

APPROCHES INFOMÉTRIQUES DE LA RECHERCHE EN ÉDUCATION

Philippe Champy

Les études bibliométriques ou scientométriques connaissent un développement croissant. Leur intérêt consiste à aider les spécialistes d'un domaine de recherche (chercheurs, experts, décideurs) à mieux appréhender dans sa structure thématique et institutionnelle, dans son environnement national et international, dans son insertion épistémologique et disciplinaire, grâce à une analyse de données fondée sur des mesures quantitatives.

Ces mesures (1) sont effectuées le plus souvent à partir d'informations bibliographiques contenues dans des banques de données informatisées (bibliométrie). Mais elles peuvent aussi prendre en considération d'autres types d'information (numériques ou financières) utiles par exemple pour cartographier les unités ou équipes de recherche et, de façon générale, pour « mesurer » l'activité scientifique (scientométrie). Les diverses méthodes de mesure d'information, à partir d'indicateurs bibliographiques ou d'une autre nature, qui concernent l'activité scientifique ou d'autres types d'activité, ressortissent à l'**infométrie**.

Le domaine de prédilection des approches infométriques appliquées à la mesure de l'activité scientifique est l'analyse de la littérature de recherche où sont diffusées les « connaissances certifiées ».

Communication documentaire

Perspectives documentaires en éducation, n° 32, 1994

Cette bibliométrie permet, entre autres, d'appréhender un contexte de recherche et d'en étudier l'évolution au fil des années (fronts de recherche, réseaux d'équipes). Pour l'heure, ce type d'étude se déploie principalement dans les domaines des sciences de la nature, de l'économie et des technologies. Or il s'agit de savoir si les méthodes éprouvées dans ces domaines peuvent être utilisées également en sciences humaines et sociales, spécialement dans le champ pluridisciplinaire de l'éducation et de la formation (2). Ce champ en pleine émergence a besoin d'être mieux reconnu et, lui-même, de mieux se reconnaître. L'infométrie peut éventuellement l'y aider. C'est l'hypothèse de départ que nous partageons avec Jacky Beillerot (cf. bibliographie).

À l'issue d'une étude exploratoire menée depuis octobre 1992 au contact notamment des séminaires de l'ADEST (3), parallèlement à notre travail de production de banques de données à l'Institut national de recherche pédagogique (cf. bibliographie), nous nous proposons de faire le point :

- 1) sur les problèmes rencontrés quant à l'importation dans le champ de la recherche en éducation et formation des approches infométriques,
- 2) sur les méthodes qui peuvent éventuellement être utilisées en tenant compte des sources disponibles, et enfin,
- 3) sur les projets qu'il pourrait être intéressant de développer à l'avenir.

Les problèmes généraux

Il est une opinion fort répandue parmi les spécialistes de scientométrie, que Rémi Barré, le directeur de l'Observatoire des sciences et des techniques (Paris), a parfaitement exprimée lors d'un séminaire de l'ADEST, selon laquelle il semble actuellement impossible de mener à bien des études scientométriques en sciences humaines et sociales. Quitte à choquer, en semblant d'entrée de jeu conclure à une impasse, nous partageons cette opinion, à la condition de préciser sa signification réelle.

Les études scientométriques canoniques, développées d'abord aux Etats-Unis, comportent schématiquement deux volets : il s'agit, d'une

part, d'analyser les références de la littérature de recherche ; d'autre part, d'analyser les notices des brevets déposés.

Le premier volet présuppose l'existence de banques de données bibliographiques qui recensent l'essentiel de la production mondiale. Or, dans les sciences de la nature précisément, ces banques existent avec un taux de couverture élevé à cause de deux phénomènes massifs qui caractérisent la littérature de recherche et sa circulation, à savoir 1°) la prééminence de l'article de périodique comme vecteur principal de la « connaissance certifiée » (ou des résultats de recherche si l'on préfère), et 2°) la domination de l'anglais comme langue de communication. En recensant systématiquement les grandes revues à comité de lecture, les banques de données spécialisées (souvent américaines) couvrent la majeure partie de la science vivante au niveau mondial selon des biais ou des lacunes qui font l'objet d'estimations rigoureuses. Le second volet est rendu également possible grâce à l'existence de banques de données nationales et internationales de brevets.

Ces deux volets d'ailleurs se complètent : le premier concerne les recherches fondamentales et appliquées, le second mesure les retombées technologiques de ces recherches, donc leur impact sur ou leur réponse à une demande économique et sociale.

Les sciences humaines et sociales (SHS) présentent une tout autre configuration que celle, ici très grossièrement décrite, d'une science mondialisée, anglophone et exposant ses résultats au rythme des périodiques. La prééminence de l'article de périodique est beaucoup moins vraie dans les SHS où l'ouvrage, la thèse, le rapport de recherche et la littérature grise demeurent des vecteurs majeurs de communication des résultats. Dans les SHS, l'anglais n'est pas non plus la langue de communication quasi exclusive puisque les chercheurs s'expriment et écrivent encore très majoritairement dans leurs langues nationales (des variations existant sans doute d'un pays à l'autre). Les aires culturelles (et linguistiques) structurent encore fortement la circulation des résultats de recherche en zones plus ou moins étanches. Les banques de données bibliographiques existantes reflètent ces réalités, rendant donc difficiles les études bibliométriques d'ampleur comparable aux études canoniques développées dans les sciences de la nature. Ainsi, lors de comptage par exemple, comment traiter sur le même plan, sans opérer de pondération (selon quels principes ?), une référence d'article et une référence de thèse ?

Comment donc résoudre la question de l'hétérogénéité de la production littéraire qui pose déjà, en amont de l'analyse, la délicate question de son recensement exhaustif ? À l'évidence, en SHS, la veille informationnelle sur les revues ne suffit pas à rendre compte de la science vivante. Comment aussi situer dans un environnement international des données qui proviennent de zones relativement cloisonnées ? Pour ces diverses raisons parmi de nombreuses autres, il n'est pas possible de prétendre « mesurer la recherche en sciences humaines et sociales » à travers la bibliographie qu'elle produit, du moins de façon globale et mondiale à l'instar des études scientométriques canoniques. La principale difficulté réside dans le manque de banques de données qui puissent constituer, pour tous les types de documents issus du travail scientifique, des sources suffisamment fiables, dans l'aire culturelle francophone notamment.

Un autre problème majeur concerne l'évaluation de l'impact de la recherche en éducation et formation sur le système d'éducation et de formation. En SHS, il n'existe pas l'équivalent des brevets qui permettraient de prendre la mesure des domaines d'application où les résultats de recherche pourraient être exploités et trouver des débouchés. Il existe diverses catégories d'« infralittérature » s'inspirant plus ou moins de la recherche qui peuvent faire l'objet d'étude, en particulier sur les références scientifiques qu'elle revendique. Il peut également exister des recensements d'innovations dont la référence à un cadre de recherche peut être plus ou moins explicite. Mais le problème semble ne pas pouvoir être traité par analogie avec ce qui se passe à cet égard dans les sciences de la nature, la source « brevets » étant absente.

De ces remarques générales, il ressort donc qu'il est impossible de réaliser en SHS, et spécialement en sciences de l'éducation et de la formation, des études similaires aux études scientométriques canoniques. À partir d'une banque de données américaine comme ERIC, il est sans doute possible d'effectuer des études bibliométriques en éducation et formation pour l'aire culturelle anglophone. C'est une hypothèse à ne pas écarter mais à confirmer (4). Pour l'aire culturelle francophone, il n'y a pas l'équivalent de cette source. Pourtant, nous pensons que certaines méthodes peuvent être appliquées sur des sources plus restreintes. Nous pensons également qu'un ensemble d'études s'appuyant sur ces sources ou créant leurs propres sources

ad hoc pourrait permettre de dégager une série de conclusions d'ordre scientométrique sur le champ considéré.

Les méthodes et les sources

Il n'y a aucune raison théorique qui permette de penser que certaines méthodes d'analyse infométrique seraient inapplicables sur des corpus de données bibliographiques (ou d'informations textuelles ou numériques) concernant la recherche en éducation et formation. Les méthodes de la bibliométrie statistique et multiparamétrique que décrit Dutheil (1992, p. 255-256) semblent applicables aux banques de données existantes. La question préalable concerne l'évaluation du rapport coût-efficacité. Qu'attendre de telles méthodes d'analyse si les sources ne sont pas fiables ou si leur caractère restreint rend possible une évaluation qualitative directe sans le détour par une phase de traitement quantitatif ? Cette question d'ordre pratique ne peut être traitée qu'à la lumière d'une évaluation des sources existantes à mener. Nous nous contenterons ici de les mentionner.

Si l'on passe en revue les sources françaises existantes, nous disposons de plusieurs banques de données. Sur un plan généraliste (touchant à l'ensemble de la recherche en éducation et formation), il faut mentionner tout d'abord la partie Sciences de l'éducation de la banque de données FRANCIS produite par l'Institut de l'information scientifique et technique (INIST) du CNRS. Cette banque remonte au début des années 70 et traite essentiellement des articles de périodique dont une part importante en langue étrangère. La banque de données EMILE, produite depuis 1985 par l'Institut national de recherche pédagogique (INRP), traite essentiellement les références non périodiques. Elle est compatible avec FRANCIS sur le plan de l'indexation. Ces deux banques de données forment la principale source généraliste qui peut être exploitée sous un angle bibliométrique. La banque de données produite à l'université de Paris-X Nanterre par Jacky Beillerot concernant les thèses en éducation de 1969 à 1989 avec un complément de 1990 à 1992 a déjà donné lieu à une étude sectorielle (cf. bibliographie) tout à fait intéressante. Au lieu d'une approche hyperglobale, multidisciplinaire et multidocummentaire, cette étude propose une approche monographique centrée sur un seul type de document, ce qui lève l'objection évoquée plus

haut concernant les pondérations à introduire pour comparer statistiquement les différents types de documents.

Sur un plan spécialisé, nous disposons de diverses micro-banques de données, à l'INRP en particulier, dans les domaines de la didactique du français langue maternelle (DAF), de la formation des enseignants (PROF), des manuels scolaires (EMMANUELLE), etc. Ces banques permettent d'appréhender un sous-domaine très circonscrit de la recherche en éducation et formation. Elles possèdent plusieurs atouts pour la scientométrie : elles visent à l'exhaustivité, sont parfois ouvertes à la francophonie, ont une profondeur de plusieurs décennies, possède une indexation qui ajoute une valeur d'information supplémentaire. Un recensement des différentes banques de données plus ou moins spécialisées de cette nature dans les pays francophones permettrait de dresser un inventaire et de montrer les secteurs particulièrement actifs dans l'étude rétrospective de la littérature de recherche.

À côté de ces sources bibliométriques, nous disposons d'une banque de données factuelles EMILE2, coproduite par l'INRP et l'INIST, qui recense les recherches et les unités de recherche françaises en éducation et formation depuis 1974. Cette banque constitue la base d'un *who's who* de la recherche française en éducation et formation en intégrant non seulement les composantes thématique (mots-clés, mots du titre ou des résumés) et biographique (auteurs/chercheurs), comme dans les banques de données bibliographiques, mais aussi la composante institutionnelle (affiliation des chercheurs). EMILE2 alimente pour la partie française le réseau européen EUDISED du Conseil de l'Europe (Strasbourg) dont la banque de données intègre l'information sur les recherches en éducation en provenance de 33 pays européens.

Les projets

Ce survol des sources permet de dessiner quelques axes de travail pour l'avenir.

Un premier axe concerne l'application de méthodes bibliométriques plus ou moins sophistiquées (statistiques ou multiparamétriques) sur les sources existantes lorsqu'elles sont jugées suffisamment fiables dans leur totalité ou pour une partie de leur

contenu seulement. L'objectif est de constituer un ensemble de monographies qui permettent, par touches et comparaisons successives, de dresser un tableau de la recherche en éducation et formation.

Le deuxième axe concerne la production d'une banque de données spécifique qui permettrait de développer un aspect des méthodes scientométriques particulièrement intéressant d'un point de vue scientifique : l'analyse des citations. En sélectionnant les quelques revues généralistes du domaine, il s'agirait de constituer une banque de données recensant pour chacun des articles parus les références bibliographiques citées. Cette banque pourrait remonter à une période d'une vingtaine d'années. Outre les réseaux intellectuels (les noyaux et les constellations), les influences entre disciplines et entre aires culturelles, que l'analyse des citations mettrait en valeur, il serait possible d'étudier aussi les différents types de citations en usage, les codes à l'honneur en la matière, etc.

Enfin, une exploitation systématique de la banque de données EMILE2 pourrait permettre de dresser des cartographies institutionnelle, géographique, disciplinaire et thématique de la recherche française en éducation et formation et d'apporter des indications sur la manière avec laquelle les recherches sont effectuées (nombre de chercheurs par recherche, lien entre les unités de recherche, etc.).

Ces différents axes et projets supposent une collaboration entre tous les partenaires intéressés à développer ce type d'approches en sciences de l'éducation, en même temps qu'une validation auprès d'experts reconnus des champs de recherche considérés. C'est pourquoi nous terminons par un appel à proposition et à collaboration avec la conviction qu'une meilleure visibilité de ces champs très divers est autant souhaitable que nécessaire.

Philippe CHAMPY

Service Banques de données

Institut national de recherche pédagogique

Notes

- (1) «Les méthodes bibliométriques et scientométriques sont des méthodes d'évaluation et non des méthodes de mesure. En effet, une mesure est un fait objectif quantitatif dont la précision peut être déterminée par un calcul d'erreur. En revanche, les résultats bibliométriques sont relatifs et ne valent que par leur interprétation, phénomène éminemment subjectif, conjoncturel et conjecturel.» (Dutheuil, 1992, p. 253.)

- (2) Cet article est une version remaniée du texte d'une communication présentée à la Deuxième Biennale de l'éducation et de la formation, le 12 avril 1994, à la Sorbonne.
- (3) L'Association pour le développement et l'étude de la scientométrie et de la technométrie (ADEST) est animée par Jean-Pierre Courtial du Centre de sociologie de l'innovation (Ecole des mines, Paris). L'article de Christian Dutheuil (cf. bibliographie) donne un aperçu des courants et acteurs français.
- (4) La revue internationale *Scientometrics* ne mentionne aucune étude à partir de la banque de données ERIC depuis le début de sa parution en 1978.

Bibliographie

Sur la bibliométrie et la scientométrie

- *Bibliométrie et documentation : évaluation stratégique du développement scientifique et technologique*. Paris : ADBS Editions, 1991.
- CALLON, Michel, COURTIAL, Jean-Pierre et PENAN, Hervé. *La Scientométrie*. Paris : PUF, 1993. (Que sais-je ?).
- COURTIAL, Jean-Pierre. *Introduction à la scientométrie : de la bibliométrie à la veille technologique*. Paris : Anthropos-Economica, 1990.
- DESVALS, Hélène et DOU, Henri, dir. *La Veille technologique : l'information scientifique, technique et industrielle*. Paris : Dunod, 1992. En particulier : PENAN, Hervé. Analyse des citations : principes et application à la théorie microéconomique, p. 277-330.
- DUTHEUIL, Christian. Bibliométrie et scientométrie en France : l'état de l'art. *Documentaliste - Sciences de l'information*, 1992, vol. 29, n° 6, p. 251-261.
- Revues : *Les Cahiers de l'ADEST*, *Revue française de bibliométrie*, *Scientometrics*

Sur les sciences de l'éducation

■ *Études générales*

- ASSOCIATION DES ENSEIGNANTS ET CHERCHEURS EN SCIENCES DE L'ÉDUCATION (AECSE). *Les Sciences de l'éducation : enjeux et finalités d'une discipline*. Paris : AECSE, 1993. (Distribué par l'INRP).
- CHARLOT, Bernard. *Les Sciences de l'éducation en 1993 : état des lieux et perspectives de développement*. Paris : CORESE, 1993.
- PLAISANCE, Eric et VERGNAUD, Gérard. *Les Sciences de l'éducation*. Paris : La Découverte, 1993. (Repères).

■ *Etudes bibliosciptométriques*

- BEILLEROT, Jacky. *Les Thèses en sciences de l'éducation : bilan de vingt années d'une discipline (1969-1989)*. Nanterre : Université de Paris-X Nanterre, 1993.
- HASSENFORDER, Jean. La communication internationale et la diffusion des revues en sciences de l'éducation. *Brises*, octobre 1983, n°3, p. 67-72.
- HASSENFORDER, Jean. La diffusion des revues en Sciences de l'éducation en France. *Perspectives documentaires en sciences de l'éducation*, 1983, n° 2, p. 9-18.
- *Recherches en éducation et formation : répertoire 1986-1991*. Paris : INRP ; Vandœuvre-lès-Nancy : INIST, 1993.
- SCHRIEWER, Jürgen et KEINER, Edwin. Communication Patterns and Intellectual Traditions in Educational Sciences : France and Germany. *Comparative Education Review*, 1992, vol. 36, n° 1, p. 25-51.
- TOUX, Claire-Hélène. Les thèmes traités par les articles de la *Revue française de pédagogie* et de la revue *Les Sciences de l'éducation pour l'ère nouvelle* à travers les tables de ces revues de 1967 à 1990. *Perspectives documentaires en éducation*, 1993, n° 29, p. 77-99.

Sur les banques de données issues de la recherche française en éducation et formation

- CHAMPY, Philippe. L'usage des banques de données à l'Institut national de recherche pédagogique : problématique et réalisations. *Perspectives documentaires en éducation*, 1991, n° 23, p. 107-128.
- CHAMPY, Philippe. Les banques de données de recherche : le cas de l'INRP. In *Continuités et ruptures : recherches et innovations dans l'éducation et la formation*. Paris : APRIEF, 1993, p. 180-183.

Sur la banque de données américaine ERIC

- MALRIEU, Denise. La base de données ERIC : évaluation de son utilisation et discussion des choix du système. *Brises*, octobre 1983, n°3, p.53-62.
- STONEHILLE, Robert M. et BRANDHORST, Ted. The Three Phases of ERIC. *Educational Researcher*, avril 1992, vol. 21, n°3, p. 18-22.

