

DES AXES DE RECHERCHE POUR UN PROGRAMME INFOMETRIQUE DE VEILLE SCIENTIFIQUE ET TECHNIQUE

W.A. TURNER
CERESI/CNRS
B.P. 132, 75960 Paris Cedex 20.

RESUME

La veille scientifique et technique est à la mode. Les décideurs semblent de plus en plus convaincus que la position concurrentielle du laboratoire, de la firme ou de l'institution qu'ils dirigent en dépendent. Les producteurs de banques de données peuvent espérer de ce fait conquérir des nouveaux marchés pour leurs produits d'information. Quant aux documentalistes, ils sont appelés à acquérir des nouveaux savoir-faire, ainsi que témoigne la multiplication de cours visant à les initier aux techniques de la bibliométrie. Cependant, toute mode est sujette à caution. Elle implique qu'après une période initiale d'engouement, l'intérêt tombe et les investissements se font ailleurs. C'est la raison pour laquelle nous nous proposons dans cet article de considérer la question suivante : la veille scientifique et technique est-elle une mode ou un axe de recherche légitime?

ABSTRACT

The need to monitor scientific and technological trends is now generally recognized. Techniques designed to carry out bibliometric analysis of large scientific and/or patent databases are now being widely developed. Decision-makers are showing considerable interest in experimenting their usefulness for policy planning purposes. Database producers can thus expect new markets for their information products. Documentalists will have to acquire new skills, as it appears through the numerous bibliometric courses now taught to them. But as every fashion, after an initial burst one can expect a loss of interest and a reorientation of investments in other areas. In this paper, we will attempt to identify the basis for transforming current interest into a full-fledged, long-term research program.

INTRODUCTION

La compétitivité est de plus en plus comprise comme une question d'écoute: écoute des marchés, des technologies et des recherches scientifiques, mais aussi écoute des hommes qui, par leurs stratégies individuelles, font le lien entre ces différentes considérations et déterminent ainsi collectivement les voies d'innovation de demain (Riboud, 1989). D'un côté, la collecte des informations;

de l'autre côté, des structures de prise de décision. Entre les deux, une série de traitements visant à mettre en forme l'information pour aider à raccourcir les délais entre l'analyse d'une situation et l'adoption d'un plan d'action.

La veille scientifique et technique peut être définie en termes d'une recherche de la meilleure adéquation possible entre l'offre et la demande d'un produit d'information. Ce produit est celui qui aidera un système productif (1) à prendre les décisions qui sont essentielles pour défendre ou améliorer sa position concurrentielle (Avenier, 1989). Afin d'évaluer l'état des recherches dans ce domaine, nous nous proposons d'adopter trois points de repères.

Le premier est conceptuel. Il s'agit de préciser la nature de la demande pour les produits de veille scientifique et technique. Etant donné que se pose le problème d'une production "sur mesure", il s'agit de savoir si on peut néanmoins identifier des catégories de besoins qui sont relativement stables. Le second repère est technique et concerne les conditions de production d'indicateurs qui sont susceptibles de satisfaire les différents besoins identifiés. Enfin, le troisième est scientifique. Un investissement dans la mise en place d'un dispositif de veille scientifique et technique se justifie en grande partie par un calcul de rentabilité. Une relation directe est supposée exister entre l'information et l'efficacité d'un processus de prise de décisions. Mais comment déterminer si cette hypothèse est fondée ou non?

I. LE CADRE CONCEPTUEL

L'identification des besoins à satisfaire dans le cadre d'un programme de veille scientifique et technique qui est fondé sur la bibliométrie suppose qu'on parvienne à :

- distinguer entre différents types de contextes décisionnels;
- préciser l'intérêt des publications pour le développement d'indicateurs permettant de décrire l'activité scientifique et technique;
- comprendre le rôle que les résultats d'une analyse bibliométrique peut jouer en vue de réduire les incertitudes liées à l'évolution rapide de l'environnement concurrentiel des systèmes productifs.

Les points de vues sont divergeants sur ces différentes questions. Nous nous efforcerons de donner un aperçu des débats en cours.

(1) La notion de système productif est employé pour désigner l'activité sociale qui consiste à mettre en oeuvre un savoir-faire spécifique pour transformer des flux captés à l'entrée du système en une série de produits qui peuvent être mis en circulation à la sortie. De ce fait, elle doit être considérée comme une notion générale, pouvant s'appliquer à la description de l'activité d'un individu qui s'efforce de valoriser ses propres savoir-faire, d'un groupe qui, comme le laboratoire de recherche, s'intéresse prioritairement au traitement des flux immatériels ou l'organisation qui, comme tout système productif, souhaite que sa production rencontre une demande.

La description du contexte décisionnel

Dans la littérature qui concerne l'évaluation de la recherche, une distinction est habituellement établie entre deux contextes de prise de décision (Chabbal, 1987). Le premier est celui qui concerne le lancement, la poursuite ou l'arrêt d'un programme de recherche. Les décisions de ce type supposent une évaluation des performances qui ont été obtenues ou qui peuvent être escomptées. Cette évaluation est entreprise dans le cadre d'objectifs clairement définis. Quant au deuxième type d'évaluation, il s'agit, au contraire, d'aider à déterminer ces objectifs dans un environnement qui se caractérise par une évolution rapide des recherches, des technologies et des marchés. Le but n'est plus de mesurer les performances mais de décider des orientations de la R et D qui sont susceptibles de contribuer à l'établissement de *la stratégie future d'une entreprise, d'un laboratoire ou d'un pays*.

Une première distinction importante est donc celle qui existe entre des évaluations de performance et des évaluations stratégiques et prospectives.

L'exploitation des bases documentaires

La signification des indicateurs tient de la représentation qu'on se fait du phénomène observé. Les indicateurs bibliométriques servent à étudier l'activité scientifique et technique. Il convient donc de poser la question du rôle que jouent les publications dans un processus de développement des connaissances. Sont-elles le point final d'un processus de recherche, le moment où les chercheurs mettent en forme de façon définitive les résultats de leur travail, afin de les présenter à leurs collègues et au monde extérieur pour évaluation (de Solla Price, 1963)? Sont-elles, au contraire, l'expression d'un "vote", un indicateur des micro-décisions qui pourraient entraîner la réorganisation de la structure thématique des recherches de demain (Turner, 1990(1)?

Ce débat est important quant à ses conséquences sur le plan statistique. Dans le premier cas, la publication est considérée indépendamment de ses conditions de production. Par exemple, le bibliométricien ne tiendra aucun compte du fait que l'acte d'écrire revient à prendre position dans un champ scientifique qui se caractérise par des conflits, des structures de domination, des écoles de pensées (Bourdieu, 1975). Il assimilera l'augmentation du nombre de publications à l'accumulation d'un stock de connaissances, chaque publication étant considérée comme la pierre qu'un individu ou une équipe, travaillant à un moment donné dans un lieu géographique précis, ajoutent à la construction de l'oeuvre collective de la science. Un élément discret, donc, qui peut être compté et agrégé en fonction des catégories d'un plan de classement, ou selon divers critères temporels, géographiques et institutionnels.

Dans le deuxième cas la situation est différente. Il s'agit de comprendre la publication comme une contribution à la dynamique d'évolution d'un champ scientifique, comme un facteur de consolidation ou de remise en cause des structures existantes. Le "vote" consiste à privilégier certains sujets qui sont jugés plus porteurs que d'autres, à s'aligner sur les positions des uns et à s'opposer aux positions des autres, à mettre dans la balance, en somme, une force de travail qui peut contribuer soit à renforcer, soit à renverser le rapport de forces dominant du champ (Latour, 1989). Le bibliométricien ne peut, dans cette optique, considérer les publications comme des éléments discrets à compter et à agréger comme il entend. Entre les publications existent des relations de ressemblance et de dissemblance qui précisent leur contribution à l'économie générale d'un champ de recherche. Il convient d'essayer de les décrire à l'aide d'outils d'analyse multidimensionnelle appropriés (Courtial, 1990).

Réduire les incertitudes du contexte décisionnel

Les bases de données constituent une source d'information sur les résultats de recherche déjà obtenus. Elles sont donc utiles lorsqu'il s'agit de comparer les résultats obtenus lors de la mise en oeuvre des programmes avec les résultats escomptés, à condition bien sûr que les délais de mises à jour des bases soient réduits au maximum (Bauin, 1990). Cependant, lors du lancement d'un programme nouveau ou de la redéfinition des contenus de recherches existants, il faut préparer l'avenir. Comment établir un lien entre ce qui a déjà été fait et ce qui doit encore être entrepris en vue de rester compétitif ou de le devenir davantage. La difficulté réside dans les incertitudes qui pèsent sur la qualité des analyses stratégiques et prospectives. Ces incertitudes sont grandement augmentées dans des situations en évolution rapide lorsque, comme c'est souvent le cas actuellement, tout change: les marchés, les savoir-faire et les technologies. Comment s'assurer que les décisions prises aujourd'hui porteront leurs fruits demain?

Il n'y a évidemment pas de recettes à fournir en réponse à cette question. Cependant, deux types de réponse méritent attention. Le premier découle de la nature des lois bibliométriques. Celles-ci donnent lieu à des distributions hyperboliques, c'est-à-dire que l'allure de la courbe obtenue est fortement asymétrique, le plus souvent ayant la forme d'une courbe en "J" (de Solla Price, 1976). L'interprétation qu'on donne à ces courbes est la suivante: un petit nombre d'items observés (les journaux, par exemple) est responsable de la production de la majorité du produit obtenu (i.e. les publications effectivement consultées) et vice versa, la grande majorité d'items considérés ne contribue que faiblement à la production de ce produit.

Quand on emploie des techniques quantitatives pour analyser une activité scientifique, on observe qu'effectivement un petit nombre d'équipes a généralement tendance à dominer le secteur par le nombre de leurs publications; ces publications sont, par ailleurs, souvent très fortement citées indiquant que

leurs travaux ont un impact scientifique reconnu. Il serait normal de conclure dans ces conditions qu'une décision judicieuse de politique scientifique consisterait à accorder à ces équipes les ressources nécessaires pour développer leurs travaux, leur succès passé étant un bon indicateur des performances à venir. Ce faisant, cependant, les moyens disponibles dans un champ de recherche se concentreraient sur un nombre limité d'élus, cette tendance entraînant des effets pervers puisqu'elle peut conduire à écarter des recherches originales et potentiellement innovatrices.

Pourquoi supposer que la performance passée est un bon indicateur de l'efficacité future? Rien ne permet de dire que les outils conceptuels et méthodologiques élaborés pour traiter un problème posé aujourd'hui vont encore être utiles pour traiter des problèmes qui se poseront demain (Mulkay, 1979). Une politique scientifique devrait être capable d'identifier et de soutenir des axes de recherches qui ne s'inscrivent pas dans les préoccupations dominantes du champ. Cependant, une telle politique comporte plus de risques. Des techniques d'analyse multidimensionnelle de données peuvent éventuellement aider à les réduire. Par exemple, les cartes stratégiques que génèrent les programmes LEXIMAPPE (2) sont fondées sur une convention de représentation destinée à classer les thèmes de recherche caractéristiques d'un fichier documentaire comme étant centraux ou périphériques par rapport à l'organisation générale du domaine étudié. Cette convention infographique vise à éviter les effets pervers que peut entraîner une politique de concentration de moyens sur un nombre limité d'élus, sans pour autant tomber dans le piège de la non-pertinence. L'objectif est d'aider à mieux comprendre les mécanismes déterminant la dynamique d'évolution d'un champ de recherche. Des décisions pourraient alors être prises en tenant compte des contraintes structurales d'une situation donnée.

II. LES TECHNIQUES DE L'ANALYSE BIBLIOMETRIQUE

La collecte des informations

Il est rarement suffisant d'exploiter des banques de données existantes pour développer des indicateurs. La qualité des résultats dépend, en premier lieu, de la construction d'une base "ciblée" ayant une couverture suffisante pour être statistiquement représentative de l'activité qu'on s'efforce de mesurer. Le problème de la construction d'un fichier de travail suppose une analyse critique et comparative des politiques de collecte documentaire mise en oeuvre par les producteurs de banques de données. Ces politiques peuvent éventuellement être

(2) Ces programmes ont été développés en collaboration par l'Ecole des Mines et le CNRS. Plusieurs articles employant cette méthode ont été publiés dans les Cahiers de l'ADEST, Numéro spécial consacré à la Scientométrie en France, CSI/Ecole des Mines, Paris, juin 1990.

la source des bruits ou des silences systématiquement reproduits au niveau de la collecte documentaire.

La création des listes d'autorité

Les banques de données bibliographiques comportent des nombreuses sources d'erreurs statistiques qui doivent être éliminées avant de pouvoir construire des indicateurs fiables. Un travail documentaire doit normalement être entrepris pour développer des listes d'autorité afin de normaliser le contenu des champs documentaires (noms de laboratoires, de journaux, etc...). Ce travail est de plus en plus un facteur de compétitivité pour les producteurs de banques de données. Avec la mise en service par les serveurs des commandes statistiques, tels que MEMTRI de Télésystèmes ou le ZOOM de l'ESA, on peut désormais envisager de faire de la bibliométrie en ligne. La part de ce nouveau marché qui reviendra à chaque producteur dépendra de la qualité de sa base, qualité qui passe par la constitution et le maintien des listes d'autorité.

L'analyse des contenus

Trois techniques de constitution des *fichiers thématiques* sont généralement employées. Il s'agit d'une interrogation à partir des codes de classement, des mots-clés (ou dans le cas du Science Citation Index, des citations) ou du texte intégral moyennant des outils de traitement du langage naturel disponibles sur les serveurs. Chaque technique a ses intérêts et ses inconvénients. La constitution des plans de classement et la réalisation d'une indexation documentaire (par mots-clés ou par citation) sont des pratiques subjectives. Il est très difficile de mesurer l'impact de cette subjectivité sur la qualité des fichiers thématiques. Le développement rapide des banques de documents primaires est peut-être une solution. Elles présentent non seulement l'intérêt d'une mise à disposition plus rapide des informations en vertu de la suppression du travail d'indexation qui retarde la mise à jour des bases bibliographiques; mais elles suppriment également le recours aux lexiques préalablement établis qui ne se recourent que partiellement avec le langage de la science en action. Cela dit, il faut bien reconnaître que les outils de traitement du langage naturel disponible auprès des serveurs sont assez rudimentaires. Ils se servent de techniques statistiques pour aider à structurer l'accès documentaire, alors qu'au niveau de la recherche, on considère que l'utilisation de descripteurs conceptuels et de règles d'exploitation sont nécessaires.

Les méthodes statistiques

Il convient de faire une distinction entre les méthodes quantitatives et les méthodes qualitatives d'analyse statistique. Les premières nécessitent une définition préalable du contenu d'un fichier thématique. Ce contenu doit en

principe correspondre aux limites du champ de recherche à analyser. Les techniques exposées dans les sections 2.1., 2.2. et 2.3. ci-dessus ont pour objectif d'assurer cette correspondance et, donc, d'éliminer les sources de bruits et de silences qui sont susceptibles de compromettre la qualité des résultats statistiques obtenus.

Dans le domaine des analyses stratégiques et prospectives, il est souvent impossible de partir d'une définition préalable des recherches menées dans un champ donné. L'objectif dans ce cas est précisément de modifier les contenus de la recherche en introduisant des nouveaux sujets d'étude, en créant des liens entre secteurs qui s'ignoraient auparavant, ou encore en favorisant l'émergence de nouveaux thèmes. On ne peut pas dire a priori quelles sont les limites d'une activité de recherche, l'objectif de l'analyse étant justement de les fixer. C'est une problématique de ce type qu'implique le recours à des méthodes qualitatives d'analyse statistique. Les statistiques de base nécessaires pour utiliser ces techniques sont le nombre de documents dans un fichier N , l'occurrence C_i de chaque objet i (les mots, les codes de classement, les citations,...) et la cooccurrence C_{ij} d'un pair d'objets. Utilisant ces statistiques, tous les indices d'association entre deux objets peuvent être calculés. La qualité du résultat suppose que l'indice utilisé pour l'analyse soit localement stable et homogène. Du point de vue des analyses qualitatives, la stabilité locale d'un indice d'association joue le même rôle que la constitution d'un fichier de test représentatif dans le cadre des analyses quantitatives (Michelet, 1988).

L'infographie

La veille scientifique suppose la possibilité de développer des indicateurs bibliométriques qui sont synthétiques, faciles à comprendre et à suivre dans le temps. Le problème infographique a une double dimension. La première est technique et concerne l'agrégation. Quelle est la perte d'information qu'entraîne le processus d'agrégation conduisant à l'adoption d'indicateurs synthétiques? La seconde est graphique : comment peut-on tirer profit de l'évolution des nouvelles technologies de l'information pour visualiser la dynamique de l'évolution scientifique et technique à partir de l'exploitation des flux documentaires?

L'infométrie

L'information est un facteur d'aide à la prise de décision. Cette hypothèse de travail n'est généralement pas remise en doute lorsqu'il s'agit d'entreprendre un programme de veille scientifique et technique. Au contraire, il semble aller de soi que la qualité des décisions prises sera fonction de la qualité des informations servant à leur élaboration. Ceci est tellement vrai qu'à l'heure actuelle le marché pour les produits de veille scientifique et technique est caractérisé par une économie de l'offre. Des investissements importants sont envisagés pour réunir

des compétences et créer des enseignements permettant de développer des dispositifs de veille scientifique et technique en France.

Cependant, on connaît les pièges d'une économie de l'offre. Les enquêtes les montrent en ce qui concerne l'industrie des bases de données, une industrie qui, elle aussi, est bâtie sur le principe qu'une offre d'information rencontre toujours une demande. Ces enquêtes convergent toutes sur le fait que les bases documentaires constituent la solution la moins employée lorsqu'il s'agit de trouver une source d'information permettant de faire avancer un projet scientifique ou technologique (Allen, 1985; Michel, 1990). On peut se demander si l'offre de produits bibliométriques dans le domaine de la veille scientifique et technique aura plus de succès. En tout cas, il est clair qu'il faut se méfier des évidences. La relation exacte entre information et prise de décision est complexe et nécessite qu'on s'efforce d'évaluer sa nature exacte. C'est l'objectif de l'infométrie : il s'agit d'essayer de déterminer les conditions d'une adéquation entre une offre et une demande d'un produit d'information.

III UNE APPROCHE SCIENTIFIQUE DES PROBLEMES DE LA VEILLE SCIENTIFIQUE ET TECHNIQUE

Les études bibliométriques donnent lieu à des comparaisons. Le problème technique posé est d'éviter qu'une variable non-contrôlée se trouve avoir des distributions systématiques dans les différents fichiers documentaires analysés pour fausser ces comparaisons. Par exemple, on sait qu'en utilisant des bases de données pour des comparaisons internationales, la langue de publication est un biais. Le nombre d'articles des pays qui ne publient pas la totalité de leurs résultats en anglais est sous-estimé par rapport à celui des pays anglophones. Mais, même en supposant qu'il soit possible de constituer des fichiers homogènes, quelles conclusions peut-on tirer d'une telle comparaison. En montrant, par exemple, que l'activité mesurée par le nombre de publications françaises est plus faible dans un secteur-clé que celle de ses principaux concurrents, qu'apprend-on? Faut-il encourager un investissement supplémentaire dans le secteur afin de remonter la pente?

Lorsque cette question est soumise aux experts, elle sert à alimenter un débat politique. Dans un tel contexte, l'intérêt de l'analyse quantitative ne découle pas de son objectivité supposée. Il réside dans la possibilité qu'elle laisse aux protagonistes de l'utiliser aussi bien pour défendre leurs propres prises de position que pour mettre en doute celles de leur adversaires (Palumbo, 1987). Ce n'est pas un facteur discriminant d'aide à la décision. Afin qu'elle le devienne, une meilleure compréhension du fonctionnement des différents systèmes de recherche comparés est nécessaire. Un nombre relativement faible de publications peut s'expliquer par des ressources insuffisantes au départ, ou par une mauvaise exploitation de celles qui sont disponibles ou, enfin, par un

manque de fiabilité au niveau du choix des indicateurs et des dispositifs de collecte de l'information mis en oeuvre.

Les raisons de la non-pertinence d'une démarche bibliométrique sont donc nombreuses. Elles trouvent leur source dans la difficulté d'interpréter des résultats observés indépendamment d'une bonne compréhension de la façon dont un système transforme les ressources dont il dispose en "outputs". Il ne suffit pas de constater la co-variance des courbes mesurant la quantité d'entrées et de sorties des différents systèmes étudiés. Il faut essayer d'élucider les conditions qui déterminent la production de publications dans un contexte donné. Partant d'une compréhension de ces conditions, on peut prendre position sur la question-clé de la veille scientifique et technique : que faut-il savoir pour agir? Le problème de l'établissement d'une meilleure adéquation entre l'offre et la demande de produits d'information suppose la formulation d'une théorie sur l'utilité de ses produits pour la définition et la mise en oeuvre des stratégies cognitives et sociales (3).

CONCLUSION

La veille scientifique et technique est-elle une mode ou un axe de recherche légitime? Il est encore trop tôt pour répondre clairement à cette question. Cependant, un danger existe: celui de bâtir un programme de recherche sur un lieu commun. Ce danger est aggravé dans le contexte actuel par un double phénomène: l'existence de nouvelles technologies et de nouvelles méthodes de traitement de l'information d'une part, et un problème grandissant pour déterminer le positionnement institutionnel des services documentaires aux sein des entreprises d'autre part. Des raisons techniques et sociologiques existent pour investir dans les programmes de veille scientifique et technique. Cependant, la justification économique des tels investissements est ailleurs, se situant plutôt au niveau de l'amélioration des processus de prise de décision. Tant qu'on n'est pas en mesure d'évaluer l'efficacité de ces investissements, le risque est grand de voir tomber l'intérêt qui existe pour la bibliométrie à l'heure actuelle. C'est la raison pour laquelle il semble urgent d'essayer de bâtir une théorie des usages de l'information dans des processus de prise de décision.

Bibliographie

Allen, T.J. *Managing the flow of technology*, M.I.T. Press, Cambridge, 1984.

(3) C'est cette prise de conscience qui a été à l'origine de la création d'une Cellule de Recherche en Sciences de l'Information (CERESI) par la Direction de l'Information Scientifique et Technique du CNRS. voir Turner 1990(2).

- Bauin, S. "Etude de faisabilité de la production d'indicateurs bibliométriques pour le CNRS et les universités", UNIPS/CNRS, avril 1990.
- Bourdieu, P. "La spécificité du champ scientifique et les conditions sociales du progrès de la raison", *Sociologie et Sociétés*, Vol. 7, No. 1, Mai 1975.
- Chabbal, R. "Organisation de l'évaluation de la recherche à la CEE", Rapport à la Commission des Communautés Européennes, DGXII, Bruxelles, 1987.
- Courtial, J.P. Introduction à la scientométrie: de la bibliométrie à la veille technologique, *Anthropos-Economica*, 1990.
- de Solla Price, D.J. "A general theory of bibliometric and other cumulative advantage processes", *Journal of the American Society for Information Science*, Vol. 27, No. 5, 1976, 292-306.
- de Solla Price, D.J. *Big Science, Little Science*, Columbia University, New York, 1963.
- Latour, B. *Science en action*, Editions La Découverte, Paris, 1989.
- Michel, J. "Deux enquêtes sur l'information dans l'entreprise", *Documentaliste-Sciences de l'Information*, Vol. 27, No. 3. mai-juin 1990.
- Michelet, B. *L'analyse des associations*, Thèse de Doctorat, Université de Paris VII, 1988.
- Mulkay, M. *Science and the sociology of knowledge*, George Allen and Unwin, Londres, 1979.
- Palumbo, D.J. (ed.) *The politics of program evaluation*, Sage Yearbooks in Political and Public Policy, Vol. 15, Sage Publications, London, 1987.
- Riboud, A. "Europe technologique, industrielle et commerciale", Rapport au Commissariat Général du Plan, La France, l'Europe. Xème Plan 1989-1992, Documentation Française, Paris, 1989.
- Turner, W.A., Michelet, B., Courtial, J.P. "Scientific and technological information banks for the network management of research", *Research Policy*, 19, North Holland, Amsterdam, 1990(1), 467-475.
- Turner, W.A. "Un programme infométrique pour le développement des recherches en science de l'information", *Documentaliste*, vol. 27, No. 3, mai-juin 1990.