



**HAL**  
open science

# L'analyse bibliométrique comme outil d'aide à la mise en place d'un groupe de recherche en physique fondamentale

Marie-Gabrielle Suraud, Luc Quoniam, Hervé Rostaing, Henri Dou

## ► To cite this version:

Marie-Gabrielle Suraud, Luc Quoniam, Hervé Rostaing, Henri Dou. L'analyse bibliométrique comme outil d'aide à la mise en place d'un groupe de recherche en physique fondamentale. *Revue Francaise de Bibliometrie*, 1994, 13, pp.100-120. hal-01547116

**HAL Id: hal-01547116**

**<https://hal.science/hal-01547116v1>**

Submitted on 26 Jun 2017

**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.

## **L'analyse bibliométrique comme outil d'aide à la mise en place d'un groupe de recherche en physique fondamentale.**

**M. G. Suraud<sup>1</sup>; L. Quoniam<sup>2</sup>; H. Rostaing<sup>2</sup>; H. Dou<sup>2</sup>**

**1. LERASS, IUT-A, 115 Route de Narbonne, 31077 Toulouse cedex**

**2. CRRM, Centre Saint Jérôme, 13397 Marseille cedex 13**

### **I Introduction**

La veille stratégique s'est récemment développée pour répondre aux besoins des industries en matière d'informations élaborées et d'aide à la décision dans le champ de l'innovation <sup>1,2</sup>. La nécessité incontournable pour les entreprises de "ne pas se contenter de produire et de vivre sur ses acquis induit une prise de conscience croissante de l'intérêt de maîtriser parfaitement les informations concernant l'industrie..."<sup>3</sup>. Les premiers travaux de veille stratégique affichaient donc un objectif de compétitivité industrielle. Ils se fondaient sur des masses d'information variées tant par le contenu que par leur forme. Par la suite ces activités se sont étendues à des domaines plus larges, comme celui de la recherche scientifique. Dans le cadre des activités de veille stratégique, l'outil bibliométrique s'est imposé et représente désormais un maillon crucial de la chaîne de traitement<sup>4</sup>. Fondé sur l'analyse statistique des publications scientifiques et/ou des brevets déposés, cet outil est utilisé pour mettre en évidence certaines caractéristiques du développement industriel ou de l'activité de recherche scientifique (détermination de fronts de recherche, de groupes de collaboration, de "collèges invisibles" ou de domaines à fort potentiel technologique...). La pertinence des analyses statistiques et donc le succès de la bibliométrie tient au caractère codifié et systématique des grandes masses d'informations qu'elle traite, en particulier des informations issues des banques de données.

Le développement des banques de données, et en parallèle celui des micro-ordinateurs, a ouvert l'accès à des données documentaires de masse. L'imposante quantité de données ainsi disponible ne représente pas toujours une information facilement exploitable et directement utilisable. La nécessité de synthétiser les documents bibliographiques afin d'en extraire une information spécifique et adaptée à la demande a favorisé le développement de techniques d'analyse statistique particulières. "Les données ne fournissent pas l'information nécessaire à la décision. Ce qu'il est nécessaire de voir, ce sont les relations que l'ensemble des données construit. L'information utile à la décision est faite de relations d'ensemble... L'information utile n'est pas un accroissement de la quantité d'information mais, tout au contraire, une réduction de cette quantité par des groupes pertinents. Ces regroupements ne correspondent pas aux catégories qui ont permis d'imaginer les tableaux des données mais sont de nouveaux regroupements définis par l'ensemble des relations que les données entretiennent entre elles"<sup>5</sup>. Cette citation de J. Bertin résume parfaitement la nature et les objectifs du traitement bibliométrique.

Si en bibliométrie la matière première est généralement la référence bibliographique, en revanche les méthodes de traitement de ces références varient considérablement en fonction de l'information recherchée et de la problématique posée. Plusieurs méthodes d'analyse de références bibliographiques ont ainsi été développées. Comme exemple, nous pouvons citer les travaux exploitant la base de données *Science Citation Index* de l'*Institute for Scientific Information* qui, par analyse des citations et des co-citations cherchent à déterminer, respectivement, l'impact de certains travaux scientifiques et la formation de front de recherche, de même que la création de cartographies de la science<sup>6</sup>. D'autres études fondées sur "l'analyse des mots associés" ont pour but de caractériser différents thèmes de recherche correspondant à des domaines spécifiques. La plupart des méthodes utilisées en bibliométrie font intervenir les techniques classiques d'analyse statistique factorielle comme l'ACP, l'AFC ou l'ACM... souvent difficiles d'emploi et dont l'interprétation reste délicate. Notre méthode de traitement, qui sera développée en détail par la suite, s'affranchit de l'utilisation de telles méthodes. Elle fonctionne essentiellement sur le principe de l'enchaînement des co-occurrences d'apparition de formes (une forme étant définie comme un élément documentaire d'une référence bibliographique<sup>4</sup>). Cette procédure a déjà été utilisée et décrite dans de nombreux travaux<sup>7</sup>. Elle présente entre autres avantages celui d'une lecture "directe" des résultats, ce qui en facilite l'exploitation.

L'étude présentée ici s'inscrit dans le champ des études bibliométriques appliquées au domaine de la recherche "fondamentale" en physique. Elle représente le

fruit de la collaboration entre le LERASS, laboratoire de sciences humaines et sociales, le CRRM, laboratoire spécialisé dans les travaux de bibliométrie et des physiciens appartenant à différents laboratoires nationaux. Bien que les études de type bibliométrique soient de plus en plus nombreuses depuis ces dernières années, très peu d'entre elles concernent ce domaine d'activité. L'originalité de ce travail tient donc entre autre à son objet d'étude; il s'agit d'accompagner la mise en place d'un Groupement de Recherche (GdR) du CNRS en physique sur le thème de la fragmentation. La nécessité de mener une étude de veille technologique tient à la nature inhabituelle du contexte scientifique dans lequel s'inscrit le GdR de physique. En effet, d'une part, le thème du GdR est large, puisqu'il concerne plusieurs sous-disciplines de physique. Or les sous-disciplines de physique concernées par ce projet constituent des communautés qui travaillent de façon totalement indépendantes et qui ne communiquent que très difficilement entre elles. D'autre part c'est un thème en pleine évolution, au sens où le concept central est relativement peu stabilisé. Dans chaque sous-discipline sont donc développés des problématiques et des outils très spécifiques à chacune d'entre elles. Dans ce contexte la littérature est particulièrement vaste, peu normalisée du point de vue scientifique donc documentaire et par conséquent difficile à cerner. Or la création du GdR nécessite la connaissance de l'état actuel des recherches pour définir l'ensemble des travaux concernant la fragmentation dans toutes les sous-disciplines, et pour évaluer leur degré de maturité.

## **II La fragmentation**

L'étude de la fragmentation en physique est un phénomène relativement nouveau. La difficulté de l'étude d'un tel sujet tient, ainsi que nous l'avons dit, à son caractère "pluridisciplinaire" mais aussi à la complexité du processus de fragmentation lui-même. En effet, si de nombreuses sous-disciplines de physique, comme la physique des particules, la physique atomique, la physique de la matière condensée, la physique nucléaire ou l'astrophysique... s'intéressent au processus de fragmentation, le concept fait encore actuellement l'objet de controverses.

En l'état actuel des choses créer un groupe de réflexion intégrant autour de ce thème plusieurs communautés constitue pour les physiciens un projet ambitieux. Pour qu'une structure aussi hétérogène puisse fonctionner avec le maximum d'efficacité, il est nécessaire que ses membres puissent acquérir une vision exhaustive et synthétique des différentes recherches menées sur la fragmentation dans des sous-disciplines qu'ils connaissent mal et avec lesquelles les contacts sont souvent difficiles. Il est donc

important de faire une synthèse de l'état actuel des recherches pour déterminer les différents thèmes développés autour de la fragmentation, évaluer leur degré de développement afin de juger de l'opportunité et de la possibilité de la mise en commun des savoirs et savoirs-faire acquis dans les différentes sous-disciplines de physique.

Dans un tel contexte, une analyse de type "veille stratégique" offre aux chercheurs un panorama des travaux menés sur ce thème, ce qui devrait permettre d'intégrer au groupe de recherche les équipes les plus impliquées dans le domaine de la fragmentation. **La veille stratégique ne cherche cependant pas à se substituer au travail "d'évaluation" de la recherche qui incombe aux scientifiques eux-mêmes. Elle propose aux chercheurs des outils d'analyse, des clés d'accès à une meilleure caractérisation des différents domaines d'activité.**

### III Constitution du corpus de données

De façon générale, en bibliométrie "...la partie la plus délicate, le maillon le plus faible de la chaîne, est la constitution du corpus de travail à partir des banques de données qui représentent la matière première."<sup>8</sup>. Quelque soit le traitement ultérieur appliqué au corpus de données, force est de constater que les résultats ne sont que le reflet du corpus de références bibliographiques initialement sélectionné. Dans le contexte de l'étude présentée ici, la constitution du corpus pose de nombreux problèmes, et il est indispensable d'en évaluer les biais éventuels. Le choix de la base de donnée, l'équation d'interrogation, le choix des champs de la base interrogée sont les principaux biais qui interviennent à ce stade du travail.

#### 1) *Choix des sources*

Le sujet à traiter est très peu tourné vers la recherche appliquée ou technologique, il concerne principalement des travaux en physique "fondamentale". Dans ce domaine, *Physical abstract* possède une très bonne couverture. Il recense plus de 4200 revues scientifiques de physique, parmi les plus importantes dans ce domaine. Le choix de la base de données bibliographique *INSPEC* de *Physical abstract* spécialisée en électronique, informatique, physique et technologie de l'information s'avère donc pertinent et présente entre autres avantages celui d'être en accès libre sur CD-Rom à la bibliothèque centrale de l'Université Paul Sabatier. Cela a facilité considérablement la

mise en place du corpus de données. L'interrogation de la base *INSPEC* a porté sur les 4 dernières années, soit de 1990 à 1993 inclus.

## 2) Equation d'interrogation

### 2a) Regard sur l'indexation documentaire dans le cas de la fragmentation

La disparité des recherches menées autour de la fragmentation par les différents groupes de physiciens d'appartenances variées témoigne de la difficulté d'établir pour le moment un consensus. Dans chaque sous-discipline de physique, la problématique se pose différemment selon l'objet physique étudié. Les méthodes d'analyse et la terminologie sont spécifiques à chacune d'entre elles. Cela est caractéristique d'un sujet de recherche en phase de développement.

Pour la physique des particules ou la physique nucléaire par exemple, l'intérêt porté à la fragmentation ne remonte qu'à ces 5 dernières années tout au plus. Dans un congrès récent D. H. E. Gross commence sa contribution ainsi : "During the last two years multifragmentation of hot nuclei has become a well established phenomenon..."<sup>9</sup>. La formulation et l'emploi de termes spécifiques pour décrire un phénomène marque une étape importante du développement de la recherche. Les travaux s'inscrivent dès lors dans un processus de maturation qui n'est certes pas linéaire, fait de controverses et de consensus ponctuels mais cependant caractérisés par la reconnaissance de l'objet comme objet d'étude à part entière. On observe alors un nombre croissant de travaux consacrés à ce nouvel axe de recherche. Peu à peu, ces travaux font l'objet d'une normalisation et d'une caractérisation de plus en plus spécifique. Au niveau de la démarche documentaire cela conduit à terme, à une modification de l'indexation des banques de données bibliographiques par l'introduction de nouveaux termes contrôlés et de nouveaux codes de classification.

Actuellement, les termes d'indexation contrôlés et les codes de classification attribués au thème de fragmentation, dans la mesure où ils existent, demeurent particuliers à chaque sous-discipline. Le premier niveau de classification de la banque de données *INSPEC* génère 10 classes thématiques qui correspondent à ce que nous avons appelé dans ce texte les "sous-disciplines" de physique (voir annexe 1). Dans chacune de ces classes les termes caractérisant la fragmentation sont différents et ne font pas toujours référence au concept lui-même. Ces termes peuvent être des termes descriptifs de l'état du système après fragmentation; par exemple, "grain size" et "droplet" en physique de la matière condensée. Ou bien ils peuvent faire référence à la technique utilisée pour fragmenter; par exemple "mechanical alloying" en physique des matériaux. Ce n'est qu'en

1992 qu'un code de classification spécifique au processus de fragmentation lui-même apparaît. Ce code est *A2570N* et correspond à "*fragmentation and relativistic collisions*". Il est intéressant de noter que c'est uniquement en physique nucléaire qu'apparaît dans la classification d'*INSPEC* une terminologie voisine du terme générique "fragmentation". Par ailleurs, les autres sous-disciplines ne bénéficient pas encore de termes véritablement spécifiques.

### *2b) Les termes choisis pour l'équation d'interrogation*

Pour constituer l'équation de recherche, il faut inventorier, dans chaque secteur de recherche, non seulement les termes utilisés par les physiciens mais aussi les types de travaux dans lesquels les recherches sur la fragmentation s'inscrivent. Un important travail a été engagé avec les experts de chacune des sous-disciplines concernées par cette étude. Une réflexion a été menée pour caractériser globalement le concept de fragmentation afin de recenser l'ensemble des travaux pouvant relever de l'étude de ce phénomène. Cela a permis d'établir dans chaque sous-discipline une liste de termes correspondant plus directement à ce concept, de même que des termes "secondaires", reflets du développement d'une activité liée à ce concept (voir tableau 1). Par exemple, le terme de "percolation", qui est un terme général, semble cependant caractéristique des travaux relatifs à la fragmentation dans le domaine de la physique théorique.

L'équation d'interrogation constituée avec les experts a confirmé que si le phénomène de fragmentation semble être de plus en plus reconnu comme "sujet d'étude à part entière", il se caractérise cependant par une divergence de points de vue qui lui interdit pour l'instant l'acceptation unanime d'une terminologie commune aux différentes sous-disciplines. L'indexation des documents fait donc nécessairement l'objet de variations importantes qui dépendent du jugement d'appréciation de l'indexeur lui-même. Y. Courrier qui a travaillé sur les problèmes posés par l'opération d'indexation en dit ceci: "...deux indexeurs choisiraient pour le même document très peu de descripteurs identiques, parfois moins de 50%. Ce sera aussi le cas pour un même indexeur à deux périodes différentes"<sup>10</sup>.

**Tableau 1 : Termes utilisés dans l'équation d'interrogation**

Fragment ?	Rupture	Nucleation
Multifragment ?	Large scale structure	Percolation
Mass distribution	Disaggregation	Decrepitation
Charge distribution	Explos ?	Jet
Size distribution	Break up	Intermittency
Mass spectra	Spinodal	Droplet
Hadronization	Disruption	Mechanical alloying
Velocity distribution	Multiplicity	Ball milling
Sputtering	Statistical distribution	Particle size
Rate equation	Disassembly	Power law

*2c) Le corpus de références bibliographiques*

La base bibliographique *INSPEC* a 36 champs qui caractérisent le contenu du document dont les principaux sont : le titre, "TI" (Title); le résumé, "AB" (Abstract); les termes contrôlés, "IT" (Controlled Term); les codes de classification, "CC" (Classification Code); les termes non contrôlés, "FT" (Free Term). La classification et les termes contrôlés correspondent à une indexation moins facilement évolutive au sens où ils font référence à un répertoire de termes prédéterminés. Restent le titre et le résumé, qui sont le plus souvent le fait du chercheur lui-même et qui traduisent peut-être plus directement l'évolution des recherches en cours. Le logiciel disponible à la bibliothèque de l'université Paul Sabatier permet de faire une recherche multichamps.

Notre souci, à ce stade, est d'éviter un silence qui serait fatal à cette étude. Afin de s'assurer de la pertinence et de l'exhaustivité des références recueillies par interrogation de la base *INSPEC*, nous avons procédé à un premier télédownload correspondant à une équation initialement constituée. Nous avons extrait quelques références de ce premier télédownload, caractéristiques de chacune des sous-disciplines concernées, pour examiner le vocabulaire couramment utilisé par les chercheurs. Dans ces quelques références, nous avons analysé les champs : "IT"; "FT", "TI" et "AB". Cette analyse a mis en évidence que dans le cas des publications s'inscrivant dans un axe original, l'indexation contrôlée est peu précise et reste très générale. Dans ces cas ce sont les champs "non contrôlés" : "TI"; "AB", "FT" qui sont

véritablement porteur d'information. L'analyse des références extraites de ce premier téléchargement a ainsi permis de répertorier un certain nombre de mots nouveaux que nous avons introduits dans l'équation initiale pour procéder à une nouvelle interrogation. Nous avons ainsi continué par itérations successives jusqu'à l'obtention d'une saturation satisfaisante; c'est-à-dire jusqu'à ce que l'apport de nouveaux termes n'engendre pas une augmentation significative du nombre de références. Etant donné l'ampleur du domaine à étudier (8 sous-disciplines ont été répertoriées), cette démarche nous assurait un minimum de silence. De plus, nous avons utilisé dans la plupart des cas des mots tronqués ainsi que notre logiciel nous le permettait (voir tableau 1). Dans le tableau 1 le point d'interrogation symbolise la troncature opérée; par exemple, "Explos ?" recouvre "exploded" ou "explosion". Enfin il est important de préciser que l'interrogation a porté sur l'ensemble des champs indexés de la base *INSPEC*.

Pour limiter notre corpus aux références bibliographiques significatives et éviter un bruit trop important nous avons filtré notre corpus en couplant l'équation de recherche avec les termes de la classification correspondant aux domaines de physique (c'est-à-dire les sous-disciplines) susceptibles de développer des études dans le champ de la fragmentation.

Notre étude, restreinte à quatre années de publication, produit un corpus de références extrêmement important. Au niveau mondial de 1990 à 1993 il représente plus de 10 000 références. Dans la mesure où cette étude s'adressait à des physiciens dont le but est de constituer un groupement de recherche du CNRS limité à un recrutement de scientifiques français nous avons pu réduire notre travail à l'activité française. Nous avons ainsi téléchargé 2267 références bibliographiques.

#### **IV Traitement de préparation**

Nous n'avons interrogé qu'une base bibliographique, notre corpus présente donc la configuration d'un fichier téléchargé homogène. Ceci facilite grandement les traitements ultérieurs. Dans la base *INSPEC*, les champs sont nombreux mais parfaitement répertoriés, la saisie est régulière et toujours présentée dans un ordre identique. Cette uniformité comporte des avantages évidents d'exploitation lors du passage au traitement automatique.

Le traitement et l'analyse des données a été réalisé au CRRM à Marseille. Pour les opérations de reformatage nous avons utilisé les logiciels commercialisés par

I+K (Infotrans). Ce travail préliminaire au traitement des données est crucial et délicat. Il s'agit de créer un corpus parfaitement uniforme pour être directement exploitable par analyse automatique des données. Certains champs sont sélectionnés puis "épurés" pour ne comporter que des éléments bibliographiques directement exploitables. Nous avons ainsi constitué un "nouveau" corpus dont les références étaient totalement normalisées. Elles comportaient uniquement les champs documentaires : auteur "AU"; termes contrôlés "IT" et termes non contrôlés "FT", suffisants aux besoins de notre analyse.

En résumé, par interrogation d'*INSPEC*, nous avons constitué, un corpus de références bibliographiques en élargissant notre recherche à tous les champs indexés dans cette base, puis, nous avons réduit et normalisé chacune des références aux seuls champs "IT", "AU" et "FT".

## **V Méthode d'analyse**

### *1) Sa pertinence par rapport à la question posée*

Plusieurs méthodes d'analyse de données s'offraient à nous pour déterminer les caractéristiques de notre corpus documentaire. Or les physiciens demandent l'élaboration d'une cartographie des différents groupes de recherche français travaillant directement ou indirectement sur le phénomène de fragmentation, ainsi que les problématiques de recherche abordées par ces groupes. L'objectif est de déterminer les collaborations constituant des groupes de travail autour de thèmes bien définis. Pour cela nous avons choisi de travailler sur le champs "AU", ce qui nous permettait de constituer les groupes d'auteurs collaborant et publiant ensemble sur des thèmes de recherche communs.

### *2) Présentation de la méthode*

L'idée est de former des groupes d'auteurs ayant publié ensemble. Les groupes d'auteurs sont constitués par liens successifs. Deux auteurs sont liés l'un à l'autre s'ils ont co-publié au moins deux fois ensemble en quatre ans. Ce seuil imposé à deux publications co-signées assurait l'existence d'une collaboration "raisonnable" entre ces auteurs. Nous obtenons ainsi des groupes d'auteurs ayant tous publié au moins deux fois avec un auteur du même groupe. Pour 2267 références initialement traitées nous avons obtenu 212 groupes pouvant être constitué de 2 auteurs à 25 auteurs et 2 groupes plus importants constitués de 41 et 450 auteurs. Il reste après traitement 386 "formes isolées" qui représentent les auteurs n'ayant jamais publié au moins deux fois avec un

autre auteur du corpus initial. Par ailleurs, pour éliminer trop de bruit et éviter de constituer des groupes trop grands nous n'avons pris en compte que les auteurs ayant publié au moins deux fois dans la période considérée (soit quatre années). Cela se justifie si l'on fait l'hypothèse qu'un chercheur véritablement impliqué dans l'étude du phénomène de fragmentation aura produit au moins deux publications en quatre ans dans ce domaine. Nous avons donc éliminé les auteurs à fréquence 1, c'est-à-dire n'apparaissant qu'une fois dans le corpus des 2267 références. Dans un deuxième temps, pour chacun des groupes d'auteurs ainsi constitué, nous avons associé à chaque auteur du groupe, un terme contrôlé, "IT" et un terme non contrôlé "FT". Les "IT" ou les "FT" sont associés automatiquement à chaque auteur d'un groupe dès lors qu'ils sont cités au moins deux fois dans les publications de cet auteur.

Cette méthode d'analyse du corpus a été choisie pour sa pertinence par rapport à la question posée par les physiciens, mais aussi pour sa simplicité. En effet les données ne subissent pas des transformations importantes pouvant introduire des biais souvent difficiles à évaluer. Cette méthode fondée sur le traitement statistique des fréquences de coapparition de noms d'auteurs échappe au problème crucial de la validité sémantique d'association statistique de mots. Ici, la constitution d'un groupe est fondée sur la collaboration entre auteurs, aboutissant à une publication. De plus, cette méthode n'impose pas de limite de taille de groupe, le critère de troncature est le nombre minimum de co-publications entre deux auteurs, qui est imposé ici à deux. Le risque était grand cependant d'obtenir des groupes beaucoup trop importants, par chaînage successifs. L'expérience faite ici est de ce point de vue particulièrement concluante. D'une part nous obtenons des groupes de taille très satisfaisante, de 2 à 25 auteurs par groupe, d'autre part, l'analyse des termes contrôlés et des termes non contrôlés associés à chaque groupe représente effectivement des thèmes de recherche bien définis, "scientifiquement homogènes". Ces points seront discutés en détail lors de la présentation et l'analyse des résultats.

## **VI Analyse des résultats**

La qualité du traitement effectué s'est mesuré à l'homogénéité "scientifique" des "IT" associés à chacun des groupes. En effet, les résultats montrent que les "IT" correspondant à un groupe d'auteurs sont bien spécifiques d'une problématique de recherche qu'il est aisé d'identifier. Ainsi une rapide analyse des "IT" de l'ensemble des 214 groupes a permis aux experts de supprimer les groupes "hors sujet"; 52 groupes ont

été ainsi conservés sur lesquels nous avons approfondi notre analyse. Deux remarques s'imposent à ce stade. Tout d'abord il faut préciser que en plus des 52 groupes conservés, une quarantaine d'autres groupes ont été identifiés comme correspondant à une problématique ne relevant que très marginalement de la fragmentation. Ces groupes marginaux n'ont pas été conservés dans la suite de cette analyse. Cependant, les experts sont convaincus de l'intérêt de ces groupes et pensent pouvoir en faire une utilisation ultérieure. Par ailleurs, le nombre de groupes effectivement retenu peut paraître faible par rapport au nombre de groupes constitués (52 sur 214). Ce résultat n'a rien de surprenant, il reflète les précautions prises au moment de la constitution du corpus pour minimiser le silence, ce qui dans cette étude était crucial, comme nous l'avons mentionné précédemment.

Seuls les deux groupes les plus importants, contenant 41 et 450 auteurs, sont "pollués". En effet, les "IT" qui leur sont associés correspondent à des sous-disciplines de physique totalement différentes, ils ne reflètent pas un thème de recherche unique. Ces deux groupes ont été retraités "manuellement".

#### *1) Le cas des deux groupes "pollués"*

Pour retraiter ces 2 groupes nous nous sommes fondés sur la liste des "IT" associée à chacun d'eux. Le groupe contenant initialement 41 auteurs a été réduit par analyse des "IT" à 36 auteurs significatifs d'un même domaine de physique. La pollution dans ce cas est due à des auteurs homonymes. Pour le groupe contenant initialement 450 auteurs le problème est double. D'une part, comme dans le cas précédent, le biais introduit par l'homonymie apparaît très nettement et à plusieurs reprises, d'autre part, les auteurs à forte fréquence (la fréquence pour un auteur donné est le nombre de fois que cet auteur apparaît dans le corpus des 2267 références) sont des chercheurs appartenant à deux laboratoires Caennais directement dépendants de l'accélérateur d'ions lourds de Caen, le GANIL (Grand Accélérateur National d'Ions Lourds). Le GANIL est un laboratoire d'accueil qui reçoit de nombreuses équipes nationales et internationales. Les 2 laboratoires Caennais, qui comportent eux-mêmes un grand nombre de chercheurs, environ 40 au total, établissent donc de nombreuses collaborations. Par ailleurs, les chercheurs "extérieurs" venant travailler ponctuellement sur l'accélérateur d'ions lourds développent eux-mêmes d'autres projets d'études indépendamment du GANIL. La situation particulière de ces deux laboratoires travaillant auprès d'un accélérateur national ouvert aux équipes étrangères explique l'étendue du réseau de collaboration.

Pour analyser ce groupe de 450 auteurs nous n'avons retenu que les paires d'auteurs à forte fréquence, c'est-à-dire des fréquences 15 à 6 inclus. Une fréquence de paires est le nombre de fois que 2 auteurs ont publié ensemble. Cette opération a permis

de restreindre le groupe à 70 auteurs. A partir de ces 70 auteurs et de leur fréquence de paire nous avons constitué le graphe des paires<sup>11</sup>. Cette procédure a généré 9 nouveaux groupes. Chaque groupe ainsi formé comporte désormais des auteurs dont les "IT" sont "scientifiquement homogènes". Il est cependant important de noter que le critère de troncature n'est plus, pour ce groupe, de 2 mais de 6 (ne font parti d'un groupe que les auteurs ayant publié au moins 6 fois avec un autre auteur du groupe).

En conclusion le traitement de ces deux groupes à 41 et 450 auteurs a engendré 10 nouveaux groupes contenant de 36 à 2 auteurs. Désormais les thèmes de physique associés à chacun de ces 10 groupes sont parfaitement identifiables. Nous considérons donc dans la suite de cette analyse, un ensemble de 60 groupes (50 + 10), bien que les modes d'identification de ces groupes diffèrent, comme nous venons de l'indiquer.

## *2) Détermination des secteurs d'activité*

L'analyse des "IT" des 60 groupes d'auteurs a conduit à la détermination des différentes sous-disciplines impliqués dans la "fragmentation" ainsi qu'à leur classement (voir tableau 2). Il ne faut pas oublier que la répartition en groupe présente une certaine inhomogénéité au niveau de l'interprétation. En effet, dans une discipline fortement structurée et avec des équipements lourds et honorables, comme dans le cas de la physique nucléaire expérimentale, les groupes rassemblent un nombre de chercheurs important et correspondant donc à un investissement de recherche conséquent. Par contre dans une discipline plus "légère", comme la physique des matériaux, les équipes sont généralement plus réduites. Ainsi les 2 sous-disciplines qui contribuent le plus en nombre de groupes et de façon similaire, c'est-à-dire la physique nucléaire et la physique des matériaux, ne représentent pas nécessairement des poids scientifiques comparables. Il serait intéressant d'exploiter d'autres indicateurs, tels que le nombre de publications par sous-discipline, le taux de citations des auteurs selon les sous-disciplines ou les budgets affectés à ces groupes de collaboration, ce qui permettrait d'analyser plus finement les contributions relatives des sous-disciplines.

**Tableau 2 : répartition des groupes d'auteurs par sous-discipline**

Sous-disciplines de physique	Nombre de groupes A	Nombre de groupes B	TOTAL	Pourcent %	Nombre d'auteurs
Physique des matériaux	14	0	<b>14</b>	23	51
Physique nucléaire	7	5	<b>12</b>	20	137
Physique théorique	9	1	<b>10</b>	17	33
Physique de la matière condensée	8	0	<b>8</b>	15	53
Astrophysique	6	0	<b>6</b>	10	31
Physique moléculaire	2	3	<b>5</b>	7	22
Physique "macroscopique"	3	0	<b>3</b>	5	7
Physique des particules	2	0	<b>2</b>	3	8
<b>TOTAL</b>	51	9	<b>60</b>	100	342

Groupes **A** : groupe d'auteur dont la fréquence de co-publication est au moins égale à 2.

Groupes **B** : groupe d'auteur dont la fréquence de co-publication est au moins égale à 6.

Nombre d'auteurs : représente le nombre total d'auteurs obtenu par sous-discipline pour l'ensemble des groupes de cette sous-discipline.

### 3) Regard sur les termes contrôlés

Pour chacune des 8 sous-disciplines, les listes des termes contrôlés ont été établies (voir annexe 2). Le nombre de termes contrôlés par sous-discipline varie considérablement. Ce nombre passe de 5 pour la physique des particules à 84 pour la physique de la matière condensée (voir tableau 3). Faisons abstraction pour le moment de la physique "macroscopique" et la physique théorique qui sont les 2 seules sous-disciplines ne se référant pas à une physique spécifique, ce qui n'est pas le cas pour les 6 autres sous-disciplines.

Pour la physique des particules, la physique nucléaire et l'astrophysique le nombre d'"IT" associés n'est respectivement que de : 5, 20 et 20. Ce résultat n'est pas surprenant et reflète une certaine réalité. En effet, ces physiques correspondent à un domaine de recherche où l'objet d'étude est bien identifié. Dans chacun de ces 3 cas, les différentes recherches développées sur le thème de la fragmentation s'inscrivent dans des axes convergents et très similaires. Par ailleurs chacune de ces 3 communautés est très structurée et la communication entre chercheurs est très soutenue. Enfin, pour ces 3 sous-

disciplines le pourcentage d'"IT" utilisés par au moins 2 groupes d'auteurs varie de 35% à 45%. Ce pourcentage élevé confirme dans ces 3 cas l'utilisation d'une terminologie convergente.

En revanche la physique des matériaux, la physique de la matière condensée et la physique moléculaire sont des sous-disciplines caractérisées par un champ de recherche très étendu. Les objets d'étude dans chacun de ces domaines sont divers. Les communautés sont donc relativement "éclatées" et développent des problématiques de recherche variées. Le nombre atteint de 49 "IT" en physique moléculaire, pour seulement 5 groupes, en témoigne (voir tableau 3). De plus, inversement au cas précédent, le pourcentage d'"IT" communs à au moins 2 groupes est faible pour ces trois sous-disciplines, il n'est au maximum que de 27%.

**Tableau 3. Termes contrôlés communs à au moins 2 groupes d'auteurs**

Sous-disciplines	groupes	"IT" total	"IT" communs	pourcent %
Physique des matériaux	14	62	17	27
physique nucléaire	12	20	9	45
Physique théorique	10	41	8	20
Physique de la matière condensée	8	84	14	17
Astrophysique	6	20	7	35
Physique moléculaire	5	49	8	16
Physique générale	3	9	0	0
Physique des particules	2	5	2	40
<b>Total</b>	60	290	<b>65</b>	

La disparité du nombre d'"IT" dans les sous-disciplines peut correspondre aux différences de structure des communautés scientifiques, différence que nous évoquons à propos de la répartition des groupes. Dans les disciplines fortement structurées, de physique lourde (physique nucléaire, physique des particules) les physiciens ne participent en général qu'à une seule thématique de recherche. Dans une discipline plus légère (physique des matériaux, physique de la matière condensée) les physiciens peuvent être impliqués dans plusieurs thématiques de recherches. Dans ce cas

le chaînage par nom d'auteurs, va introduire outre des termes propres à la fragmentation des termes qui n'y sont pas nécessairement liés.

## VII Conclusion

La méthode d'analyse utilisée ici a permis de répondre dans de très bonnes conditions à l'attente des physiciens. Ce travail a suscité un vif intérêt de la part des chercheurs impliqués dans la création du GdR.

Il représente pour les experts une base de travail. En effet, parmi les 60 groupes d'auteurs que cette analyse a permis d'identifier, plusieurs étaient a priori inconnus des physiciens chargés de constituer le GdR. Ainsi d'autres laboratoires ont pu être contactés, ce qui permet éventuellement d'élargir le champ des thématiques proposées pour le GdR et d'inclure de nouvelles équipes de recherche. Il est difficile à l'heure actuelle de conclure définitivement, les contacts sont en cours. Ce qui ne laisse aucun doute cependant, c'est le fait que cette analyse bibliométrique a effectivement ouvert de nouvelles pistes d'investigation aux chercheurs.

Du point de vue de la bibliométrie l'analyse développée est fort concluante. Les groupes constitués sont parfaitement homogènes sur le plan de la thématique scientifique, seuls 2 groupes sur 52 sont pollués et ont nécessité un traitement supplémentaire. Pour les autres, les logiciels développés au CRRM ont permis d'effectuer un travail totalement automatisé. Constituer des groupes sur le seul critère de la co-publication d'auteurs, et ne faire intervenir que dans un deuxième temps les termes contrôlés significatifs du thème de recherche de chacun des groupes est une méthode qui introduit peu de biais statistiques ou sémantiques et conduit à des résultats très satisfaisants.

Il est important de préciser qu'il n'est pas question à ce stade d'introduire des connaissances sur l'état du débat scientifique. Ce travail est avant tout une "grille de lecture" pour les physiciens dont l'objectif est de faire une synthèse de l'état des recherches en fragmentation. "Il convient donc de prendre les résultats pour ce qu'ils sont: des indices et non des preuves. Les méthodes bibliométriques... sont des méthodes d'évaluation et non des méthodes de mesure"<sup>12</sup>.

- 1 . "Maîtriser l'information critique",  
F. Jakobiak, *Les éditions d'organisation*, Paris, 1988
- 2 . "Pratique de la veille technologique"  
F. Jakobiak, *Les éditions d'organisation*, Paris, 1991
- 3 . "De l'information documentaire à la veille technologique pour l'entreprise :  
enjeux, aspects généraux et définitifs"  
F. Jakobiak H. Dou  
dans "La veille technologique", H. Desvals, H. Dou (sous la direction de),  
*Dunod*, Paris, 1992, p 1 à 46
- 4 . "Veille technologique et bibliométrie : concept, outils et application"  
H. Rostaing, *Thèse* soutenue à l'université d'Aix-Marseille III, Janv 1993
- 5 . "La graphique et le traitement graphique de l'information"  
J. Bertin, *Flammarion*, Paris, 1992
- 6 . "Citation indexing-its theory and application in science, technology and  
humanities"  
E. Garfield, John Willey & son, New York 1979
- 7 . "Information stratégique en chimie. Analyse topologique automatique de la base  
Chemical abstract"  
H. Dou, P. Hassanaly, L. Quoniam, A. La Tela, *Revue Française de Bibliométrie*  
Vol 7 1990
- 8 . "Du corpus documentaire à l'interprétation des résultats de l'analyse de  
données"  
C. Dutheil, *Revue Française de bibliométrie appliquée*, n° 6, Fev 1990
- 9 . "Proceeding of the International Workshop on Dynamical Fluctuations and  
Correlations in nuclear collisions"  
D. H. E. Gross, *Nuclear Physics A545*, 1992 p 187c
- 10 . "Analyse et langage documentaire"  
Y. Courrier, *Documentaliste-Sciences de l'information*, Vol 13, 1976

**11.** "Veille technologique et information documentaire"

H. Dou, P. Hassanaly, L. Quoniam, A. La Tela, *Documentaliste-Sciences de l'information*, Vol 27, n° 3, 1990

**12.** "Bibliométrie et scientométrie en France, état de l'art",

C. Dutheil, *Documentaliste-Sciences de l'information*, Vol 29, n° 6, 1992

## Annexe 1. Classification INSPEC

Code de classification en physique	Terme contrôlé correspondant	Sous-discipline de physique correspondante
A00	Général	Physique "macroscopique"
A10	The physique of elementary particles and fields	Physique des particules
A20	Nuclear physics	Physique nucléaire
A30	Atomic and molecular physics	Physique moléculaire
A40	Classical areas of phenomenology	
A50	Fluids, plasmas and electric discharges	
A60	Condensed matter: structure, thermal and mechanical properties	Physique des matériaux
A70	Condensed matter : electronic structure, electrical, magnetic and optical prop.	Physique de la matière condensée
A80	Cross-disciplinary physics and related areas of science and technology	
A90	Geophysics, Astronomy and Astrophysics	Astrophysique

Remarque : La sous-discipline que nous avons nommé "physique théorique" dans ce texte, n'est pas introduite dans ce tableau. En effet, la physique théorique correspond à l'ensemble des disciplines.

## Annexe 2. Termes contrôlés communs à au moins 2 groupes

### Physique nucléaire

Termes contrôlés / Groupes	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
Direct reaction and scattering							1	3				2	6
Fission								9	9			2	20
Heavy ion-nuclear reaction	29	5	2		13	3	26	32	9	2	3	2	126
Nuclear charge					10		2						12
Nuclear matter	20						1	25	3				49
Nuclear react and scat theo								18	1			2	21
Proton production	8							3					11
Statistical theory of nuclear reaction and scattering	24			3	10	3	19	27	1	2	3	2	94

### Physique de la matière condensée

Termes contrôlés / Groupes	1	2	3	4	5	6	7	8	TOTAL
Amorphisation	3				2				5
Carbon		5				1			5
Ceramics	2		5						8
Cristal atomic struc. of inorg. comp.	4		2					2	8
Drops		4			3				7
Ferromagnetic of substance	2					5		3	5
Heat transfer		1				1			2
Lithium	1		3						4
Magnetic properties of fine particles	6							2	8
Nucleation	4	1					1	1	7
Powder metallurgy	2				1				3
Powder technology	1		3						4
Silicon	4		3	2		5			14
Silicon	4		3	2		5			14

## Physique théorique

Termes contrôlés / Groupes	1	2	3	4	5	6	7	8	9	10	TOTAL
Aggregation in substances		1		3							4
Critical phenomena						1			2		3
Fluctuation		2						1			3
Fractal		1	1	2				2		2	8
Lattice theory and statistics								4		2	6
Percolation	1		2	4	4	2		4	2	2	21
Phase transformation		1						1			2
Random process				2		2		14		2	20

## Physique des matériaux

Termes contrôlés / Groupes	1	2	3	4	5	TOTAL
Aluminium			2	2		14
Barium						5
Ceramics		2		2	2	14
Densification				2		3
Grain growth						3
Heat treatment	2					4
Iron	9					20
Magnetisatic behavior	1					16
Mechanical alloying	4					6
Particle size	2					4
Powder technology	6	4		2	5	27
Scanning e- microscop exam of mat	3					4
Silicon	2			1		5
Sintering	3				2	9
Transm e- microscopic exam of material	2					8
X-ray diffract examination of material	4					9
Zirconium	1					8

## Physique moléculaire

Termes contrôlés / Groupes	1	2	3	4	5	TOTAL
Antimony			1		5	6
Atomic clusters			9	6	4	19
Ionisation potential			8	1	1	10
Metal cluster			9	5	5	19
Molecular clusters				1	1	2
Molecular photoionisation			9	2		11
Sodium			3	5		8
Time of flight mass spectra			13	2		15

## Astrophysique

Termes contrôlés / Groupes	1	2	3	4	5	6	TOTAL
Astrotechniques	2					2	4
Cluster of Galaxies	2	2	5	2	8	2	21
Cosmology			2	1	5	1	5
Galaxies		2	1	1	4		8
Red-Shift	1				4		5
X-ray astrophysique observation			1	2			3
X-ray source			4	2			6

## Physique des particules

Termes contrôlés / Groupes	1	2	TOTAL
Elementary particle inclusive interaction	1	3	4
Parton model	1	4	5