

**Supplément : Pour une recherche en appui à l'action publique :  
leçons de l'expérience du ministère de l'environnement (1990–2016)**

## Analyser la qualité de l'évaluation des risques : l'exemple des perturbateurs endocriniens★

Laura Maxim<sup>1,\*</sup>, Sandrine Blanchemanche<sup>2</sup>, Natalia Grabar<sup>3</sup> et Jeroen Van Der Sluijs<sup>4,5,6</sup>

<sup>1</sup> Études interdisciplinaires des sciences et de la technologie, Institut des Sciences de la communication, UMS 3665, CNRS, université Paris-Sorbonne, UPMC, 75013 Paris, France

<sup>2</sup> Sociologie, Économie publique, AgroParisTech, INRA, Université Paris-Saclay, 75231 Paris cedex 05, France

<sup>3</sup> Traitement automatique des langues, CNRS, Université de Lille, UMR 8163 – Savoirs Textes Langage, BP 60149, 59653 Villeneuve d'Ascq cedex, France

<sup>4</sup> Sciences de l'environnement, Centre for the Studies for the Sciences and the Humanities, University of Bergen, Bergen, Norway

<sup>5</sup> Department of Chemistry, University of Bergen, Bergen, Norway

<sup>6</sup> Copernicus Institute of Sustainable Development, Environmental Sciences, Utrecht University, Utrecht, The Netherlands

Le traitement de l'incertitude dans l'expertise<sup>1</sup> est associé à des enjeux politiques importants. De nombreux épisodes dans l'histoire des interfaces science-politique démontrent que l'absence de considération pour l'incertitude peut conduire à des controverses et à l'érosion de la confiance dans les institutions scientifiques et politiques (Van der Sluijs, 2002)<sup>2</sup>. Par ailleurs, ignorer l'incertitude peut conduire à des décisions politiques de mauvaise qualité, avec d'importantes conséquences sanitaires et économiques (European Environmental Agency, 2013). Ce type de débats autour du rôle politique de l'incertitude a conduit au développement de démarches méthodologiques, pour l'évaluer et la rendre communicable. Leurs racines conceptuelles se trouvent dans les études théoriques et empiriques sur le rôle de la science dans la gouvernance

du risque (e.g., Nowotny *et al.*, 2001 ; Wynne, 1992), réalisées entre autres par le courant de pensée dit de l'école « post-normale » (Funtowicz et Ravetz, 1990).

Cette appellation fait référence à un rapport entre enjeux de gouvernance et degré d'incertitude dans la connaissance des objets/systèmes :

- quand les enjeux et l'incertitude sont réduits, la science « appliquée » ou « normale » est considérée comme adéquate pour appréhender la réalité ;
- quand les deux ont une intensité moyenne, on parle de « consultance professionnelle » : en plus de l'application de méthodes validées, l'expérience de l'expert intervient dans le jugement (il en va ainsi des chirurgiens, des architectes...);
- quand les deux sont élevés, il est besoin de se référer à de nouvelles pratiques scientifiques ; ce sont celles-ci qui sont réunies sous le terme de science « post-normale » (Fig. 1).

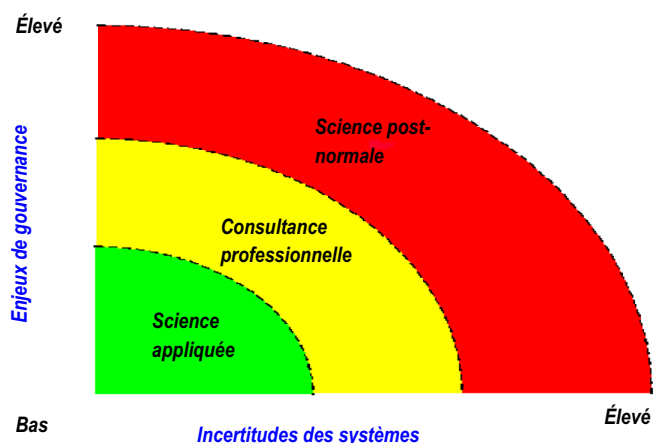
La science post-normale préconise d'accorder une attention particulière à l'incertitude, afin d'améliorer la qualité de la connaissance scientifique utilisée dans la prise de décisions et la gouvernance des risques environnementaux complexes (Funtowicz et Ravetz, 1993). Cette proposition se traduit notamment par un ample mouvement de recherche qui propose des outils d'évaluation qualitative ou quantitative de l'incertitude pouvant être utilisés par les experts pour permettre aux acteurs sociaux, aux responsables des politiques publiques et au grand public d'en prendre la mesure. Peu présent en France, ce courant fait l'objet d'une ample littérature anglo-saxonne.

★ Voir dans ce supplément le texte d'introduction de Marcel Jollivet.

\* Auteur correspondant : [laura.maxim@cnrs.fr](mailto:laura.maxim@cnrs.fr)

<sup>1</sup> On emploie ici le mot expertise pour nommer les connaissances scientifiques utilisées pour prendre des décisions politiques (*science for policy*).

<sup>2</sup> Le cas décrit par Van der Sluijs est illustratif du lien très étroit entre qualité et crédibilité, dans l'expertise : en 1999, un statisticien de l'Institut national néerlandais pour la santé publique et l'environnement (RIVM) a accusé l'institut de « mentir et tromper » dans son rapport sur l'état de l'environnement dans le pays, en raison de son élaboration à partir de modèles mathématiques peu validés et trop rarement sur la base de mesures réelles. Reprise par la presse grand public, cette accusation s'est transformée en scandale autour des chiffres fournis par le RIVM, dont la crédibilité en a sensiblement souffert.



**Fig. 1.** De la science « normale » à la science post-normale (reproduit et traduit par nos soins, d'après Funtowicz et Ravetz, 1993).

Dans le projet Dico-Risque<sup>3</sup>, nous nous sommes appuyés sur cette littérature pour l'appliquer à l'analyse de l'incertitude inhérente à l'évaluation du risque chimique. Dans le cas des perturbateurs endocriniens (PE) en général, et du Bisphénol-A (BPA) en particulier, le défi pour tout outil d'évaluation de la qualité de la connaissance est de pouvoir rendre compte des évaluations divergentes de scientifiques venant de disciplines différentes (ex. : endocrinologie, toxicologie...) et ayant des affiliations différentes (industrie, recherche académique, experts des agences sanitaires au niveau national et européen, gouvernements) tout en permettant d'avoir une vue synthétique de la qualité globale de l'étude. La littérature montre que les mêmes études et les mêmes données brutes peuvent être interprétées de manières différentes par divers experts dans des contextes variés et que cela peut conduire à des conclusions complètement opposées en termes d'évaluation du risque (Rudén, 2001). Dans ces cas, une « moyenne » de ces avis divergents n'aurait aucun sens.

L'évaluation de la qualité d'une étude (ex. : un test toxicologique) peut être très importante pour le choix des études dites « critiques » dans les agences sanitaires. En effet, ce sont ces études qui fournissent par la suite les valeurs limites des effets (ex. : NOAEL<sup>4</sup>) incluses dans

l'évaluation du risque, avec des conséquences directes sur les résultats finaux et donc sur la prise ultérieure de décisions. Les substances comme le BPA pouvant être présentes dans de nombreux produits de consommation, les conséquences d'une mauvaise évaluation de la qualité des études peuvent donc influencer la régulation des risques et finalement toucher des milliers ou de millions de personnes. Pour cette raison, l'évaluation de la qualité des études utilisées dans des contextes réglementaires doit être particulièrement robuste, ce qui n'est pas le cas aujourd'hui.

L'hypothèse centrale de notre travail est que le rôle de l'expertise est non pas de fournir « la connaissance parfaite », mais la « meilleure connaissance scientifique disponible » à un moment donné. Adopter cette hypothèse nous conduit à élargir la définition de l'incertitude, classiquement réduite au « manque de connaissance », pour y inclure :

- une dimension technique (ex. : l'utilisation d'instruments de mesure précis) ;
- une dimension méthodologique (ex. : l'utilisation de la meilleure connaissance et des meilleures pratiques scientifiques pour la construction du protocole d'expérimentation, le choix des hypothèses (lorsque les connaissances manquent) et des méthodes (lorsque plusieurs sont disponibles) ;
- une dimension normative (interprétation des données brutes, considérations portant sur le niveau de la preuve) ;
- une dimension communicationnelle (liée à l'exhaustivité et la compréhensibilité de la communication des résultats).

L'objectif principal du projet Dico-Risque a été de développer une « boîte à outils » d'analyse de l'incertitude et de la qualité de la connaissance, pour explorer et comprendre les interdépendances entre les dimensions scientifiques et les dimensions politiques des controverses autour des risques du BPA et des risques chimiques en général. Après une description des méthodes et les outils développés, nous en tirerons quelques leçons sur les formes de relations entre recherche et politiques publiques que ce travail a induites...

## Méthode

Le corpus de travail réunit les documents d'expertise qui servent à l'évaluation des risques sanitaires du BPA dans les agences sanitaires (EFSA<sup>5</sup>, ANSES<sup>6</sup>), notamment des études toxicologiques et des rapports d'évaluation

<sup>3</sup> Le projet Dico-Risque (développement d'une boîte à outils pour l'analyse de l'incertitude et de la qualité de la connaissance, dans les évaluations des risques des perturbateurs endocriniens : application à l'étude de cas du Bisphénol-A) a été financé par le programme national de recherche sur les perturbateurs endocriniens, dans le cadre de son appel à projets 2010 (Projet PNRPE 11-MRES-PNRPE-4-CVS-30).

<sup>4</sup> NOAEL : no-observed-adverse-effect-level. Il s'agit de la plus haute dose testée qui ne présente pas un effet adverse, selon Dorato et Engelhardt, 2015.

<sup>5</sup> European Food Safety Authority.

<sup>6</sup> Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.

des risques. Dans ce corpus, l'incertitude est présente de deux manières : y faire référence faisant partie des règles de présentation des études et des rapports, elle y est explicitement exprimée en termes linguistiques spécifiques ; elle est par ailleurs implicitement contenue dans les hypothèses sur lesquelles reposent ces mêmes documents, du fait de leurs limites intrinsèques par rapport à la complexité de l'objet de la recherche.

Pour la construction des outils d'analyse de l'incertitude propres à ce projet, trois méthodes ont été associées. Deux portent sur l'incertitude explicite :

- une analyse du contenu textuel et sémantique du corpus à l'aide du traitement automatique des langues (TAL). Les travaux des chercheurs en TAL sont motivés par le fait que le traitement manuel des données langagières devient de plus en plus difficile étant donné leur volume toujours grandissant. Ces travaux s'appuient sur des théories et méthodes issues de la linguistique et de l'informatique. La méthode a été utilisée pour construire une série de classes d'incertitude définies par l'équipe du projet (exemple : incertitude associée au choix inadéquat de la taille d'un échantillon, dans une étude toxicologique) et y associer les phrases des textes du corpus ;
- la méthode Holyrisk développée à l'Inra qui est basée sur le codage à la main des rapports d'évaluation du risque (produits par des agences de régulation du risque en Europe et aux États-Unis), pour y identifier l'expression de l'incertitude. Par différence de la méthode précédente (TAL), Holyrisk mobilise une typologie d'incertitudes agrégée, ayant un nombre limité de classes, dont l'ambition est d'être applicable à plusieurs domaines d'évaluation du risque (chimique, alimentaire...).

La troisième méthode, nommée Qualichem, porte sur l'incertitude implicite. Elle consiste en une enquête auprès d'experts<sup>7</sup>, associée au développement de typologies originales de l'incertitude. Ces typologies, validées avec les experts interrogés, permettent de repérer les aspects qui peuvent influencer la qualité finale de l'évaluation du risque (par exemple, le choix d'une méthode d'analyse statistique pour interpréter des résultats d'une expérimentation).

## Les résultats et leur utilité pour les responsables des politiques publiques

De manière générale, les résultats du projet confirment qu'il existe une grande hétérogénéité interindividuelle entre les scientifiques en ce qui concerne l'appréciation de la qualité des mêmes études, à l'aide des mêmes critères.

<sup>7</sup> En anglais, *expert elicitation*

Le projet a permis la production de trois outils de caractérisation de l'incertitude, chacun issu de l'application d'une des trois méthodes présentées (enquête auprès d'experts, TAL et Holyrisk) :

- pour l'évaluation de l'incertitude implicite et de la qualité de la connaissance, l'outil intitulé Qualichem<sup>8</sup>, a été développé. Cet outil a été validé avec vingt-huit scientifiques spécialistes du BPA, sur cinq études de cas (deux études *in vivo*, deux études épidémiologiques et une revue de littérature) (Maxim et Van der Sluijs, 2014). Dans un groupe d'experts, l'outil permet d'inclure tous les points de vue des participants, y compris les avis minoritaires, sur une base comparable et reproductible d'une étude à l'autre. Notre démarche est de faire ressortir à la fois les points de consensus et les divergences afin de faciliter et structurer les discussions dans les groupes d'experts ;
- pour la caractérisation de l'incertitude explicite – autrement dit, exprimée dans les documents d'expertise – trois approches en TAL ont été testées (Encadré). Elles l'ont déjà été avec succès par le passé, dans le domaine biomédical (Grabar et Hamon, 2010).

Notre recherche montre que l'approche par l'apprentissage automatique peut très bien identifier des significations complexes de l'incertitude et peut donc être utile pour faire des suggestions à des analystes humains.

Un résultat de l'expérience que nous avons acquise mérite tout particulièrement de retenir l'attention. Il est tiré de notre participation aux groupes de travail d'agences sanitaires : contrairement à une hypothèse faite au départ du projet, les incertitudes exprimées semblent conduire à renforcer les mesures de gestion des risques, au lieu de bloquer l'action. Plus encore, les décisions les plus fortes sont prises sur la base de documents citant le plus d'incertitudes. Cependant, le petit nombre de documents de gestion des risques du BPA ne permet pas une analyse statistique sur ce point.

Enfin, cette recherche nous a permis de réaliser des développements méthodologiques qui se sont avérés utiles pour le groupe de travail « perturbateurs endocriniens » de l'Anses, auquel la coordinatrice de ce projet a participé en qualité d'experte.

Au-delà de la mise au point de l'outil Qualichem et ces différentes avancées tant sur le fond que méthodologiques, un autre résultat intéressant est à mettre au crédit du projet Dico-Risque. En effet, l'analyse des rapports entre l'expression linguistique de l'incertitude

<sup>8</sup> Qualichem est un outil qui permet d'évaluer le niveau global d'incertitude dans une étude ou un rapport. Il fournit un indicateur de confiance agrégé des experts dans la qualité globale d'une étude et en propose une représentation graphique facile à interpréter par les non-spécialistes.

### Encadré. Les trois approches en traitement automatique des langues (TAL)

Les approches TAL ont pour objectif d'assigner les phrases des textes aux classes d'incertitude définies par l'équipe du projet (exemple : incertitude associée au choix inadapté de la taille d'un l'échantillon, dans une étude toxicologique). À partir d'exemples annotés à la main, des algorithmes automatiques généralisent ces exemples et proposent des modèles qui permettent d'effectuer la catégorisation automatique des phrases :

1- L'approche à base de règles repose sur les annotations et un ensemble de règles, comme par exemple la co-occurrence des mots-clés et des marqueurs de doute ou d'approximation (Grabar *et al.*, 2012).

Exemples d'informations linguistiques et sémantiques contenues dans les textes :

- le doute au sujet des résultats obtenus, leur interprétation... (*e.g., possible, hypothetical, should, can, may, usually*);
- la négation, relative par exemple au fait que l'étude est incomplète dans la description de la méthode ou des résultats (*e.g., no, not, neither, lack, absent, missing*);
- les limitations, relatifs par exemple à la taille insuffisante de l'échantillon traité, au faible nombre de doses testées... (*e.g., only, shortcoming, small, insufficient*);
- l'approximation, associée par exemple à l'imprécision dans la description du protocole de recherche d'une étude toxicologique (*e.g., approximately, commonly, considerably, estimated*).

2- L'approche par apprentissage supervisé exploite les données annotées manuellement et permet de les généraliser afin de créer un modèle de langue.

3- L'approche par recherche d'information est basée sur la détection de mots-clés, leur voisinage et leur significativité pour une phrase donnée.

exprimée dans les rapports d'expertise et la décision politique montre qu'il existe des différences nationales entre les évaluations effectuées aux États-Unis et en Europe : on trouve plus de « précautions » exprimées dans les évaluations américaines que dans celles de l'UE. En outre, les analyses montrent que, contrairement à une vision réaliste de la connaissance qui postulerait que l'incertitude diminuerait avec l'accumulation de connaissances, on observe l'inverse : la quantité d'incertitude exprimée par les experts augmente avec le nombre de références citées, rejoignant ici une

approche cognitiviste. La comparaison entre l'expression linguistique de l'incertitude pour les perturbateurs endocriniens et pour les contaminants montre que pour les premiers, l'expertise repose plus sur des variables de jugement que celles concernant les contaminants. Cela souligne l'ambiguïté restante dans le domaine des perturbateurs endocriniens et le poids important du jugement d'expert.

### Références

- Dorato M.A, Engelhardt J.A., 2015. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s), *Regulatory toxicology and pharmacology*, 42, 3, 265-274, <https://doi.org/10.1016/j.yrtph.2005.05.004>.
- European Environmental Agency (EEA), 2013. Late lessons from early warnings: science, precaution, innovation, Report, Luxembourg, Publications Office of the European Union, <http://www.eea.europa.eu/publications/late-lessons-2>.
- Funtowicz S., Ravetz J. 1990. *Uncertainty and quality in science for policy*, Dordrecht, The Netherlands, Kluwer Academic Publishers, 231 p.
- Funtowicz S., Ravetz J.R., 1993. Science for the post-normal age, *Futures*, 25, 739-755.
- Grabar N., Hamon T., 2010. Exploitation of linguistic indicators for automatic weighting of synonyms induced within three biomedical terminologies, in Safran C., Reti S., Marin H.F. (Eds.), *Studies in Health Technology and Informatics*, MEDINFO 2010, 160, Netherlands, IOS Press BV, 1015-1019, <http://ebooks.iospress.nl/publication/13595>.
- Grabar N., Maxim L., Hamon T., 2012. Towards an automatic detection of the chemical risk statements, in *Proceedings of The Fourth Swedish Language Technology Conference (SLTC 2012)*, Lund, Sweden, 33-34.
- Maxim L., Van der Sluijs J., 2014. Qualichem in vivo: A tool for assessing the quality of in vivo studies and its application for Bisphenol A. *PLOS One*, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087738>.
- Nowotny H., Scott P., Gibbons M., 2001. *Re-thinking science, knowledge and the public in an age of uncertainty*, Cambridge, Polity, 288 p.
- Rudén C., 2001. Interpretations of primary carcinogenicity data in 29 trichloroethylene risk assessments, *Toxicology*, 169, 3, 209-225.
- Van der Sluijs J.P., 2002. A way out of the credibility crisis around model-use in Integrated Environmental Assessment, *Futures*, 34, 133-146.
- Wynne B. 1992. Uncertainty and environmental learning, *Global Environmental Change*, 2, 111-27

**Citation de l'article** : Maxim L., Blanchemache S., Grabar N., Van Der Sluijs J., 2017. Analyser la qualité de l'évaluation des risques : l'exemple des perturbateurs endocriniens. *Nat. Sci. Soc.* 25, S72-S75.