

# L'analyse statistique au service du patrimoine scientifique



Au fil du temps, des désaccords se font jour au sein de la communauté scientifique ou de la société en général. Notamment sur la signification de certains mots, certains termes, qui au cours de l'histoire a évolué, changé. Pour évaluer ces désaccords concernant une part du patrimoine scientifique, la recherche philosophique s'essaye à une nouvelle discipline : l'analyse statistique.

**L**e développement des sciences s'est accompagné d'une production croissante d'articles scientifiques, particulièrement depuis le XIX<sup>e</sup> siècle. « Si on prend la seule revue *Nature* qui a commencé son activité en 1869, précise Charles Pence, Chargé de cours à l'UCLouvain et bénéficiaire d'un Mandat d'Impulsion Scientifique FNRS, il faudrait lire 10 articles par jour durant une centaine d'années pour obtenir une connaissance complète de ses archives ! Face à ce problème, nous avons imaginé, avec mon promoteur de thèse, une méthode de recherche alternative à la méthode "classique". »

## Rassembler de grandes quantités d'articles

Pour travailler à la signification du concept de « valeur sélective » au sein du débat sur l'évolution des espèces de Charles Darwin, Charles Pence s'est attelé à rassembler le plus d'articles scientifiques possibles liés aux débats scientifiques au tournant du XIX<sup>e</sup> et du XX<sup>e</sup> siècle. « Au début, les revues étaient réticentes à octroyer un accès à leur publication. Il fallait passer des contrats très pointus et le service juridique de l'université a été mis à contribution maintes fois. C'était le début de ce type de recherche, aujourd'hui les revues sont plus enclines à proposer un accès. » Les contrats signés, Charles Pence finit par disposer d'une impressionnante base de 300.000 articles, dont ceux de la revue Nature. Restait à analyser le contenu de ce patrimoine...

« J'ai d'abord tenté de faire des recherches via de grands tableaux Excel, mais cela s'est rapidement soldé par un échec, avoue le philosophe des sciences. Comme j'aime la programmation depuis l'enfance, j'ai fini par développer un outil informatique permettant d'effectuer des recherches dans le corps des articles. Les recherches simples via des dates ou des auteurs ne posent pas beaucoup de difficultés. Le challenge, par contre, résidait à trouver des informations sur la signification de termes particuliers. Afin de suivre l'évolution du terme dans le temps ou ses connexions avec un terme d'intérêt et ainsi analyser comment ce concept est utilisé. »

## Préparer les données

Évidemment, la méthode de Charles Pence n'est pas sans risque. À l'instar du data mining, « une grande partie du travail consiste à préparer les données, les nettoyer et s'assurer d'avoir un ensemble d'informations qui permet de formuler une hypothèse et de la tester, souligne-t-il. En outre, un grand nombre de données ne s'apparente pas à de l'objectivité. Nous ne disposons jamais de tous les articles mais bien d'une sélection, aussi large soit-elle, qui suppose donc un biais. Il faut prendre garde à ne pas faire un travail qui ne sert qu'à confirmer une hypothèse de départ ».

Si cette approche statistique nouvelle ne vise pas à remplacer celle de lire un plus petit ensemble de textes sur un nombre plus limité d'auteurs voire un concept spécifique, « elle permet de

formuler de nouvelles hypothèses et de les tester en parallèle d'une méthode plus conventionnelle », assure son concepteur.

Cette approche permet, par exemple, d'évaluer les désaccords qui se font jour au sein de la communauté scientifique ou de la société en général. « Le concept d'"espèce" est un excellent exemple de désaccord, note Charles Pence. Il remonte à Aristote, mais la science a évolué depuis lors et a connu de grands bouleversements, notamment avec le séquençage ADN. Comment rassembler à la fois des oiseaux et des bactéries sous ce concept d'espèce afin qu'il soit opérant ? Dans ce cas, c'est certain, ils sont trop différents pour que le concept fonctionne ! »

La classification du vivant par classes, ordres, familles, etc., et la quantification de sa diversité reposent en effet sur des termes aux acceptions anciennes. « Nous nous retrouvons souvent avec des mots qui ont une histoire très longue et dont la signification est devenue brouillée par les découvertes scientifiques. Par exemple, au sein de la communauté scientifique, certains soutiennent qu'il faut abandonner le concept de biodiversité tandis que d'autres militent pour son maintien. Il existe aujourd'hui une dizaine de définitions différentes du terme « espèce ». Notre approche statistique permet d'avoir un regard sur la distribution dans le temps de ces désaccords sur les différentes significations attachées à un concept. »  
Voire même de réaliser une cartographie d'un terme avec les associations les plus courantes ou pertinentes !

## Avoir un rôle dans le débat scientifique et sociétal

« Ces concepts sont au cœur de débats et d'enjeux sociétaux. La biodiversité en est un bel exemple. La mesure classique de la biodiversité, à lier avec les problèmes écologiques que nous connaissons, se base sur le nombre d'espèces. Mais cela ne suffit pas à définir la biodiversité ! Il faudrait



 **Charles Pence**, Bénéficiaire d'un Mandat d'Impulsion Scientifique FNRS, UCLouvain

ajouter les liens phylogénétiques entre espèces, voire les clades ou les familles. »

L'approche statistique au sein de la philosophie permettra, espère Charles Pence, d'élargir le rôle du philosophe autant au sein de la communauté scientifique que dans la société. « Aujourd'hui, la philosophie des sciences peut jouer un rôle descriptif et aussi normatif. Cette dernière approche est notre challenge pour les prochaines années : comment intervenir dans la démarche scientifique ? Le scientifique est plus absorbé par sa recherche et ses résultats que par des problèmes de définition. C'est là où se situe l'intérêt de notre démarche et le rôle du philosophe. La recherche sur le sens et l'histoire de certains termes nous permet de clarifier le sens des concepts qui sont utilisés par les chercheurs. En standardisant certains de nos résultats – c'est le travail que va commencer un postdoctorant d'ici peu –, nous pourrions jouer un rôle concret au sein de la recherche scientifique. Tout en ayant un impact dans les débats et les questions sociétales autour de la biodiversité. »

 Adrien Dewer

“  
**Notre approche statistique permettra d'élargir le rôle du philosophe au sein de la communauté des chercheurs.**