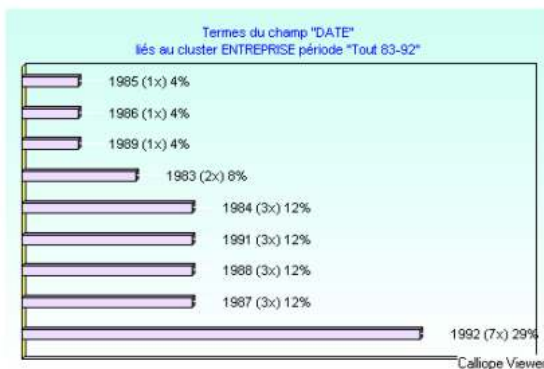
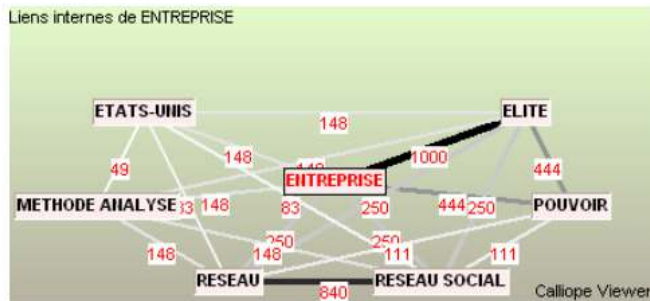
- 13 Ce cluster, suite thématique de « Qualité donnée », regroupe 22 articles sur 83, soit 26% du corpus. Un retour sur les dates (voir graphique ci-dessous) montre que cela concerne aussi en majorité des articles parus en 1991.
- 14 On peut noter que le terme « Analyse factorielle correspondance » est uniquement lié à « Enquête ». C'est le terme 'trait d'union' avec le cluster « Analyse factorielle » (cluster Figure 8). On peut supposer que ces deux termes sont synonymes², pourtant, ils n'apparaissent pas systématiquement dans les mêmes articles. D'où l'importance de la validation des termes extraits par Calliope-Extract suivant une méthode fondée sur la statistique³, par un expert du domaine, qui décide si ces deux termes sont équivalents ou pas.

Figure 8 : Composition du cluster « Analyse Factorielle » et ses liens internes



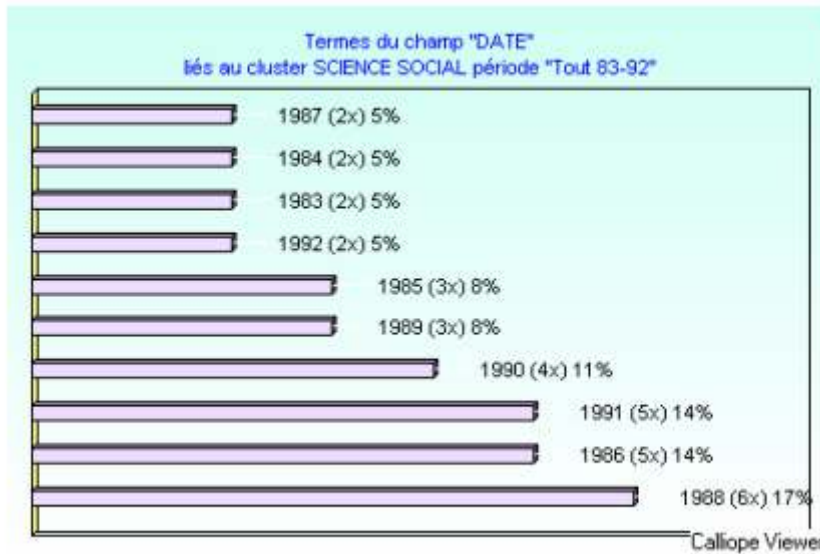
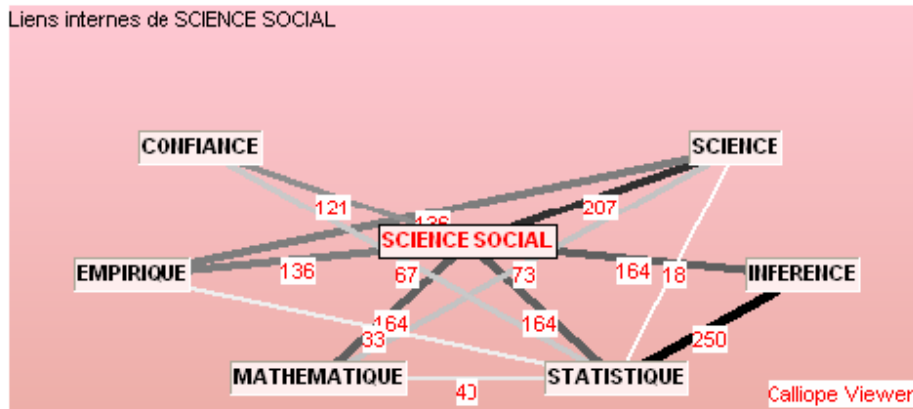
- 15 Le cluster « Entreprise », un peu singulier sur le diagramme stratégique car ne se rapportant pas directement à une méthodologie statistique mais à un objet d'étude, regroupe 24 articles sur 83, soit 29% du corpus. Un retour sur les dates (voir graphique après la Figure 9 ci-dessous) montre que cela concerne en majorité des articles parus en 1992.

Figure 9 : Composition du cluster « Entreprise » et ses liens internes



- 16 D'autre part, un retour sur les articles concernés, montre que cela traite principalement des méthodes d'analyse de réseaux sociaux en entreprise, principalement aux Etats-Unis (9 articles sur 24, soit 37%).

Figure 10 : Composition du cluster « Science sociale » et ses liens internes



- 17 Ce cluster, au croisement des médianes de la centralité et de la densité (voir Figure 2), représente les articles traitant de méthodes statistiques, outils de recherche en science sociale. La distribution des articles se rapportant à ce cluster sur les années étudiées, montre que cela a fait l'objet d'au moins un article par an (voir graphique ci-dessous).

Période 1993-2000

- 18 La période 1993-2000 montre une certaine évolution des thèmes (voir Figure 3). Le terme « Questionnaire », qui figurait dans le cluster dominant « Enquête » en 1983-1992, est devenu le terme central du cluster le plus central et le plus dense de cette période. Sa composition (figure 11), avec les termes « Enquête », « Qualité donnée » et « Interview », montre qu'il a « avalé » le cluster « Qualité donnée » de la période précédente, tout en incluant deux nouveaux termes « Pré-test » et « Cognitif ». L'autre cluster « Grande-Bretagne » du quadrant 1 est nouveau venu, c'est un objet d'étude. Quant à « Qualitatif » toujours du quadrant 1, il se trouvait en 1983-1992 dans « Méthodologie », cluster peu structuré du quadrant 3 (voir Figure 14).

- 19 En fait, l'analyse des clusters de cette période montre le glissement des thématiques dans les articles du *BMS*, des méthodologies d'analyse vers les objets de recherche.

Figure 11 : Composition du cluster « Questionnaire » et ses liens internes

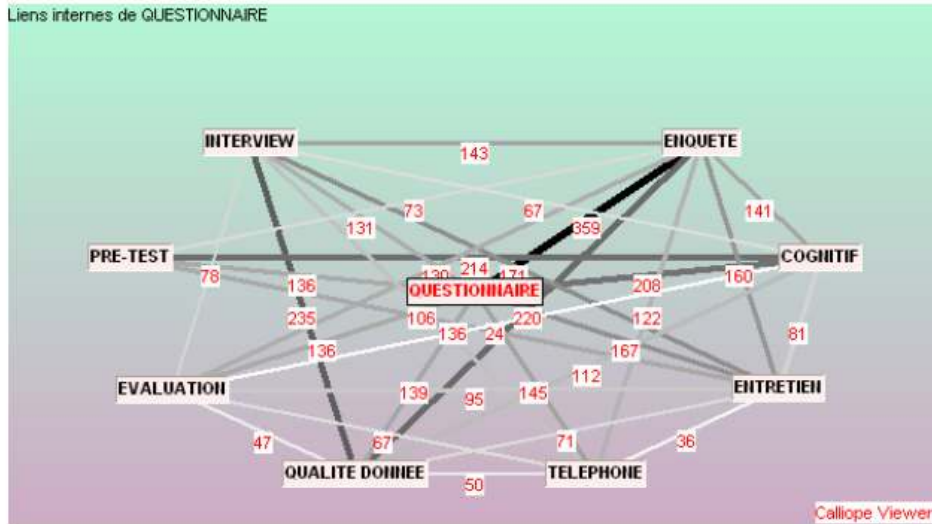


Figure 12 : Composition du cluster « Grande-Bretagne » et ses liens internes

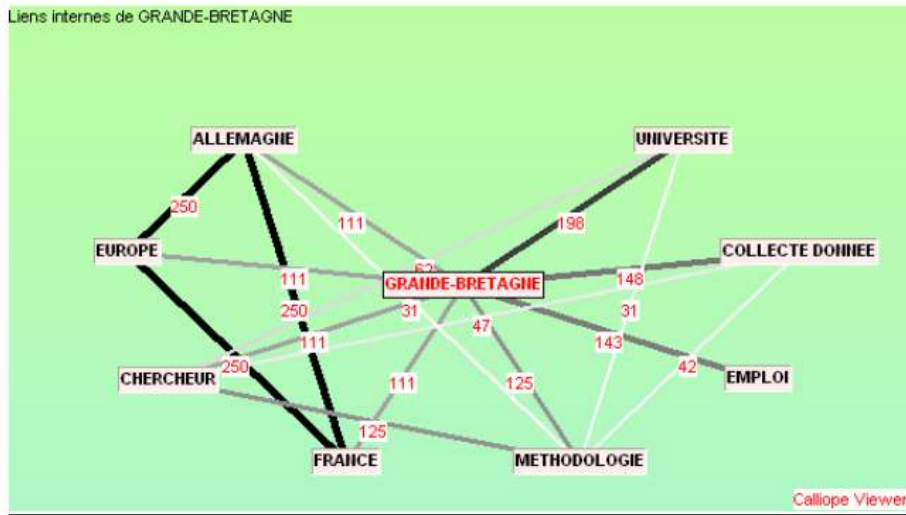


Figure 13 : Composition du cluster « Qualitatif » et ses liens internes

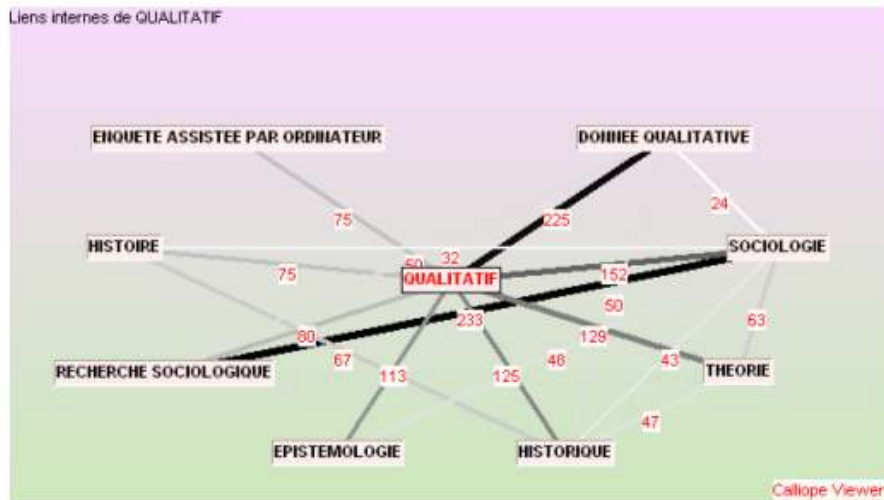
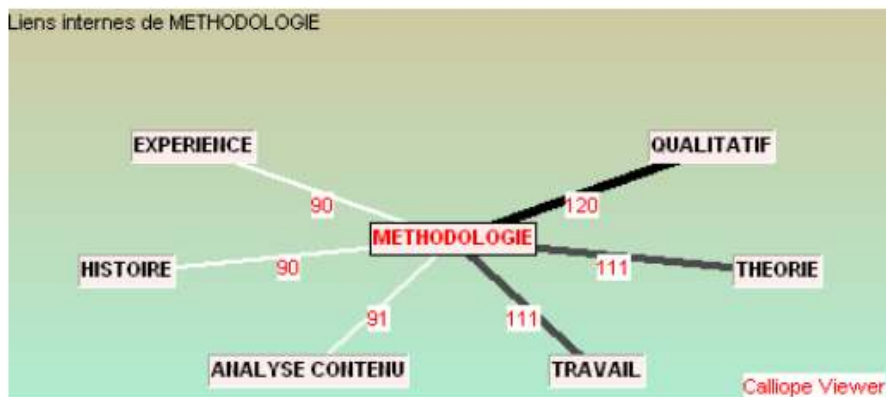


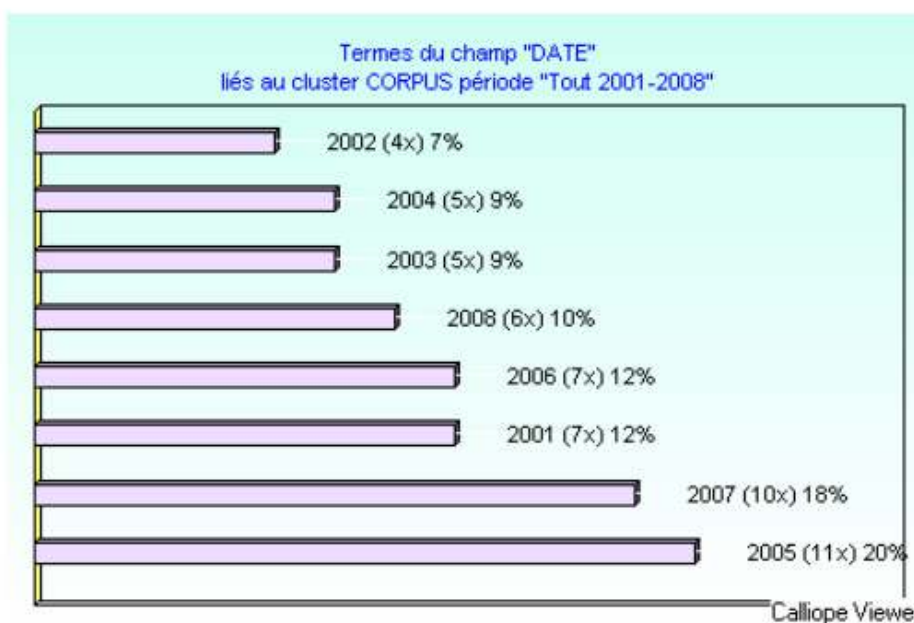
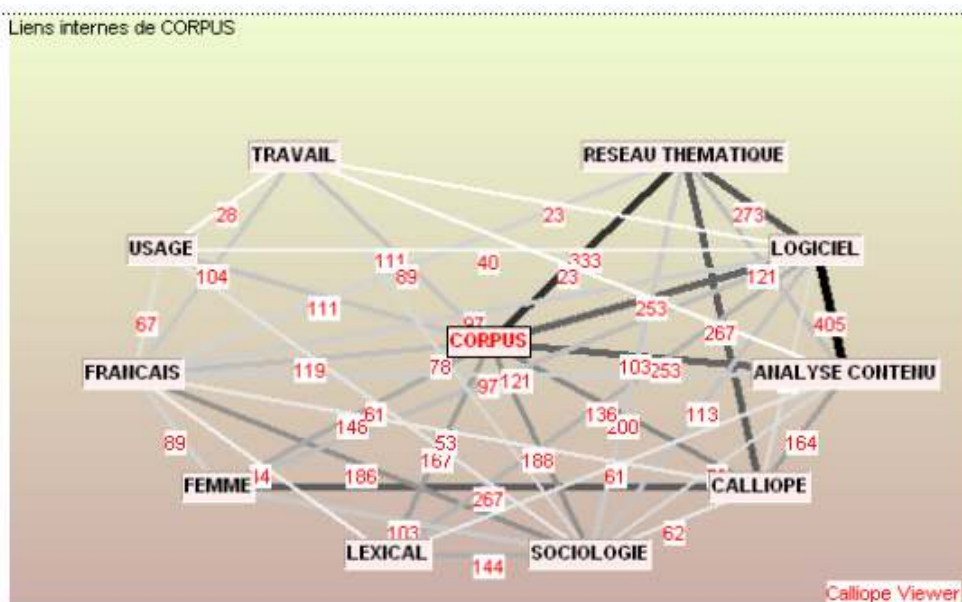
Figure 14 : Composition du cluster « Méthodologie » et ses liens internes



Période 2001-2008

- 20 Dans la période actuelle (2001-2008), les clusters les plus centraux et les plus denses sont dans l'ordre : « Corpus », « Enquête », « Statistique » et « Individu » (voir Figure 4).

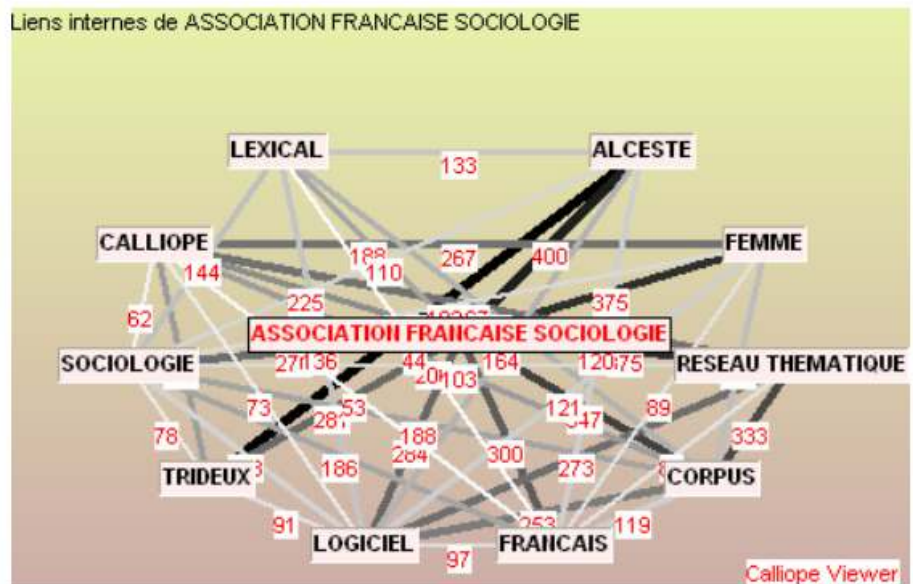
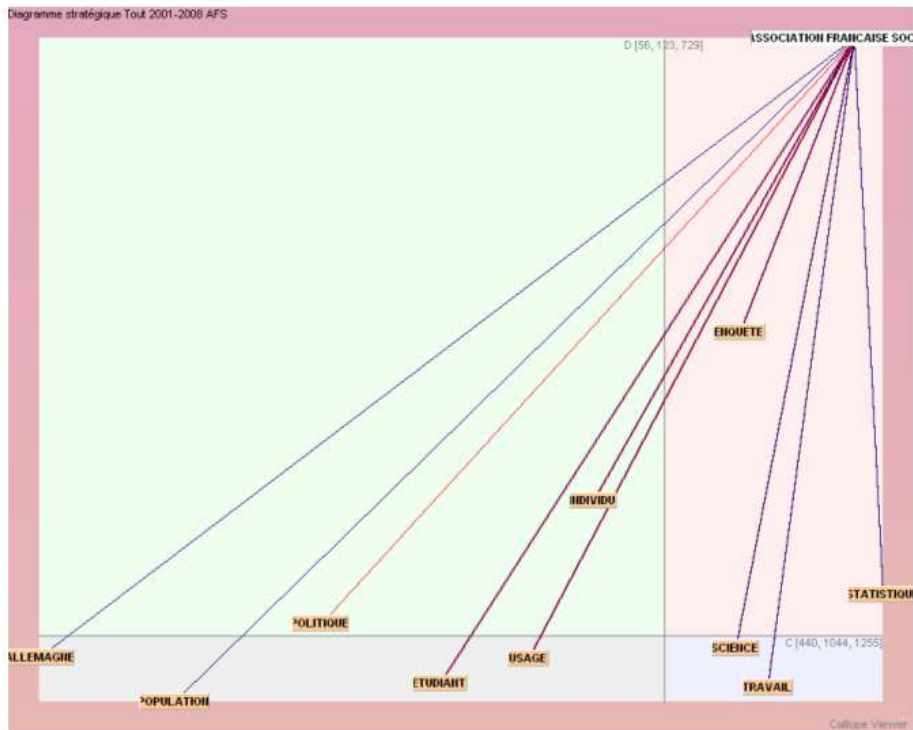
Figure 15 : Composition du cluster « Corpus » et ses liens internes



- 21 Le détail de ce cluster montre que les thématiques des articles du BMS ont glissé de problématiques méthodologiques au cours des deux périodes précédentes, aux outils informatiques de traitement de données (« Logiciel », « Calliope »), avec l’incursion des objets de recherche (« Femme », « Français », « Travail ») dans ce premier cluster.
- 22 Un retour sur les dates (voir graphique ci-dessous) montre que cette évolution a commencé dès 2001.
- 23 Nous tenons à faire remarquer, qu’en 2005 et 2006, sept articles du BMS ont été consacrés aux logiciels d’analyse de contenu et plus spécifiquement l’analyse des résumés des présentations des deux premiers congrès de l’Association Française de Sociologie (AFS). Ces articles ont également servi de base pour le livre *Analyses textuelles en sociologie*

(Demazière *et al.*, 2006). On peut se demander alors, pourquoi le terme AFS n'apparaît pas au moins dans un cluster du diagramme stratégique ? Nous avons fait le choix de ne pas valider ce terme pour l'analyse avec Calliope de cette période, car la proximité thématique des sept articles dus à un seul évènement, aurait quelque peu biaisé la structure du diagramme stratégique sur la période 2001-2008. C'est ce que montrent les figures ci-dessous du diagramme stratégique et du cluster AFS, obtenus en validant ce dernier (Figure 16).

Figure 16 : Diagramme stratégique du BMS 2000-2008 avec « AFS »



- 24 Le cluster AFS se détache des autres sur le diagramme stratégique ; c'est un attracteur (de Saint Léger et van Meter, 2005 ; Demazière *et al.*, 2006 ; van Meter et de Saint Léger, 2007).
- 25 Le cluster « Enquête » semble être assez clairement la continuation du cluster « Questionnaire » de la période précédente. Il regroupe 69 articles sur 111 (soit 62%), qui sont uniformément répartis sur la période analysée (voir graphique ci-dessous).

Figure 17 : Composition du cluster « Enquête » et ses liens internes

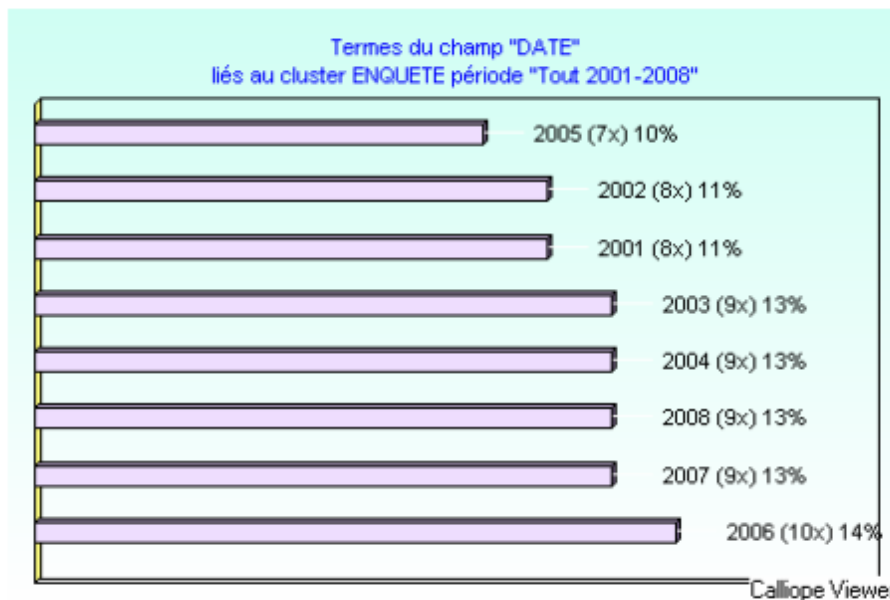
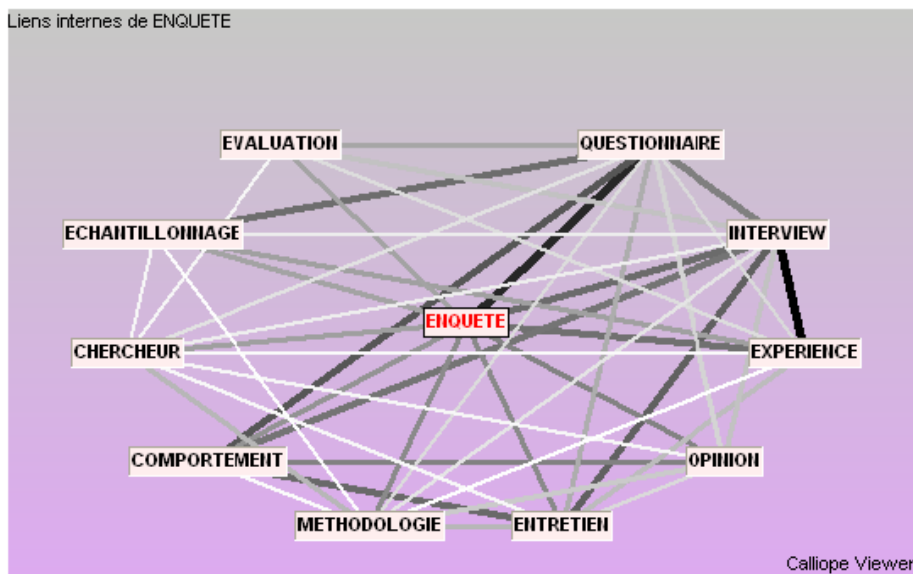
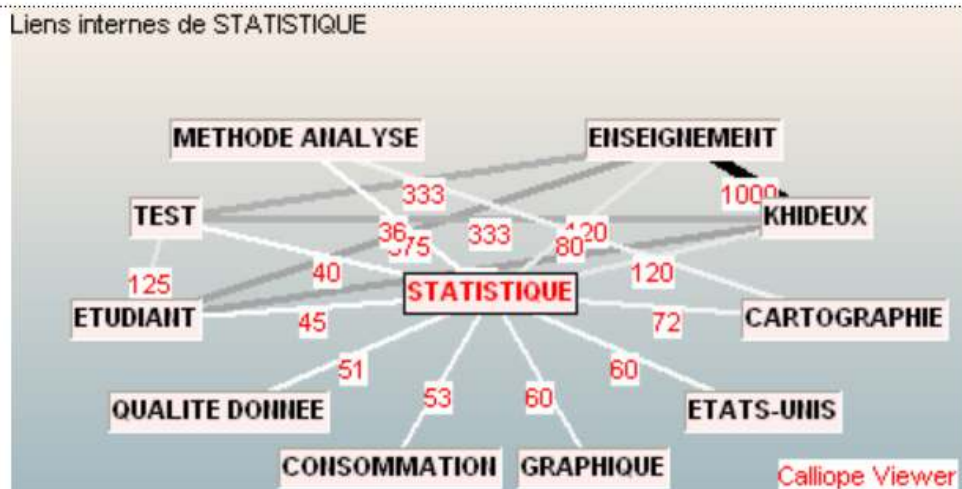
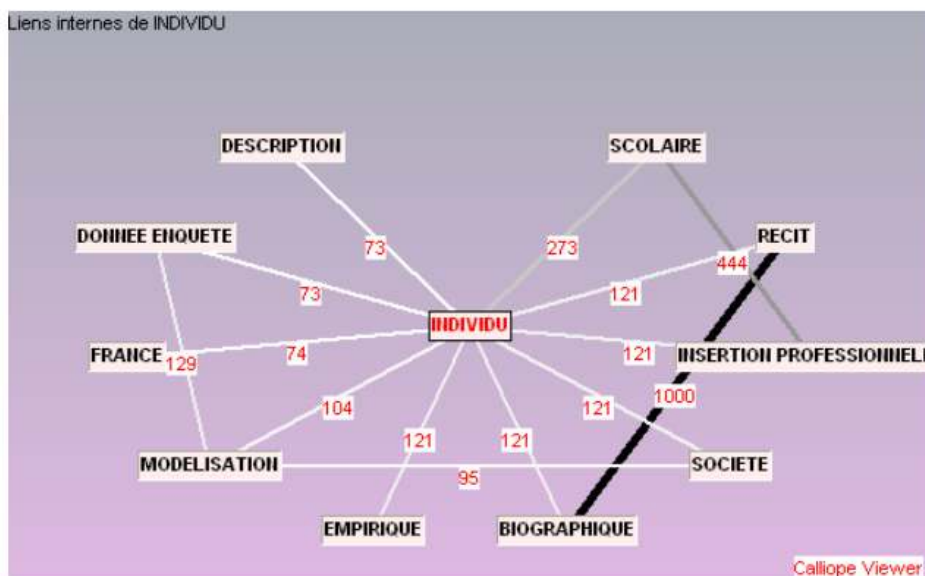


Figure 18 : Composition du cluster « Statistique » et ses liens internes



26 Ce cluster « Statistique », reprend certains aspects du cluster « Qualité donnée » de la première période (voir Figure 5).

Figure 19 : Composition du cluster « Individu » et ses liens internes



27 Ce cluster « Individu » introduit un nouveau thème pour la période 2001-2008 qui concerne des méthodes biographiques et de récit, fondées sur la mise en œuvre d'entretiens narratifs.

Conclusion : Analyse des tendances

28 L'analyse des tendances avec Calliope sur les trois périodes successives (voir Figure 20) montre l'importance croissante des outils informatiques dans l'analyse des données (« Alceste », « Calliope »). Parallèlement, les objets d'études (« Etudiant », « Femme », « Consommation », « Web ») prennent aussi de plus en plus d'importance dans les articles par rapport aux termes de méthodologies d'analyse.

- 29 A l'inverse, les courbes ci-dessous (Figure 21) montrent l'évolution du poids des termes déclinants (« Analyse correspondance », « Analyse donnée », « Classification », « Collecte donnée », « Informatique », « Tableau croisé », « Traitement donnée »). Ils sont liés à la méthodologie d'analyse et au recueil de données. Cela porte à croire, une fois de plus, qu'au cours de ces années, les thématiques des articles du BMS se sont déplacées des méthodologies aux objets d'étude. Peut-on dire alors que les méthodologies sont maintenant bien cernées ? Cette évolution, est-elle due en partie au progrès de l'outillage informatique ?

Figure 20 : Poids des termes « émergents » de 1983-1992 à 2001-2008 en passant par 1993-2000

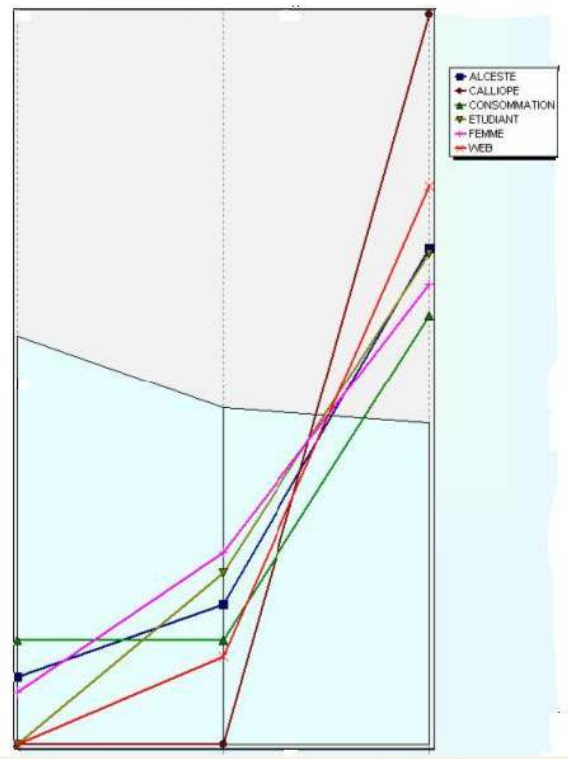
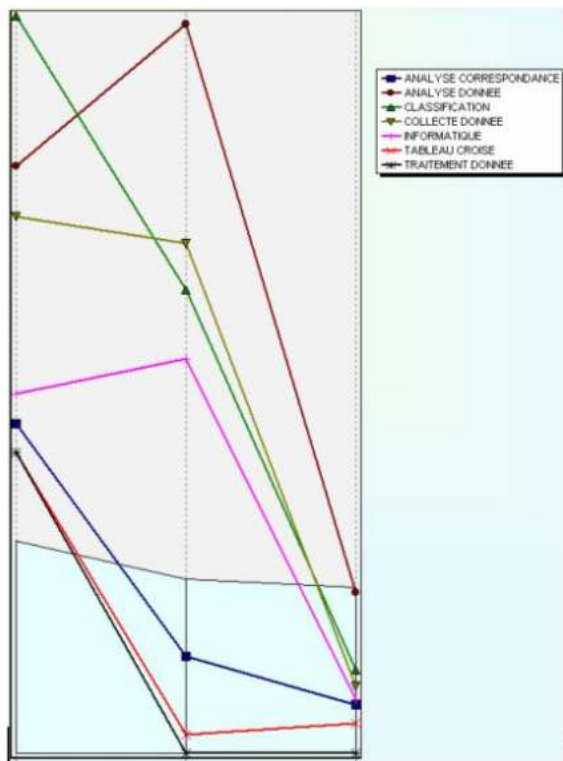


Figure 21 : Poids des termes « déclinants » en 1983-1992, 1993-2000 et 2001-2008



BIBLIOGRAPHIE

Michel Callon, Jean-Pierre Courtial et William Turner (1991), La méthode Leximappe : Un outil pour l'analyse stratégique du développement scientifique et technique, in D. Vinck (sous la coordination de), *Gestion de la recherche. Nouveaux problèmes, nouveaux outils*, Armand Colin.

Mathilde de Saint Leger (1997), *Modélisation de la dynamique des flux d'information, vers un suivi des domaines de connaissance*, Thèse de doctorat en Science de l'Information (CNAM).

Mathilde de Saint Léger et Karl M. van Meter (2005), Cartographie du premier congrès de l'ASF avec la méthode des mots associés, *Bulletin de Méthodologie Sociologique (BMS)*, n. 85, January.

Didier Demazière (2005), Des logiciels d'analyse textuelle au service de l'imagination sociologique, *Bulletin de Méthodologie Sociologique (BMS)*, n. 85, January.

Didier Demazière, Claire Brossaud, Patrick Trabal et Karl van Meter (2006), *Analyses textuelles en sociologie : Logiciels, méthodes, usages*, Rennes, Presses universitaires de Rennes.

Karl M. van Meter et Mathilde de Saint Léger (2008), Co-Word Text Analysis Applied to Political Science: 2006 International Political & "Parapolitical" Headlines, *Bulletin de Méthodologie Sociologique (BMS)*, n. 97, January.

Karl M. van Meter et Mathilde de Saint Léger (2007), Preliminary Co-Word Analysis of the 2006 Congress of the Association Française de Sociologie, *Bulletin de Méthodologie Sociologique (BMS)*, n. 93, January.

Karl M. van Meter, Philippe Cibois et Mathilde de Saint Léger (2004), Correspondence & Co-Word Analysis of Ten Years of *BMS* articles (1993-2003), *Bulletin de Méthodologie Sociologique (BMS)*, n. 81, January.

NOTES

1. Un article est lié à un cluster si au moins 2 de ses termes s'y trouvent
2. Un retour sur les termes dans leur contexte, montre que « Analyse factorielle » apparaît dans 12 articles alors que « Analyse factorielle correspondance » apparaît dans 7.
3. La méthode des segments répétés

RÉSUMÉS

Les titres, résumés et mots clefs des 315 articles publiés dans les numéros 1 à 99 du *BMS* entre 1983 et 2008 ont été divisés en trois périodes : 1983-1992 (85 articles), 1993-2000 (119 articles) et 2001-2008 (111 articles). Les trois corpus ont été analysés individuellement par Calliope, une analyse par co-occurrence de mots clefs dans des unités de texte qui produit des « cluster » thématiques. Elle construit aussi des diagrammes stratégiques qui montrent la structure thématique générale pour chaque période avec des axes de densité (liens internes à un « cluster ») et de centralité (liens externes entre « clusters »). Ces diagrammes, de ceux des « clusters » les plus dominants sont présentés et interprétés.

How have the themes of the first 99 issues of the *BMS* evolved, An analysis with a text-mining software: The titles, abstracts (in French) and key words (in French) of the 315 articles published in issues 1 to 99 of the *BMS* from 1983 to 2008 were divided into three periods: 1983-1992 (85 articles), 1993-2000 (119 articles) and 2001-2008 (111 articles). These three corpora were individually analyzed with Calliope, an analysis by co-occurrence of key words in units of text, which produces thematic clusters. It also constructs strategic diagrams, which show the general thematic structure of each period with its axes of density (ties within a thematic cluster) and centrality (external ties between clusters). These diagrams, and those of the more dominant clusters, are presented along with their interpretation.

INDEX

Mots-clés : BMS, Analyse textuelle, Calliope, Analyse par mots associés, Bibliométrie

Keywords : Text analysis, Co-Word Analysis, Bibliometrics

AUTEUR

MATHILDE DE SAINT LÉGER

CNRS-IPAM, mathilde.de-saint-leger@cnrs-dir.fr