

La fièvre de l'évaluation

Entretien avec Yves GINGRAS

Depuis une trentaine d'années, l'évaluation quantitative de la recherche agite le monde scientifique : établissements, laboratoires, revues, mais aussi chercheurs sont évalués, mesurés et classés par le moyen d'indicateurs bibliométriques. Yves Gingras porte un regard d'historien et de sociologue sur les controverses actuelles autour de l'évaluation de la production scientifique.

Yves Gingras est professeur au Département d'histoire de l'Université du Québec à Montréal et titulaire de la Chaire de recherche du Canada en histoire et sociologie des sciences. Il vient de publier *Les dérives de l'évaluation de la recherche, du bon usage de la bibliométrie* (Raisons d'agir, 2014), ainsi qu'un ouvrage collectif sur les controverses scientifiques : *Controverses : accords et désaccords en sciences sociales et humaines* (éditions du CNRS, 2014).

La Vie des idées : Dans « Les dérives de l'évaluation de la recherche », vous montrez comment les outils bibliométriques, apparus dans la première moitié du vingtième siècle, n'ont été mis au service de l'évaluation de la recherche qu'à partir des années 1980. Quels en étaient les usages auparavant, et comment expliquer que leur utilisation à des fins d'évaluation ne soit apparue qu'au cours des trente dernières années ?

Yves Gingras : Comme je le montre au Chapitre 1 de l'ouvrage, l'analyse bibliométrique, c'est-à-dire l'étude de l'évolution du nombre de publications et l'analyse des références (citations) qu'elles contiennent, émerge dans un contexte de gestion des collections des bibliothèques. Par exemple, l'analyse de la distribution de l'âge des références permet de savoir s'il faut rendre disponibles facilement les publications de plus de 10 ans ou si l'on peut les élaguer ou les mettre en entrepôt. Avant les années 1960, ces études se font à la main sur des échantillons limités de revues. Quant à l'index des citations, le fameux « science citation index » inventé par Eugene Garfield, spécialiste en sciences de l'information (pas en évaluation !) au milieu des années 1950, sa fonction première était de faciliter la recherche bibliographique. C'est d'ailleurs encore un outil essentiel de ce point de vue : au-delà de l'égotisme généré par la manie de « compter ses citations », il est en effet utile de voir qui cite vos papiers car cela vous renseigne sur les personnes qui travaillent sur le même sujet que vous et qui publient dans des revues que vous ne consultez pas nécessairement. Mais aujourd'hui, avec internet, cette fonction est souvent assurée par les revues elles-mêmes qui vous envoient – sans que vous le demandiez – des courriels indiquant : « On a cité un de vos papiers ! ». Il y a quelques jours, par exemple, j'ai reçu un courriel de Elsevier dont l'objet était intitulé « Dr. Y. Gingras, your work has been cited ». Cela m'a ainsi permis de voir un article que je n'aurais probablement pas consulté, car je n'aime pas faire de recherche bibliographique ! Une fois informatisées, les données bibliométriques peuvent servir à d'autres fonctions car elles sont faciles à agréger. Ainsi, en 1975 la compagnie de Garfield (*L'Institute for Scientific Information*) commence à publier le fameux

« facteur d'impact » des revues qui permet aux bibliothèques de savoir lesquelles sont les plus citées et donc probablement les plus lues et utilisées. Cet indicateur caractérise la revue et son usage et non pas les articles eux-mêmes, d'où la dérive absurde d'utiliser le facteur d'impact d'une revue pour mesurer la « qualité » d'un chercheur. À ce compte, il serait plus cohérent d'utiliser les citations réelles obtenues par ces articles. Car il faut savoir que même dans une revue comme *Nature*, 20 % des publications accumulent 80 % des citations. C'est dire que même un article dans *Nature* n'est pas nécessairement cité. Car toutes les distributions en bibliométrie sont des distributions de type Pareto dont la loi de Lotka est un cas particulier.

Il faut en fait attendre les années 1970 pour que la bibliométrie serve à construire des indicateurs. D'abord pour les politiques scientifiques. Alors que le pourcentage du PIB consacré à la R&D pouvait servir de mesure d'intrant au système scientifique, le nombre de publications, devenu accessible au début des années 1960 dans la base de données du Science Citations Index (devenue depuis le Web of Science), pouvait dorénavant servir de mesure d'extrant de la recherche, et le nombre de citations reçues par ces publications pouvait pour sa part mesurer (de façon imparfaite bien sûr, comme tout indicateur social) l'impact de la recherche, selon un modèle simple : intrant → extrant → impact. On retrouve ce modèle dans le premier volume de la série *Science and Engineering Indicators* produite aux États-Unis à compter de 1972 par le *National Science Board* et imité depuis par d'autres pays.

Jusqu'ici on est à l'échelle des pays. Il faut attendre les années 1980 pour voir se développer, lentement cependant, des méthodes d'évaluation à l'échelle des centres de recherche et les années 2000 pour les universités. Ce que j'ai appelé la « fièvre de l'évaluation » est un phénomène qui atteint son *acme* dans la décennie 2000 et qui voit se multiplier les indicateurs concoctés de façon arbitraire et sans réflexion méthodologique sérieuse. Le classement de Shanghai des universités dites de « calibre mondial » fait irruption en 2003, « l'indice h » des chercheurs est publié par un physicien américain en 2005, pour ne prendre que les exemples les plus connus. Et depuis les années 2010 on invente à qui mieux mieux des « altmetrix » (pour « métriques alternatives ») pour compenser les faiblesses de l'analyse des citations et tenir compte du passage aux publications électroniques, mais tout cela se fait encore sans réflexion méthodologique sérieuse.

Voilà pour une périodisation rapide des usages. Comment expliquer ce passage rapide à la sur-évaluation de la recherche ? Bien sûr, on ne peut répondre de manière complète à une telle question car il y a rarement des causalités simples dans les changements sociaux. Mais deux éléments me semblent importants. D'abord, le fait qu'à compter des années 1970, la science devient massifiée et qu'il est devenu difficile de se fier seulement aux évaluations informelles du « old boy network » pour justifier les choix scientifiques. Ce n'est donc peut-être pas un hasard si, comme je le décris dans le volume, c'est une femme à qui on avait refusé une promotion qui, en 1975, utilise les citations pour montrer qu'elle était pourtant plus citée que ses collègues et méritait donc davantage une promotion que ses deux collègues qui l'avaient obtenue. On voit là pointer le problème de la subjectivité des évaluations par les pairs qui ont – c'est inévitable – des biais conscients ou non. Donc la recherche d'indicateurs dits « objectifs » a aussi favorisé la montée en puissance de l'évaluation bibliométrique des chercheurs même si, bien sûr, les chiffres ne sont pas par définition « objectifs ». En plus des effets de massification, il y a depuis les années 1990 l'offensive d'un nouveau mode de gestion des institutions universitaires et de recherche qui, après le service public, se voit imposer les règles du nouveau management public

fondé sur des batteries d'indicateurs et autres « tableaux de bord » supposés faciliter une gestion « rationnelle » et « efficiente » des institutions. Rappelez-vous la phrase de la Directrice générale du FMI, Christine Lagarde « deux de mes indicateurs sont au vert et un autre ne l'est pas encore ». Cet aspect des choses a déjà été bien analysé dans l'ouvrage *Benchmarking* d'Isabelle Bruno et Emmanuel Didier. D'ailleurs, si on se fie à l'évolution de la présence du terme « indicateur » dans Google Ngram Viewer, on voit qu'il monte en usage à compter des années 1970 mais redécolle de plus belle au cours des années 1990.

La Vie des idées : Les analyses que vous rassemblez dans votre ouvrage servent d'assise à une critique des effets pervers des palmarès des universités, pour la plupart fondés sur des indicateurs simplistes. Vous dénoncez également le cynisme des dirigeants de nombre d'établissements qui reprennent, tout en les considérant invalides, les résultats de ces classements. Comment situez-vous cette prise de position dans l'espace des débats sur la bibliométrie et l'évaluation scientifique ?

Yves Gingras : Le sous-titre de mon ouvrage est en effet : du bon usage de la bibliométrie. Comme il ne faut pas jeter le bébé avec l'eau du bain, il importe de bien distinguer les problèmes. Dans le cas de l'évaluation et des classements, le problème est que ceux qui les construisent et qui s'en servent ne connaissent pas vraiment les propriétés des indicateurs qu'ils utilisent. Parmi les nombreux chercheurs des sciences de la nature, par exemple, qui sait vraiment que « l'indice h » est en fait corrélé à plus de 0.9 avec le nombre de publications ? Cela montre qu'il mesure davantage la production que la qualité, alors qu'on dénonce habituellement l'accent mis sur le nombre d'articles au détriment de leur qualité !

Il en va de même des dirigeants d'université, surtout en Europe il faut dire, qui s'excitent à chaque nouveau classement de Shanghai sans réaliser que les fluctuations sont sans signification d'une année à l'autre. Or, cela est politiquement important car imaginez une politique des universités fondée sur des indicateurs faux ! Si un thermomètre numérique sur votre bureau indique soudainement 30 degrés, alors qu'il en fait évidemment 20, allez-vous vous mettre à transpirer ou allez-vous plus sobrement conclure que le thermomètre est défectueux ?

Il faut dire que l'évaluation de la recherche est non seulement normale mais en fait très ancienne et même que la recherche est le métier le plus évalué. À ceux qui disent qu'il faut évaluer, il faut dire : mais on ne fait que cela ! On évalue nos articles depuis 1665, nos projets de recherche depuis 1920, nos programmes de formation et notre enseignement depuis les années 1970 ! On est donc sur-évalués ! L'évaluation ne s'improvise pas et la mise au point d'indicateurs est une opération délicate qui doit obéir à des principes, comme je le suggère au chapitre 4 de l'ouvrage.

La Vie des idées : En plus de prendre part dans les controverses liées aux enjeux de l'évaluation quantitative de la recherche, vous êtes un spécialiste de l'analyse des controverses en sciences et avez récemment consacré un ouvrage collectif traitant plus particulièrement des controverses en sciences sociales et humaines. Est-ce à dire que dans ces disciplines, les controverses se forment et se résolvent de manière spécifique ?

Yves Gingras : On sait que le renouvellement de la sociologie des sciences au début des années 1970 s'est fait sur la base d'études de controverses scientifiques. Cela a permis de mettre

en évidence une dynamique complexe de développement des sciences qui fait intervenir des instruments, des arguments, des positions institutionnelles plus ou moins fortes, processus moins stylisé (et moins désincarné) que celui qui intéressait jusque-là les philosophes des sciences.

Ce qui me frappe est que l'on a peu tenté de manière systématique de faire le même exercice pour les sciences humaines et sociales qui sont pourtant aussi le lieu de nombreuses controverses. L'idée de l'ouvrage que j'ai dirigé, suite à un séminaire que j'ai tenu sur le sujet, est de lancer ce chantier en proposant de partir d'une série d'études de cas, mais aussi en exposant en introduction un plan général permettant de situer les différents types de controverses, lesquelles forment un spectre allant du débat strictement limité au sein du champ scientifique jusqu'aux controverses publiques qui font intervenir une multitude d'acteurs aux formations et aux intérêts hétérogènes. Tirer profit des études sur les sciences de la nature nous permet, je crois, de réaliser que les sciences sociales sont en fait plus dangereuses socialement que les sciences de la nature et qu'elles engendrent donc des débats sociaux souvent plus virulents quand ces recherches remettent en cause des représentations sociales bien ancrées et qui confortent des idéologies et des positions sociales qui peuvent difficilement être remises en cause, même par des études scientifiques fondées sur des données aussi « objectives » que possible (documents, statistiques, etc). La résistance sociale est donc souvent très forte face aux résultats des sciences sociales alors qu'une controverse sur les quarks ou les neutrinos n'excite pas la population et ne fait intervenir que quelques dizaines de physiciens. Cette résistance est parfois telle qu'il devient impossible d'aborder de façon froide et rationnelle certains sujets. Car il est faux de penser que les sciences sociales ne sont pas cumulatives et ne fournissent que des opinions ou des idéologies. Le problème est davantage lié à la tendance à renommer les choses pour se distinguer, ce qui donne l'illusion du renouvellement alors que derrière bien des mots à la mode qui ne sont pas des concepts, on pointe en fait souvent les mêmes réalités et les mêmes explications tout en donnant l'impression de réfuter un collègue ou même tout un paradigme! Venant de la physique, je pense souvent au fait que l'on ne peut pas imaginer un physicien qui, au lieu d'utiliser la mécanique quantique, ferait mine d'inventer un nouveau vocabulaire tout en utilisant de fait l'équation de Schrödinger sous un autre nom.

La Vie des idées : Terminons par la question plus générale des liens entre les deux objets de sociologie des sciences que sont les controverses et la bibliométrie. Leur étude conjointe peut-elle contribuer à notre compréhension du monde scientifique et de ses transformations ? Y a-t-il une possibilité de dialogue fructueux entre vos deux ouvrages ?

Yves Gingras : Dans certains cas, la bibliométrie ajoute un angle de vision irremplaçable. On parle en effet beaucoup de « communauté scientifique » mais on observe qu'au delà du mot, la plupart des travaux utilisent des méthodes classiques qui abordent des individus (archives, entretiens) et qui se fixent en fait sur quelques individus ! Newton contre Leibniz, Lavoisier, Liebig, Einstein ou encore un ou deux inconnus habilement tirés de l'oubli. Or, la seule façon de construire vraiment la structure d'un champ scientifique est d'utiliser des méthodes qui permettent d'atteindre l'ensemble des chercheurs et non pas se contenter du mot magique « communauté » tout en n'étudiant que deux ou trois représentants célèbres. Comme je l'ai montré dans de nombreux textes, on peut suivre la formation de disciplines et de spécialités de façon globale¹, mais aussi la diffusion d'une innovation comme la découverte de l'ADN avec des

¹ Gingras, Yves, « [The Transformation of Physics from 1900 to 1945](#) », *Physics in Perspective*, vol. 12, no 3 (2010)

données bibliométriques². De même, le débat entre historiens sur les mérites relatifs d'Einstein et de Poincaré dans l'invention de la théorie de la relativité en 1905 peut être éclairé par des analyses bibliométriques qui permettent d'éviter les jugements rétrospectifs sur la visibilité (ou invisibilité) de certains travaux au moment de leur publication³.

En somme, comme cela devrait pourtant être évident et aller de soi, les méthodes d'analyses retenues (quantitatives, qualitatives ou mixtes) devraient être adéquates aux questions posées. Si l'analyse d'un fond d'archives peut suffire à l'étude biographique d'une trajectoire, la saisie d'une dynamique globale d'un champ scientifique requiert des méthodes permettant l'accès à la structure même du champ. Or, l'étude des co-citations de milliers d'articles scientifiques produits sur une période donnée, permet de visualiser cette structure et son évolution temporelle. Et comme je l'ai montré pour les 17^e et 19^e siècles, l'évolution de la science peut aussi être étudiée en analysant les co-citations dans les correspondances entre savants⁴. Il n'y a donc aucune raison de créer de faux débats de méthodes, les choix ne pouvant être discutés en dehors de questions précises auxquelles ces méthodes sont censées répondre à une échelle donnée. Il aurait d'ailleurs été intéressant d'aborder dans l'ouvrage la querelle stérile opposant les méthodes dites « qualitatives » aux approches « quantitatives », tant il est surprenant que de tels faux problèmes abstraits aient pu occuper tant de chercheurs qui semblent parfois préférer débattre pour débattre que de tenter de comprendre sérieusement un problème particulier qui pose question.

Publié dans *Laviedesidees.fr*, le 16 septembre 2014.

© laviedesidees.fr

pp. 248-265.

² Gingras, Yves. « [Revisiting the “Quiet Debut” of the Double Helix: A Bibliometric and Methodological note on the “Impact” of Scientific Publications](#) ». *Journal of the History of Biology*, vol. 43, no 1, 2010, pp. 159-181.

³ Gingras, Yves, « [The Collective Construction of Scientific Memory: The Einstein-Poincaré Connection and Its Discontents, 1905-2005](#) » *History of Science*, vol. 46, no 1, mars 2008, pp. 75-114.

⁴ Gingras, Yves, « [Mapping the structure of the intellectual field using citation and co-citation analysis of correspondences](#) », *History of European Ideas*, vol. 36, no 3 (2010) pp. 330-339.