



HAL
open science

Une étude de l'incertitude dans les textes scientifiques : vers la construction d'une ontologie

François-C. Rey, Marc Bertin, Iana Atanassova

► To cite this version:

François-C. Rey, Marc Bertin, Iana Atanassova. Une étude de l'incertitude dans les textes scientifiques : vers la construction d'une ontologie. TOTh 2018 Terminology & Ontology: Theories and applications, Jun 2018, Chambéry, France. pp.229-242. hal-01755245

HAL Id: hal-01755245

<https://hal.science/hal-01755245v1>

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une étude de l'incertitude dans les textes scientifiques : vers la construction d'une ontologie

François-C. Rey*, Marc Bertin**, Iana Atanassova***

* CRIT, Université de Bourgogne Franche-Comté
Francois-c.rey@edu.univ-fcomte.fr

** ELICO, Université Claude Bernard Lyon 1
Marc.bertin@univ-lyon1.fr

*** CRIT, Université de Bourgogne Franche-Comté
Iana.atanassova@univ-fcomte.fr

Résumé. L'incertitude en sciences est une problématique de nature disciplinaire, liée à l'objet de l'étude et aux méthodologies mises en œuvre. Cependant, l'enjeu de traiter l'incertitude est essentiel sur le plan sociétal, rappelant la nécessité de disposer d'outils permettant l'analyse du discours scientifique, afin de répondre aux problématiques actuelles. La problématique que nous abordons ici est celle de la compréhension interdisciplinaire et conceptuelle de la notion d'incertitude à travers la constitution et l'étude d'un corpus d'articles scientifiques traitant du changement climatique.

1. Introduction

Dans quelle mesure l'incertitude est-elle détectable dans les publications scientifiques ? À cette problématique épistémologique, nous souhaitons apporter une réponse linguistique et informatique afin de répondre au fait que les textes scientifiques peuvent relayer une forme de subjectivité, d'incertitude, et peuvent être porteurs d'ambiguïtés, que ce soit à cause de limites théoriques ou de protocoles expérimentaux. Ce peut être particulièrement le cas dans la conclusion ou la partie de discussion d'un article, puisqu'elles servent à évoquer des perspectives, des hypothèses qui restent à examiner dans des futures recherches, des voies à explorer, des intuitions et opinions de l'auteur. Dans ces cas, l'expression de l'incertitude devient l'outil pour diriger l'attention du lecteur vers les limites de la connaissance actuelle et les domaines que la science doit encore explorer.

Cette problématique qu'est l'analyse de l'incertitude scientifique est une problématique sociétale. La physique en apporte un exemple à travers les réajustements voulus dans la définition de ses unités de mesure fondamentales que sont les *constantes physiques* : selon le [BIPM, 2006, p. 11, 14, 19 et 34], le Système international d'unités (SI), qui fournit les unités de référence (mètre, kilogramme, seconde, ampère, kelvin, mole et candela) et dont l'usage «est recommandé dans les sciences, la technologie, les sciences de l'ingénieur et le commerce», «est, bien sûr, un système évolutif, qui reflète les meilleures pratiques en matière de mesure du moment». [Newell *et al.*, 2018] illustrent les changements actuels dans les définitions des constantes physiques du SI, en montrant comment ces définitions dépendent de la connaissance des intervalles d'incertitude sur les valeurs récemment mesurées des constantes (ces valeurs mesurées et leurs intervalles d'incertitude proviennent de l'information fournie par différentes équipes de recherche). Des ajustements fondamentaux peuvent donc se produire grâce à la correcte transmission et intégration de l'information sur l'incertitude scientifique.

À l'opposé, il se trouve dans l'actualité des exemples du fait qu'une insuffisante communication, ou une incompréhension, ou une mauvaise gestion de l'incertitude scientifique peut engendrer des conséquences importantes, telles que des catastrophes industrielles ou sanitaires, ou, à plus grande échelle, le réchauffement climatique ; d'autres types de conséquences peuvent apparaître, par exemple à l'encontre des scientifiques dans leur pratique¹.

En Sciences Humaines et Sociales (SHS) les déclarations reposant sur l'objectivité ou la subjectivité sont souvent présentes dans les textes et le concept d'incertitude est étendu : imprécision, indétermination, incomplétude, ambiguïté et imprévisibilité. L'incertitude en SHS est liée à la complexité des objets d'étude sociaux et humains, à l'influence du contexte et méthodes, aux perspectives et paradigmes divers, aux controverses et à la variété de points de vue et d'interprétations, et au mode de communication des résultats. Il s'agit alors de «faire science avec l'incertitude» [Fusco *et al.*, 2015] : l'évaluer,

1 Par exemple l'infirmière SARAGA Daniel en 2003 aux Pays-Bas, («Condamnée par les maths - à tort», p 14 in *Horizons - Le magazine suisse de la recherche scientifique*, no 115, décembre 2017), qui était seulement présente lors de neuf accidents et a été emprisonnée puis réhabilitée ; et c'est le cas de volcanologues en 2012 (*Journal Le Monde*, version en ligne du 22/10/2012 : «Séisme de L'Aquila : les scientifiques condamnés à six ans de prison»), dont le rapport avait été interprété pour rassurer catégoriquement la population avant que la terre ne tremble à nouveau et qu'ils soient emprisonnés puis réhabilités.

la mesurer, la réduire, et l'intégrer dans les processus, les résultats et la communication scientifique.

Au-delà d'une diversité thématique, l'incertitude scientifique s'éloigne significativement de l'incertitude du «sens commun»: cette notion doit être définie dans sa diversité et nature, en fixant ses frontières et origines. Notre objectif est de traiter linguistiquement les phénomènes de l'incertitude dans les articles scientifiques afin de proposer un outil permettant d'identifier et classer ces phénomènes présents dans les textes. L'identification des segments textuels exprimant une incertitude a été l'objectif de plusieurs travaux. Par exemple, l'identification de phrases spéculatives dans des textes par des approches en apprentissage automatique a été abordée par [Moncecchi *et al.*, 2012] qui souligne la spécificité de la problématique de la subjectivité. Une étude récente par [Chen *et al.*, 2018] propose d'identifier des expressions introductives d'incertitude par une expansion d'un ensemble d'expressions restreint. Cependant, ces derniers travaux expriment une vision binaire de l'incertitude et ne s'intéressent pas aux différents niveaux et dimensions de l'incertitude afin de rendre compte de la complexité de cette notion. Pour exemple, [Bernhard and Ligozat, 2011] utilisent des graduations de la certitude en s'appuyant sur des catégories pour appréhender les assertions associées à une information sur des problèmes médicaux dans des comptes rendus cliniques.

2. Problématique

La production de nouvelles connaissances s'appuie sur une démarche méthodologique rigoureuse en fonction de l'objet d'étude et de son champ disciplinaire.

L'incertitude, par sa nature et définition, est dépendante du champ scientifique qui l'utilise. Par exemple en économie, [Knight, 1921] restreint l'usage du terme «incertitude» à ce qui est non quantifiable, ou non mesurable. Dans le domaine biomédical, la notion de niveau de preuve scientifique apporté par une étude selon la force du protocole mis en œuvre [Haute Autorité de Santé, 2013] est une évaluation des études selon des critères tels que : l'objectif de l'étude, le type du protocole, le facteur étudié, les critères de jugement, les facteurs de confusion et les biais, la pertinence des analyses statistiques...

En physique, science où la mesure de l'incertitude dépend des imperfections des instruments (erreurs systématiques dépendantes du temps de réponse, de la sensibilité et de l'exactitude) et des imperfections de l'opéra-

teur (erreurs accidentelles), « Le résultat de mesure est généralement exprimé par une valeur mesurée unique et une incertitude de mesure »² et la valeur numérique comporte certaines particularités : la marge d'erreur des instruments de mesure est elle-même mesurée ; l'incertitude connue sert à qualifier la qualité de la mesure ; et il faut « combiner les composantes individuelles de l'incertitude en une seule incertitude globale » pour « donner une indication quantitative sur la qualité du résultat pour que ceux qui l'utiliseront puissent estimer sa fiabilité ».

En géographie [Fusco, 2013], elle est liée par exemple au traitement des données (échelle, pluralité des sources, ...), à la définition des objets (concepts et limites d'un paysage...), à la complexité des systèmes spatiaux (interactions, passé, prospection...), à la modélisation spatiale, à la représentation de la connaissance (limites des zones inondables...), aux phénomènes spatiaux (perception, culture, intentionnalité...) et aux processus décisionnels (scénarios urbains, tel qu'avec l'impact d'une carte scolaire...); et la recherche du degré d'incertitude dépend de la question, de l'actualisation des données et du mode de représentation des résultats intermédiaires et finaux.

Sur le plan conceptuel, il faut différencier l'incertitude de la subjectivité tout en prenant en considération la modalité épistémique. Il apparaît que l'incertitude est l'un des composants normaux des résultats de la recherche scientifique, dont les dimensions peuvent être multiples dans les différents champs disciplinaires qui se l'approprient. Ce composant doit être réutilisable avec les résultats scientifiques qu'il accompagne et dont il constitue une partie intégrante essentielle. La subjectivité est aussi partie intégrante de la science. D'un point de vue définitoire, la subjectivité est considérée en philosophie par opposition à l'objectivité comme qualité (inconsciente ou intérieure) de ce qui appartient seulement au sujet pensant, alors qu'en linguistique elle est liée à la présence du sujet parlant dans son discours³. Étudiés respectivement par la logique avec le carré des logiques modales d'Aristote, par la philosophie avec Kant, et enfin par la linguistique, les concepts liés à la subjectivité sont complémentaires et parfois non-accordables, même à l'intérieur d'un même champ disciplinaire.

2 Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés (VIM, 3^e édition), JCGM 200:2012, <https://www.bipm.org/fr/publications/guides/vim.html> et Évaluation des données de mesure – Guide pour l'expression de l'incertitude de mesure, JCGM 100:2008, <https://www.bipm.org/fr/publications/guides/gum.html>

3 Source : Centre National de Ressources Textuelles et Lexicales – CNRT.

En linguistique, [Bally, 1965] en faisant «l'analyse logique des formes de l'énonciation», considère que la phrase explicite comprend deux parties complémentaires : le «dictum» qui est «le corrélatif du procès qui constitue la représentation», et le «modus» qui est «l'expression de la modalité, corrélatif à l'expression du sujet pensant» et se compose d'un verbe modal et d'un sujet modal. Les concepts et relations autour de la notion de modalité en linguistique ne sont pas consensuels comme en témoignent d'autres travaux souvent pris comme références, par ex. [Le Querler, 1966] pour le français et [Halliday and Matthiessen, 2014] pour l'anglais. Ce problème est souligné par d'autres auteurs, par ex. [Meunier, 1981] qui pour ce problème en français emploie l'expression de «nébuleuse», et [Perkins, 1983] qui pour l'anglais déclare que «paradoxically, in spite of the vastness of the available literature, it is by no means easy to find out what modality actually is». Enfin, le concept de modalité épistémique (du grec «*épistèmè*»: *science, application de l'esprit, étude*) accompagne le savoir scientifique. Il se situe à la croisée de la linguistique et de la logique et permet de différencier les expressions du certain, du plausible, du contestable et de l'exclu. Concernant l'utilisation des expressions textuelles de cette modalité, [Vold, 2008] précise que «Le recours à la modalité épistémique [...] constitue une stratégie [...] fréquemment employée dans le discours scientifique et jugée essentielle [...]». Bien que les concepts qui permettent d'aborder l'incertitude via la linguistique soient globalement instables, la voie des modalités épistémiques est à explorer pour prendre en compte la nécessité de mieux saisir l'incertitude comme composant en science.

3. Objectifs de recherche

Notre objectif (voir la figure 1) est de proposer une modélisation linguistique de l'expression de l'incertitude dans les articles scientifiques. Nous étudions les définitions des notions de subjectivité et d'incertitude ainsi que leur expression dans les différentes disciplines scientifiques, afin d'aboutir à la création d'une ontologie de l'incertitude en sciences.

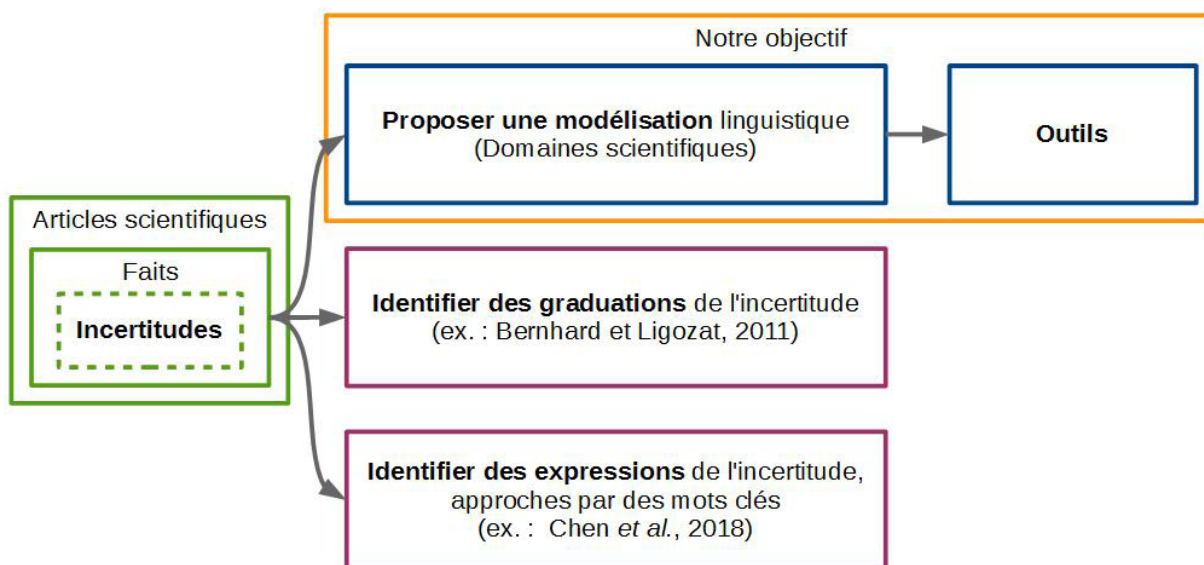


FIG. 1 – *Domaines d'analyse de l'incertitude en sciences.*

Cette recherche fait partie d'un ensemble de travaux autour de l'analyse du contenu d'articles scientifiques dans le but d'automatiser l'état de l'art de domaines scientifiques [Atanassova and Bertin, 2014, Atanassova *et al.*, 2016]. D'autres domaines d'application que les états de l'art sont envisageables, tels que : l'analyse des champs disciplinaires, du degré de validité d'affirmations scientifiques, et de la décidabilité de la prospective technique et économique ; la reconstitution de la genèse de concepts scientifiques ; la proposition d'hypothèses sur les limites des recherches ; le filtrage du factuel et du non factuel ; l'aide à la rédaction de documents scientifiques ; la communication sur les faits, les théories et les méthodes scientifiques ; la garantie de l'éthique et du principe de précaution, et des réglementations et législations. La finalité de ce travail est la création d'un outil pour l'identification et l'extraction automatique de segments exprimant de l'incertitude dans des articles scientifiques. En effet, des enjeux cruciaux dépendent des apports de la connaissance scientifique (environnement, paix, énergie...) et ces derniers doivent être correctement compris pour être pris en compte par des non spécialistes ou des non-scientifiques (climatologie, toxicologie, épidémiologie...): de nouvelles interfaces d'accès à la production scientifique, traitant l'incertitude dans les textes, devraient aussi permettre de mieux communiquer l'incertitude scientifique vers des responsables directs de chaînes de décisions qui consultent la science (politiques, ingénieurs, entrepreneurs...) et vers des professionnels

impliqués dans l'interprétation et la diffusion de la science (juristes, journalistes, enseignants...).

Term	Likelihood of the Outcome
<i>Virtually certain</i>	99-100 % probability
<i>*Extremely likely</i>	95-100 % probability
<i>Very likely</i>	90-100 % probability
<i>Likely</i>	66-100 % probability
<i>*More likely than not</i>	50-100 % probability
<i>About as likely as not</i>	33-66 % probability
<i>Unlikely</i>	0-33 % probability
<i>Very unlikely</i>	0-10 % probability
<i>*Extremely unlikely</i>	0-5 % probability
<i>Exceptionally unlikely</i>	0-1 % probability

TAB. 1 – *Echelle de probabilité de l'IPCC*

Parmi les organisations qui ont pour tâche de transmettre du savoir scientifique, certaines communiquent des informations sur des solutions qu'elles ont adoptées «sur le terrain». Par exemple, l'*Intergovernmental Panel on Climate Change (IPCC)*, organisation internationale sous l'égide de l'OMM⁴ et du PNUE⁵, communique aux auteurs de ses rapports des recommandations pratiques pour transmettre l'incertitude de manière non équivoque et consistante. À cette fin, [IPCC, 2005] et [Mastrandrea *et al.*, 2010, pour l'IPCC] recommandent aux auteurs principaux des parties du cinquième rapport de l'IPCC (*Fifth Assessment Report, AR5*) de rédiger l'information textuelle en adoptant une terminologie commune et consensuelle, prédéfinie, pour exprimer et transmettre d'une part la probabilité et d'autre part la qualité de l'incertitude. Ce choix terminologique consiste en un langage calibré (*calibrated language for describing quantified uncertainty*). En utilisant le concept de limites «floues» («*fuzzy*» boundaries), des intervalles de probabilité ont été mis en correspondance avec des expressions : l'échelle terminologique qui en résulte (voir tableau 1) indique quelle expression exacte employer⁶ pour

4 Organisation météorologique mondiale (*World Meteorological Organization, WMO*).

5 Programme des Nations Unies pour l'environnement (*United Nations Environment Programme, UNEP*).

6 [Mastrandrea *et al.*, 2010] précisent : «Quand l'information présente est suffisante, il est préférable de préciser entièrement la distribution de probabilité ou un intervalle de probabilité (par ex. : 90-95 %) sans utiliser les termes» (traduction de l'auteur).

exprimer la quantité de la mesure d'incertitude d'un résultat scientifique. Les niveaux dont l'expression est précédée par un '*' sont signalés comme moins utilisés lors de la rédaction du rapport précédent (AR4). Ce type d'échelle, qui est aussi employé par d'autres d'organisations, ne permet pas à lui seul de rendre compte de certaines dimensions de l'incertitude qu'il pourrait pourtant être nécessaire de transmettre. Il est donc accompagné ou remplacé par d'autres vecteurs de présentation de l'incertitude : par exemple, une échelle de confiance, un commentaire (séparé de l'information elle-même, pour mieux la comprendre où l'exploiter), une interface (à l'état expérimental dans le cas de la fusion d'information dans le domaine de la défense, [Davis *et al.*, 2016]).

4. Expérimentation

La prise en compte de l'incertitude diffère en concepts, en granularité et en degrés selon les domaines scientifiques, mais l'incertitude peut également être étudiée de façon transversale dans l'ensemble des champs scientifiques. Par exemple, dans le cadre d'une catégorisation manuelle de la certitude explicite dans un corpus d'articles de presse, [Rubin *et al.*, 2006] proposent des catégories de marqueurs textuels répartis dans quatre dimensions : «niveau de certitude» (absolu, haut, modéré et bas), «perspective de certitude» (point de vue de celui qui écrit et point de vue rapporté), «focus de certitude» (information abstraite et information factuelle) et «temps de référence de la certitude» (passé, présent et futur). Le tableau 2 nous offre une vue synthétique des notions et classes de l'incertitude chez différents auteurs.

Références	Notions	Classes
[Rubin <i>et al.</i> , 2006]	Niveau de certitude	absolu, haut, modéré, bas
[Rubin <i>et al.</i> , 2006]	Perspective de certitude	point de vue de celui qui écrit, point de vue rapporté
[Rubin <i>et al.</i> , 2006]	Focus de certitude	information abstraite, information factuelle
[Rubin <i>et al.</i> , 2006]	Temps de référence de la certitude	passé, présent, futur
[Dragos, 2013]	Vagueness VS precision	incomplete knowledge of the facts VS appropriate words or phrases to convey information
[NATO, 2003]	Reliability of the source & credibility of the information	from unknown to confirmed
[Goujon, 2009]	Reality	assertion, negative

[Chauveau- Thoumelin & Grabar, 2014]	Incertitude forte VS incertitude faible	influence fortement ou faiblement la fiabilité
--	--	---

TAB. 2 – *Classes de l’incertitude et notions associées*

Voici des exemples d’expressions d’incertitude identifiés manuellement dans les segments de texte de la partie discussion d’un article scientifique sur le changement climatique⁷ (le ‘*’ signale les expressions de certitude) :

- *‘Previous research has suggested that...’*
- **‘we found that all stated hypotheses were confirmed’*
- *‘The present study finds no support for these claims’*
- *‘these findings provide the strongest evidence to date that...’*
- *‘this interaction might, to some extent, be attributable to...’*
- **‘We do not dispute, however, that...’*
- *‘other recent research has suggested that...’*
- *‘the criticism might be raised that...’*
- *‘While we agree with this view and..., this shortcoming does not, however, negate the...’*

Une liste de termes-clés sur le changement climatique, a été constituée pour pouvoir reconnaître les articles traitant du changement climatique et servir à la construction d’un corpus spécifique au changement climatique. Pour cela, nous avons suivi la procédure suivante :

1. Recherche de listes de termes, expressions et acronymes du domaine (vocabulaires, glossaires) sur des sites institutionnels, encyclopédiques ou d’information ;
2. Compilation des termes de chaque source ;
3. Recomposition manuelle de la liste finale en éliminant les doublons.

Nous avons étudié les différentes relations entre l’incertitude et les objets textuels et informationnels (voir figure 2) et construit une ontologie en prenant en compte les types de raisonnement formalisés pouvant induire de l’incertitude (voir figure 3).

7 Source textuelle : « Discussion » in : van der Linden (2015) « The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence ».

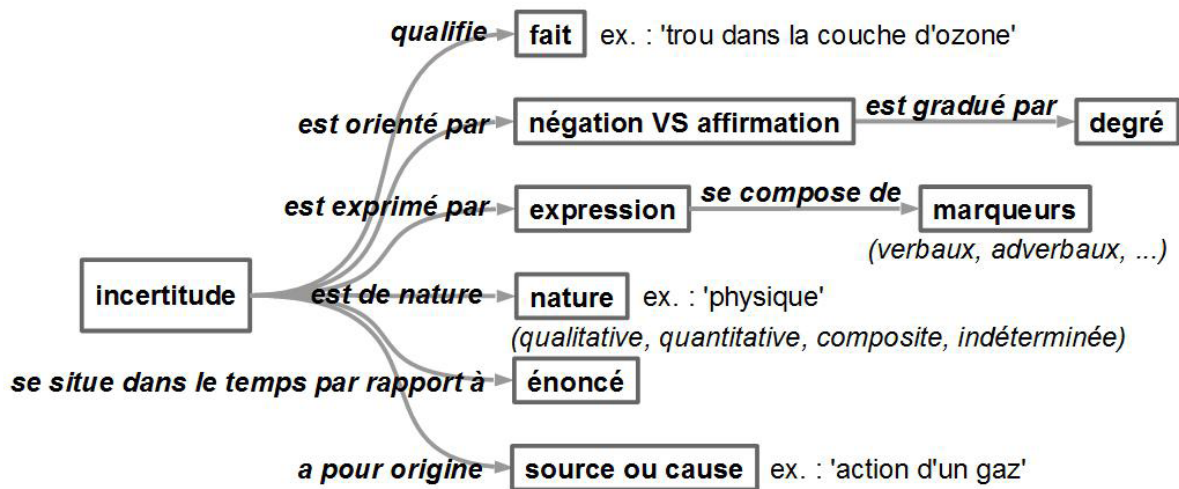


FIG. 2 – Schéma relationnel de l'incertitude scientifique.

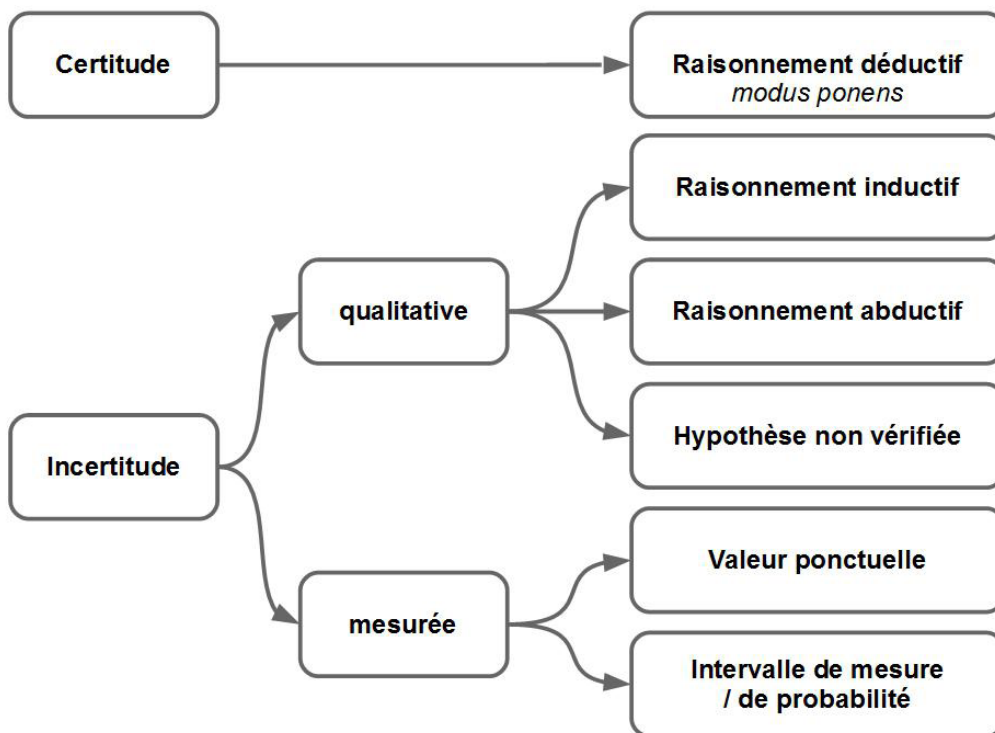


FIG. 3 – Ontologie de l'incertitude scientifique : types d'incertitude exprimée.

Une première expérimentation [Atanassova *et al.*, 2018] a été menée sur un corpus d'autour de 10 000 articles scientifiques constitué à partir de deux sous-corpus issus de *PubMed Open Access Subset*⁸: l'un des sous-corpus avec 9 463 articles en biomédecine, et l'autre avec 488 articles en physique. Dans cette expérimentation, le corpus a servi à évaluer la distribution d'expressions de l'incertitude dans les sections principales de la structure IMRaD (Introduction, Méthodes, Résultats, Discussion) des articles. Cette expérimentation et sa préparation ont aussi été l'occasion de valider à nouveau la pertinence de la problématique de l'incertitude dans le domaine scientifique du changement de climat.

5. Conclusion et discussion

Dans cet article, nous posons la problématique de l'analyse de l'incertitude en sciences, telle qu'exprimée dans les textes, et la construction de ressources linguistiques pour son repérage et classification. Nous nous sommes interrogés en premier lieu sur les origines et les causes d'incertitude en sciences. Nous avons ensuite construit une ontologie de l'incertitude. Ces résultats ont été obtenus par des observations à partir d'un corpus d'articles scientifiques traitant du changement climatique, domaine dont l'objet d'étude est d'une grande complexité et où la gestion de l'incertitude associée aux résultats scientifiques est nécessaire. Nous avons étudié les textes afin de produire une typologie de l'incertitude dans ce domaine avec une visée de généralisation. L'ontologie produite reste à être validée par l'expérimentation à grande échelle, notamment à travers la construction de ressources linguistiques pour l'annotation de corpus. Cette méthodologie nous permettra de proposer un outil pour l'analyse et l'extraction de segments textuels porteurs d'incertitude.

References

Atanassova, I. and Bertin, M. (2014). Semantic Facets for Scientific Information Retrieval. In Presutti, V., Stankovic, M., Cambria, E., Cantador, I., Iorio, A. D., Noia, T. D., Lange, C., Recupero, D. R., and Tordai, A., editors, *Semantic Web Evaluation Challenge: SemWebEval 2014 at ESWC 2014, Communications in Computer and Information Science, Communications in Computer and Information Science (Book 475)*, p. 108-113. Springer.

⁸ *Pubmed OA subset corpus*, composé d'environ 1,7 millions d'articles disponibles en XML. URL : <https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>

- Atanassova, I., Bertin, M., and Larivière, V. (2016). On the composition of scientific abstracts. *Journal of Documentation*, 72(4):636–647.
- Atanassova, I., François-C., R., and Bertin, M. (2018). Studying Uncertainty in Science: a distributional analysis through the IMRaD structure. *Conférence 7th International Workshop On Mining Scientific Publications (WOSP) - 11th edition of the Language Resources and Evaluation Conference (LREC 2018)*, 7-12 Mai 2018, Miyazaki, Japon.
- Bally, C. (1965). *Linguistique générale et linguistique française*. 4e éd. revue et corrigée. Editions Francke, Berne, Suisse.
- Bernhard, D. and Ligozat, A.-L. (2011). Analyse automatique de la modalité et du niveau de certitude : application au domaine médical. *Actes de TALN*, 1 : 433-444.
- BIPM (2006). *Le système international d'unités (SI) - 8e édition*. Technical report, Bureau international des poids et mesures (BIPM), Sèvres, France.
- Chauveau-Thoumelin, P. and Grabar, N. (2014). La subjectivité dans le discours médical : sur les traces de l'incertitude et des émotions. *Revue des Nouvelles Technologies de l'Information*, RNTI-E-26 : 455-466.
- Chen, C., Song, M., and Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *Journal of Informetrics*, 12(1):158-180.
- Davis, P. K., Perry, W. L., Hollywood, J. S., and Manheim, D. (2016). Uncertainty-sensitive heterogeneous information fusion - assessing threat with soft, uncertain, and conflicting evidence. Technical report, RAND Corporation, Santa Monica, Californie.
- Dragos, V. (2013). An ontological analysis of uncertainty in soft data. In *Information Fusion (FUSION)*, 2013 16th International Conference on, pages 1566-1573. IEEE.
- Fusco, G. (2013). L'incertitude dans les sciences géographiques : spécificités, schémas d'appréhension. 1e Table Ronde «L'incertitude : regards pluridisciplinaires», Nice, 19 juin 2013, MSHS Sud-Est, Saint-Jean d'Angély 3, Projet «Faire Science avec l'Incertitude», Axe 4 : «Territoires, systèmes techniques et usages sociaux».
- Fusco, G., Bertoncello, F., Candau, J., Emsellem, K., Huet, T., Longhi, C., Poinat, S., Primon, J.-L., and Rinaudo, C. (2015). Faire science avec l'incertitude : réflexions sur la production des connaissances en Sciences Humaines et Sociales. In *Incertain et connaissances en SHS : production, diffusion, transfert*, page 17, Nice, France. Maison des Sciences de l'Homme et de la Société Sud-Est (MSHS) - Axe 4 : Territoires, systèmes techniques et usages sociaux.

- Goujon, B. (2009). Uncertainty detection for information extraction. In Proceedings of the International Conference RANLP-2009, pages 118-122.
- Halliday, M. A. K. and Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar*. Routledge, Abingdon, Royaume-Uni, 4^e édition.
- Haute Autorité de Santé (2013). *Etats des lieux : Niveau de preuve et gradation des recommandations de bonne pratique*.
- IPCC (2005). *Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties*. Technical report, Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- Knight, F. H. (1921). *Risk, uncertainty and profit*. Houghton Mifflin Co. - The Riverside Press, Boston MA, EUA, 1st edition.
- Le Querler, N. (1966). *Typologie des modalités*. Presses universitaires de Caen.
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., and Zwiers, F. W. (2010). *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. Technical report, Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- Meunier, A. (1981). *Grammaire du français et modalités, matériaux pour l'histoire d'une nébuleuse*. DRLAV - Documentation et recherche en linguistique allemande contemporaine, Université de Paris XIII-Vincennes, (25):119–144.
- Moncecchi, G., Minel, J.-L., and Wonsever, D. (2012). Improving speculative language detection using linguistic knowledge. In Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics, pages 37–46. Association for Computational Linguistics.
- NATO (2003). *NATO STANAG 2511 Intelligence Reports*.
- Newell, D. B., Cabiati, F., Fischer, J., Fujii, K., Karshenboim, S. G., Margolis, H. S., Mirandés (de), E., Mohr, P. J., Nez, F., Pachucki, K., Quinn, T. J., Taylor, B. N., Wang, M., Wood, B. M., and Zhang, Z. (2018). The codata 2017 values of h, e, k, and NA for the revision of the SI. *Metrologia*, 55(1):L13.X
- Perkins, M. R. (1983). *Modal expressions in English*. Frances Pinter.
- Rubin, V. L., Liddy, E. D., and Kando, N. (2006). Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In Shanahan, J. G., Qu, Y., and Wiebe, J., editors, *Computing Attitude and Affect in Text*:

Theory and Applications, chapter 7, pages 61–76. Springer Netherlands, Dordrecht.

Vold, E. T. (2008). Modalité épistémique et discours scientifique - Une étude contrastive des modalisateurs épistémiques dans des articles de recherche français, norvégiens et anglais, en linguistique et médecine. PhD thesis, Université de Bergen, Norvège.

Abstract

The problem of uncertainty in sciences is related to the discipline, to the object of study and to the methodologies used. However, taking into account scientific uncertainty is an essential challenge for society, together with the need of creating tools for the analysis of scientific discourse. The problem that we address in this paper is the definition, typology and understanding of uncertainty, through an interdisciplinary view. We aim at studying a corpus of scientific papers on climate change in order to construct an ontology of uncertainty in science.