



Ontologie de l'incertitude et annotation sémantique d'un corpus autour du changement climatique

François-C. Rey

► To cite this version:

François-C. Rey. Ontologie de l'incertitude et annotation sémantique d'un corpus autour du changement climatique. Linguistique. Université Bourgogne Franche-Comté, 2022. Français. <NNT : 2022UBFCC007>. <tel-03780719v2>

HAL Id: tel-03780719

<https://hal.science/tel-03780719v2>

Submitted on 28 Feb 2023

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.



HAL Authorization

THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE

PREPAREE A L'UNIVERSITE DE FRANCHE-COMTE

Ecole doctorale n° 592

Lettres, Communication, Langues, Arts - LECLA

Doctorat en sciences du langage

Par

M. Rey François-Claude G. J. H.

**Ontologie de l'incertitude et annotation sémantique d'un corpus
autour du changement climatique**

Thèse présentée et soutenue à Besançon, le 20 mai 2022.

Composition du Jury :

M ^{me} CARDEY Sylviane	Pr ^e émérite, Université de Franche-Comté Institut Universitaire de France (IUF)	Présidente
M. POGNAN Patrice	Pr ^e émérite, INALCO	Rapporteur
M. ROCHE Christophe	Pr ^e des universités, Université Savoie Mont-Blanc	Rapporteur
M. KEMBELLEC Gérald	MCF, Conservatoire National des Arts et Métiers	Examineur
M ^{me} ATANASSOVA Iana	MCF HDR, Université de Franche-Comté Institut Universitaire de France (IUF)	Directrice de thèse
M. BERTIN Marc	MCF, Université Claude Bernard Lyon 1	Codirecteur de thèse
M ^{me} DESFRICHES DORIA Orélie	MCF, Université Paris 8 Vincennes - Saint-Denis	Invitée

Titre : Ontologie de l'incertitude et annotation sémantique d'un corpus autour du changement climatique

Mots clefs : incertitude, articles scientifiques, ontologie, annotation sémantique, grammaire formelle, changement climatique

Résumé : Disposer d'outils permettant l'analyse du discours scientifique est essentiel pour répondre à des problématiques sociétales actuelles. L'incertitude fait partie intégrante du processus de recherche scientifique. Cette dernière s'exprime dans les textes des articles scientifiques sous la forme de marqueurs linguistiques.

La présente thèse introduit une ontologie de l'incertitude scientifique à travers l'étude

d'un corpus multidisciplinaire d'articles sur le changement climatique en anglais, en prenant en compte la diversité et la complexité de cette notion dans les différentes disciplines. Nous proposons une méthodologie pour l'annotation des expressions de l'incertitude qui s'appuie sur une grammaire formelle et des règles linguistiques. Nous avons produit un corpus annoté de 1000 phrases - Gold Standard.

Title: Ontology of uncertainty and semantic annotation of a corpus dealing with climate change

Keywords: uncertainty, scientific papers, ontology, semantic annotation, formal grammar, climate change

Abstract: Having tools for the analysis of scientific discourse is essential to address current societal issues. Uncertainty is an integral part of the scientific research process. It is expressed in the texts of scientific articles in the form of linguistic cues.

This thesis introduces an ontology of scientific uncertainty through the study of a multidisci-

plinary corpus of articles on climate change in English, taking into account the diversity and complexity of this notion in the different disciplines. We propose a methodology based on a formal grammar and linguistic rules for the annotation of expressions of uncertainty, as well as an annotated corpus of 1000 sentences - Gold Standard.

Remerciements

Je remercie les membres de mon jury de thèse, M^{me} Sylviane Cardey, Professeure émérite à l'Université de Franche-Comté et membre de l'Institut Universitaire de France (IUF), pour l'honneur qu'elle m'a fait en acceptant d'être la présidente du jury de cette thèse ; M. Patrice Pognan, Professeur émérite à l'Institut National des Langues et Civilisations Orientales (INALCO), et M. Christophe Roche, Professeur des universités à l'Université Savoie Mont-Blanc, pour l'honneur qu'ils m'ont fait en acceptant d'être les rapporteurs de cette thèse ; M. Gérard Kembel-lec, Maître de conférence au Conservatoire National des Arts et Métiers (CNAM), pour l'honneur qu'il m'a fait en acceptant d'être examinateur de cette thèse ; et M^{me} Orélie Desfriches Doria, Maîtresse de conférence à l'Université Paris 8 Vincennes - Saint-Denis, pour l'honneur qu'elle m'a fait en acceptant de distinguer la soutenance de cette thèse par sa présence en tant qu'invitée.

Je remercie, pour leur soutien bienveillant et pour leur exemplarité scientifique, M^{me} Iana Atanassova, Maîtresse de conférences HDR à l'Université de Franche-Comté et membre de l'Institut Universitaire de France (IUF), qui a accepté de diriger cette thèse, et M. Marc Bertin, Maître de conférences à l'Université Claude Bernard Lyon 1, qui a accepté de la co-diriger.

Je remercie M^{me} Aurélie Nomblot et M. Divin Angapay, dont l'expertise de la langue anglaise et la collaboration dans les débats et tâches minutieuses d'annotation de la première partie du corpus ont permis de faire avancer la réflexion et l'expérimentation pour cette thèse.

Je remercie les membres de mon comité de suivi de thèse, dont l'écoute et les questionnements ont contribué à structurer ma réflexion sur ce qu'est une thèse et son contexte scientifique.

J'adresse également mes remerciements à celles et ceux qui ont aimablement enrichi mon parcours commençant de chercheur en partageant des responsabilités et des moments joviaux en parallèle à cette thèse, et en particulier aux membres du Centre de Recherche Lucien Tesnière et Centre de Recherches Interdisciplinaires et Transculturelles CRIT, aux membres du Comité de rédaction de la revue scientifique Eclats, aux membres de l'association ThèsEnvies, aux membres de la représentation des doctorantes et doctorants, aux membres du Conseil de l'Ecole

doctorale LECLA, aux membres du Bureau du Collège doctoral UBFC y compris les responsables de la Mission doctorale, aux membres du Conseil Académique UBFC, aux membres du projet Interreg France-Suisse DecRIPT, aux membres du Comité AFNOR X03A Terminologie principes et coordination, et aux membres du COPIL Ecoresponsabilité de l'Université de Franche-Comté.

Enfin, je remercie ma famille et mes amies et amis pour leur compréhension, en particulier leur compréhension de l'adage qui invite à ne pas demander à un doctorant où en est sa thèse.

Sommaire

1	Introduction	1
1.1	Objectifs de la thèse	2
1.2	Contexte scientifique	4
1.3	Plan de la thèse	4
2	Pourquoi étudier l'incertitude ?	7
2.1	Communiquer l'incertitude scientifique	7
2.2	Pouvoir explorer l'incertitude scientifique	8
2.3	Prise en compte de l'incertitude par des organisations	10
2.3.1	La communication scientifique pour l'aide à la décision	13
2.3.2	La gestion et diffusion du renseignement	17
2.3.3	La communication scientifique auprès du public général	19
3	Le concept d'incertitude en science	23
3.1	Définitions de l'incertitude	23
3.1.1	Définition générale de l'incertitude	24
3.1.2	Sources de l'incertitude en science	25
3.2	Les modalités de l'incertitude : un point de vue linguistique	38
3.2.1	Approche des définitions des modalités	44
3.2.2	Les modalités épistémiques	48
3.3	Typologie conceptuelle de l'incertitude	52
3.3.1	Le <i>hedging</i>	53
3.3.2	Le <i>boosting</i> vs l'atténuation	59
3.3.3	Les opinions et sentiments	60
3.3.4	L'ambiguïté	62
3.3.5	L'évidentialité	64
3.3.6	L'élicitation	67
3.3.7	La prise en charge	69

3.3.8	La spéculation	70
3.4	La notion de subjectivité	71
3.4.1	Définition générale de la subjectivité	72
3.4.2	La notion de subjectivité en philosophie	73
3.4.3	La notion de subjectivité en linguistique	74
4	La prise en compte de l'incertitude dans le domaine du TAL	77
4.1	Dimensions des expressions de l'incertitude	77
4.1.1	Des dimensions des modalités	77
4.1.2	La polarité	78
4.1.3	Schéma des dimensions de l'incertitude	79
4.2	Corpus <i>ad-hoc</i> pour le traitement de l'incertitude	80
4.3	Méthodes linguistiques pour l'annotation de textes écrits	81
4.3.1	L'approche microsystemique	82
4.3.2	La méthode d'Exploration Contextuelle	84
4.4	Marqueurs des expressions de l'incertitude	86
4.5	Métadonnées de l'incertitude	89
5	Vers une ontologie de l'incertitude scientifique	93
5.1	Les ontologies informatiques et le web sémantique	93
5.2	Présence de l'incertitude dans des ontologies existantes	96
5.3	Une définition de l'incertitude	99
5.4	Les raisonnements formalisant l'incertitude	101
5.4.1	Le raisonnement abductif	101
5.4.2	Comparaison des raisonnements formalisés	105
5.4.3	Du raisonnement à l'incertitude	107
5.5	Proposition d'une ontologie de l'incertitude	109
6	Constitution de corpus pour le traitement de l'incertitude	111
6.1	Principales étapes du travail	111
6.2	Contexte de l'incertitude : les articles scientifiques	113
6.3	Choix du changement climatique comme domaine d'étude	121
6.4	Constitution d'une liste d'expressions clefs du domaine	126
6.5	Constitution du corpus d'articles du domaine	128
6.5.1	Source et critères pour les choix des articles	128
6.5.2	Composition du corpus d'articles	129
6.5.3	Exploration préliminaire	129
6.6	Constitution du corpus d'étude et du corpus d'évaluation	130

6.6.1	Extraction automatique de 500 phrases	130
6.6.2	Extraction manuelle de 500 phrases	131
6.7	Proposition d'un corpus annoté - <i>gold standard</i>	131
6.7.1	Annotation manuelle	132
6.7.2	Corpus annoté final - <i>gold standard</i>	138
6.7.3	Traçabilité de l'article d'origine de chaque phrase	139
7	Règles linguistiques pour l'annotation de l'incertitude	145
7.1	Motivations pour la création d'une grammaire formelle	145
7.2	Définition de la grammaire	147
7.2.1	Caractères réservés	148
7.2.2	Observations sur les caractères	150
7.3	Définition des règles	150
7.4	Conception des règles	151
7.5	Exemples de règles	152
7.6	Exemples de modifications de règles	153
7.7	Les expressions contenant « <i>may be</i> »	159
7.8	Algorithme d'exécution des règles	160
8	Résultats et évaluation des annotations	163
8.1	Protocole d'évaluation du système	163
8.2	Expérimentations lors de la phase de construction des règles	165
8.2.1	Expérimentation sur le corpus d'étude (partie 1)	165
8.2.2	Expérimentation sur le corpus d'étude (partie 2)	168
8.3	Evaluation	170
9	Conclusion	175
	Bibliographie	178
	Liste des tableaux	208
	Liste des schémas	210
A	Exemples de définitions	211
A.1	Exemples de définitions de la notion de modalité	211
A.2	Exemples de définitions de la notion de doute	213
A.3	Exemples de définitions de la notion d'incertitude	215

A.4	Exemples de définitions autour de la notion d'incertitude liée au changement climatique	217
A.5	Exemples de définitions de la notion de subjectivité en linguistique	218
B	Liste d'expressions clefs liées au changement climatique en anglais	219
C	Mode d'emploi fourni aux experts pour l'annotation manuelle	225
D	Mode de calcul des indices moyens	229
E	Algorithme d'annotation : exécution des règles linguistiques sur des corpus de phrases	233
F	Règles d'annotation de l'incertitude	239
G	Corpus de phrases annotées - <i>gold standard</i>	247
G.1	Corpus d'étude - partie 1 (500 phrases)	247
G.2	Corpus d'étude - partie 2 (200 phrases)	267
G.3	Corpus d'évaluation (300 phrases)	275

Chapitre 1

Introduction

Les articles scientifiques sont par nature des documents écrits qui ont pour vocation la publication et la diffusion de résultats originaux de recherches scientifiques. Ces textes jouent un rôle crucial dans la production de nouveau savoir, car ils sont le vecteur premier dans les échanges entre scientifiques, mais aussi dans la validation des produits de la recherche. La nature de la recherche scientifique ayant pour but une objectivité et une neutralité dans le raisonnement, l'expression de celle-ci dans les articles scientifiques est dotée, pour la plupart, par ces mêmes caractéristiques. Par ailleurs, ce sont des documents hautement structurés, organisés suivant des modèles établis afin de faciliter l'accès à l'information et mettre en avant l'enchaînement des idées ainsi que la structure argumentative. Parmi ces modèles, le plus utilisé dans les sciences expérimentales est IMReD (Introduction, Méthodes, Résultats et Discussion).

La production de nouvelles connaissances doit s'appuyer sur une démarche méthodologique rigoureuse en fonction de l'objet d'étude et de son champ disciplinaire. Néanmoins, ce système accepte des nuances. Par exemple, dans le domaine biomédical, la notion de niveau de preuve scientifique apporté par une étude selon la force du protocole mis en œuvre est une évaluation des études selon des critères telles 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 et les conclusions des auteurs, etc. Les articles scientifiques, à travers ce que l'on nomme les grades de recommandation : Preuve scientifique établie - Présomption scientifique - Faible niveau de preuve scientifique pour le domaine médico-scientifique, relaient donc une part d'incertitude.

Pour cela, les auteurs d'un article scientifique peuvent utiliser divers procédés linguistiques pour moduler la certitude (en anglais « hedging »). Les textes scien-

tifiques relaient donc de la subjectivité volontaire ou non, et ils peuvent exprimer des ambiguïtés, des intuitions, des affects pouvant refléter des prises de position favorables ou non. Ce peut être particulièrement le cas dans la conclusion ou la partie de discussion d'un article, puisqu'elle sert naturellement à évoquer des perspectives, des hypothèses qui restent à examiner dans de 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 vers les limites de la connaissance actuelle et les domaines que la science doit encore explorer. Dans quelle mesure cette incertitude est-elle détectable, qualifiable, réutilisable ?

Reconnaître, recueillir et présenter l'incertitude revêt une importance considérable dans les activités scientifiques, puisque cela permet de savoir, avec plus d'acuité et de perspective, quels sont les ressorts, les perceptions et les tendances de développement des recherches en cours. C'est utile pour compléter les informations affirmées dans un cadre strict, mais aussi pour comprendre les motivations des chercheurs et les aspects créatifs de la recherche. Un outil permettant d'identifier et classer ces phénomènes dans des textes aurait de nombreuses applications pour les chercheurs : produire et compléter des états de l'art, analyser un champ disciplinaire, analyser l'échelle temporelle des recherches et décidabilité de la prospective technique et économique, et automatiser la reconstitution de la genèse de concepts scientifiques au cours d'une période par une approche diachronique.

1.1 Objectifs de la thèse

La présente étude porte sur l'expression de l'incertitude dans les textes scientifiques : il s'agit d'identifier les différents types d'incertitude scientifique et de les classer, automatiquement. Cela implique d'extraire et de présenter les segments de textes porteurs d'une expression de l'incertitude scientifique. Les objectifs de ce travail peuvent être formulés selon les axes suivants :

- Etudier la notion de l'incertitude, et plus particulièrement l'incertitude scientifique, à travers ses définitions et les concepts qui sont liés dans une perspective multidisciplinaire.
- Produire une analyse linguistique de corpus d'articles scientifiques afin de rendre compte des expressions linguistiques qui sont mobilisées pour exprimer les incertitudes.

- Proposer une première ontologie de l’incertitude scientifique.
- Construire un prototype d’un outil informatique permettant l’identification et classification des incertitudes dans les articles, par une modélisation linguistique à travers des règles de repérage.

L’objectif principal dans cette thèse est l’étude de l’incertitude sur des faits scientifiques, qui est exprimée explicitement par les auteurs dans le texte des articles, par une modélisation linguistique à partir de l’étude de corpus (voir schéma 1.1¹). Notre but est également de proposer une méthode linguistique afin d’identifier automatiquement les segments textuels porteurs d’incertitude. Cela signifie identifier des expressions qui indiquent l’incertitude, en prenant en compte la complexité de cette notion et les différents types et sources d’incertitude qui peuvent être présents dans les contenus des articles.

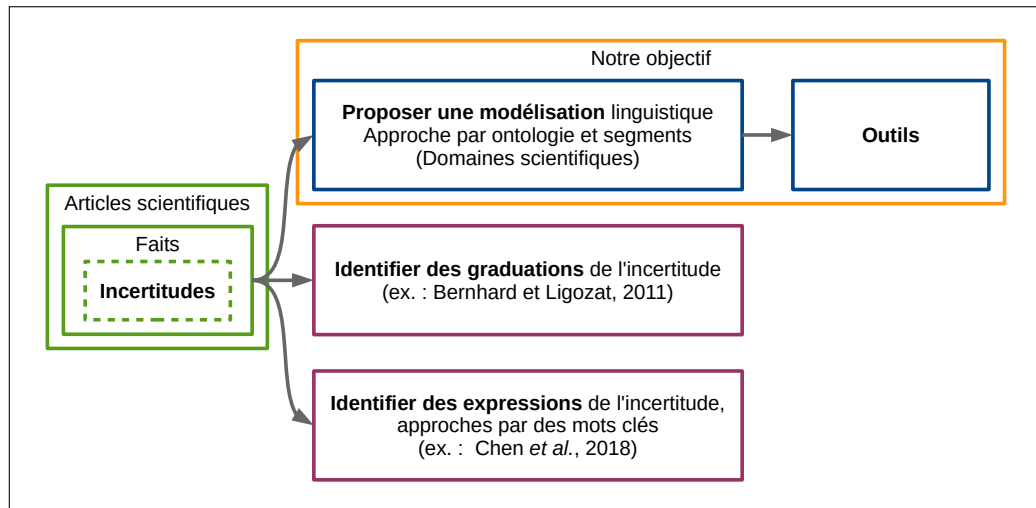


Schéma 1.1: Objectifs.

Un autre objectif complémentaire est de reconnaître et prendre en compte des particularités de l’incertitude dans les différents domaines scientifiques et d’en tirer des conséquences méthodologiques et pratiques pour leur analyse. Plus généralement, il s’agit de capturer l’univers référentiel, la sémantique et la phraséologie de l’incertitude scientifique dans une ontologie de l’incertitude, afin qu’elle puisse être exploitée pour une recherche d’information enrichie dans des textes scientifiques.

¹Références bibliographiques incluses dans le schéma 1.1 : [Bernhard et Ligozat, 2011] et [Chen *et al.*, 2018].

1.2 Contexte scientifique

Cette étude se situe dans un ensemble plus large de travaux sur les textes scientifiques menés au Centre de Recherche Lucien Tesnière, du Centre de Recherches Interdisciplinaires et Transculturelles EA 3224 (CRIT), en particulier concernant le traitement automatique et la valorisation des contenus des articles scientifiques et des langues de spécialité.

Aujourd’hui, de plus en plus d’articles sont disponibles en *open access*, ce qui rend possible la mise en place de traitements automatisés à grande échelle. De plus, le développement du traitement automatique des langues permet d’envisager la reconnaissance, l’extraction et la transformation d’informations complexes qui prennent en compte la sémantique textuelle au delà des mots clés. Il en résulte l’opportunité d’envisager le filtrage et l’exploitation massive des contenus d’articles scientifiques.

Le terrain d’expérimentation choisi ici est le domaine spécifique du changement climatique, toutes disciplines scientifiques confondues. Ce domaine définit un périmètre assez large, qui offre une possibilité d’aborder l’incertitude d’un point de vue multidisciplinaire.

Une partie des résultats présentés dans cette thèse ont été publiés dans des ateliers et conférences internationales, à comité de lecture : [[Rey et al., 2018](#), [Atanassova et al., 2018](#), [Atanassova et Rey, 2021](#)].

1.3 Plan de la thèse

Le reste du contenu de la thèse est organisé de la manière suivante :

- Le Chapitre 2 présente la motivation pour étudier l’incertitude scientifique à travers l’examen de la place de cette notion dans les processus de recherche et sa prise en compte par des organisations.
- Le Chapitre 3 détaille les définitions de l’incertitude et les concepts liés d’un point de vue linguistique.
- Le Chapitre 4 est dédié au Traitement Automatique des Langues et les travaux qui existent dans ce domaine autour du traitement de l’incertitude.
- Le Chapitre 5 propose une première ontologie de l’incertitude à travers l’examen, principalement, des différents raisonnements qui impliquent des incertitudes.

- Le Chapitre 6 explique notre démarche pour la construction des corpus d'étude et corpus d'évaluation, ainsi que pour leur annotation manuelle.
- Le Chapitre 7 présente notre approche pour la construction des règles linguistiques pour le repérage et annotation des expressions de l'incertitude dans les textes.
- Le Chapitre 8 présente des résultats des traitements et des évaluations que nous avons menées lors de notre expérimentation.
- Enfin, le Chapitre 9 propose une conclusion.

Chapitre 2

Pourquoi étudier l'incertitude ?

2.1 Communiquer l'incertitude scientifique

Quels rôles jouent la subjectivité et l'incertitude dans la science ? Comment s'exprime l'incertitude présente dans la science, mais aussi comment l'appréhender ? L'incertitude fait partie des objets, des résultats et des théories, dans toutes les disciplines : causée par des mesures imparfaites, des propositions réfutables, des modèles incomplets, etc., et fait même partie du caractère arbitraire de certains choix d'analyse scientifique [Sluijs (van der) *et al.*, 2006a, p. 5].

L'astrophysicien Jean Lilensten [Lilensten, 2018, p. 94, 96], à propos de ce que *science* signifie, explique que « l'incertitude est inhérente à la mesure », qu'elle en fait partie quelque soit la mesure et l'échelle, et que « C'est l'homme qui introduit des valeurs dites exactes qui n'ont rien à voir avec la physique ». Pour le sociologue Edgar Morin [Morin, 1980, p. 9, 224, 229 et 357], la pensée doit se nourrir d'incertitude « au lieu d'en mourir » :

« l'intelligence [...] s'aventure dans le flou, l'incertain, l'ambigu, mais doit aussi vérifier le trop certain, le trop bien connu », « la connaissance doit savoir à la fois combattre et utiliser l'incertitude », et « La problématique de la pensée complexe est, non pas d'éliminer, mais de travailler avec le paradoxe, l'incertitude, le désordre [...] ».

Allant plus loin, dans [Morin, 1999, p. 8, 12 et 46], il indique que l'incertitude est une condition de survie de la rationalité, car elle entretient la vigilance auto-critique de cette dernière. Il nomme ainsi un *principe d'incertitude rationnel* qui protège la rationalité d'une dégradation en une rationalisation ; et il met en avant

un rôle immunitaire¹ de l'incertitude².

S'intéresser aux incertitudes qui s'expriment dans un domaine scientifique, ou dans les sciences, est s'intéresser à leur diversité : les concepts, types, dimensions, degrés et prises en compte de l'incertitude diffèrent selon les domaines et les recherches scientifiques. Concernant la diversité des moyens d'expression de l'incertitude, la présente recherche se centre sur l'expression de l'incertitude dans les sciences par des moyens textuels, mais, bien que ces moyens textuels soient omniprésents, il ne sont pas exclusifs : divers moyens non textuels peuvent aussi être mis en œuvre pour exprimer l'incertitude et ses dimensions dans les sciences, par exemple les graphiques basés sur des données numériques, comme ceux présentés par [Cooke et Noortwijk (van), 1999], ou des méthodes de visualisation basées sur des données qualitatives, comme celles expérimentées par [Boukhelifa *et al.*, 2012].

2.2 Pouvoir explorer l'incertitude scientifique

À travers et en liaison avec la science, la connaissance et l'utilisation de l'incertitude est une constituante fondamentale d'activités humaines diverses, par exemple dans l'ingénierie, l'entrepreneuriat, la politique, le sport, etc. En science et dans les champs scientifiques la diversité des intérêts pour l'incertitude est remarquable par la diversité d'origine des publications à son sujet : sorte de lapin blanc suivi par Alice³ dans l'enceinte du jardin, l'incertitude constitue à la fois une cause de la recherche du savoir, un moyen de construction du savoir, une partie de ce savoir, et un objet d'étude en elle-même.

Le phénomène de nécessité scientifique de l'incertitude se reflète par exemple chez [Chen et Song, 2017, p. v-vi et 28-31] lorsqu'ils affirment qu'une connaissance n'aurait probablement pas de valeur dans un champ de recherche si elle n'était soumise à l'incertitude⁴, lorsqu'ils clament⁵ une équivalence entre expertise et connaissance de l'incertitude, et lorsqu'ils lient la compréhension des incertitudes basée sur les méthodes de mesure d'indices dans la recherche à cinq des « Grands Défis » de la science qu'ils nomment comme suit :

¹Ce terme est notre choix.

²Élimination de la « connaissance simpliste » et détoxification de la « connaissance complexe ».

³Les personnages du mathématicien Lewis Carroll, dans son roman « *Alice's Adventures in Wonderland* ».

⁴« *Knowledge that is free from uncertainty probably has no value in a research field* ».

⁵Terme employé par ces auteurs.

- **accessibilité de la littérature scientifique** (*Accessibility*) : il s'agit d'avoir accès à temps aux documents utiles au cœur de son domaine de recherche pour être à jour, mais aussi de pouvoir accéder à d'autres avancées de la connaissance que des liens rendent intéressantes puis pertinentes bien que leurs domaines de recherche soient éloignés du domaine de départ, ce qui dépend des limitations humaines et techniques face à l'apparition continue des publications et à leur évolution. Pour les auteurs, les approches qui aident les chercheurs à trouver des documents scientifiques pertinents devraient prendre en compte des liens de dimensions diverses allant de la linguistique au social ; et les résultats de ces approches devraient aussi élargir les perspectives des chercheurs sur ce que peut signifier leur recherche⁶.
- **intelligibilité de l'incertitude** (*Clarity on Uncertainty*) : il s'agit de pouvoir comprendre et communiquer l'incertitude scientifique, ce qui est crucial « en particulier dans domaines où l'information est incomplète, contradictoire ou complètement manquante », par exemple dans des domaines préoccupants généralisés tels que le virus Ebola ou le changement climatique. La nouveauté, par exemple dans des résultats d'expériences, peut avoir pour effet l'introduction a posteriori d'incertitude sur des interprétations de résultats, des théories et des affirmations précédentes. L'incertitude est aussi une méta connaissance dont l'importance est d'autant plus grande dans la recherche en cas de nouveauté ou en cas de crise, ce qui pourrait permettre d'observer des cycles et des tendances de la connaissance.
- **mise en relation de perspectives diverses** (*Connecting Diverse Perspectives*) : la découverte de relations innovantes entre divers domaines du savoir, par exemple par analogie, est particulièrement propice aux découvertes, et permet de valoriser des divergences. C'est pour cela que de nouveaux outils informatisés devraient être mis en œuvre pour élargir les perspectives des chercheurs, en s'appuyant sur la présentation des faits et sur les raisonnements.

⁶Les auteurs citent en exemple les huit questionnements d'évaluation des propositions de projets de recherche, mis en place à l'agence DARPA d'innovation pour la sécurité nationale des EUA, connus sous le nom de *Catéchisme de Heilmeier* : « *What are you trying to do? Articulate your objectives using absolutely no jargon* », « *How is it done today, and what are the limits of current practice?* », « *What is new in your approach and why do you think it will be successful?* », « *Who cares? If you are successful, what difference will it make?* », « *What are the risks?* », « *How much will it cost?* », « *How long will it take?* » et « *What are the mid-term and final "exams" to check for success?* ». Disponibles sur le site de la DARPA consulté le 23/09/2020 : <https://www.darpa.mil/work-with-us/heilmeier-catechism>.

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

- **évaluations de performances et tests de référence** (*Benchmarks and Gold Standards*) : des cas de références, diversifiés et à jour, incluant les découvertes et controverses importantes, devraient être fournis aux chercheurs pour élaborer, « tester et calibrer leurs outils de mesure et d'analyse », suivant l'exemple du biomédical.
- **intégration des méthodes scientifiques de mesure et d'analyse des données** (*Integrating Scholarly Metrics and Analytics*) : pouvoir évaluer les résultats et activités scientifiques, quelque soit leur domaine, en intégrant ensemble les aspects quantitatifs et qualitatifs, par exemple dans des indices sur l'activité des auteurs de la connaissance scientifique, est un défi. La normalisation est une piste qui a été suivie, mais c'est la facilité à alterner les perspectives qui le permettra.

2.3 Prise en compte de l'incertitude par des organisations

Dans cette section, il s'agit de dégager un aperçu d'ancrages concrets de la prise en compte de l'incertitude dans des organisations, c'est-à-dire de découvrir en partie « comment cela se passe sur le terrain ».

En fonction des besoins, les scientifiques catégorisent différemment les incertitudes. Par exemple, l'Autorité européenne de sécurité des aliments (AESA)⁷, [Hart *et al.*, 2019, p. 3], retient huit types d'expressions de l'incertitude issus du processus d'analyse d'incertitude que sa Commission scientifique décrit dans [Benford *et al.*, 2018] :

- conclusions sans réserve (« *unqualified* ») sans expression d'incertitude,
- description d'une source d'incertitude,
- description qualitative de la direction et/ou magnitude de l'incertitude,
- évaluation non concluante,
- probabilité précise,
- probabilité approximative,

⁷European Food Safety Authority - EFSA.

- distribution de probabilité,
- distribution de probabilité bi-dimensionnelle.

La Commission scientifique de l'AESA, [Benford *et al.*, 2018, tableau 1, p. 19], présente deux listes, incluses dans des conseils procéduraux d'identification des incertitudes destinés aux évaluateurs : une liste des incertitudes associées aux entrées d'évaluation :

- ambiguïté,
- exactitude et précision des mesures,
- incertitude d'échantillonnage,
- données manquantes dans les études,
- études manquantes,
- hypothèses (« *assumptions* ») sur les entrées,
- estimations statistiques,
- incertitude liée à l'extrapolation (p. ex. des limitations dans la validité externe),
- autres incertitudes,

et une liste des incertitudes associées à la méthodologie d'évaluation :

- ambiguïté,
- facteurs exclus,
- hypothèses distributionnelles,
- usage de valeurs fixes,
- relation entre des parties de l'évaluation,
- attestation (« *evidence* ») de la structure d'évaluation,
- incertitudes liées au processus de prise en compte de preuves dans les publications,

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

- jugement expert,
- calibration ou validation avec des données indépendantes,
- dépendance entre sources d'incertitude,
- autres incertitudes.

Si l'on considère l'utilisation de l'incertitude, la façon dont elle est déjà prise en charge dans la pratique aura des conséquences concrètes sur les attentes de ceux qui doivent suivre ou tenir compte de cette pratique dans des domaines d'application divers. D'un autre côté, si l'on considère la recherche comme le suggère [Poibeau, 2011, p. 173-176], le réalisme des solutions proposées en traitement automatique des langues dépendra des utilisateurs ; que ces utilisateurs soient des experts pouvant se confronter aux approches des problèmes adoptées par la recherche, ou qu'ils soient des utilisateurs finaux ayant des attentes sur les règles du langage. Cela vaut aussi si l'utilisation de l'incertitude se fait par des machines : par exemple, à propos de la gestion des ressources hydriques, [Loucks et Beek (van), 2017, chap. 7 et 8] remarquent que la prise en charge de l'incertitude par les modèles peut se répercuter sur le réalisme des résultats de simulations, ce qui importe pour les décisions qui se basent sur les résultats.

L'incertitude est un problème auquel sont confrontées en particulier certaines organisations dans leur fonctionnement, et qu'elles ont choisi de traiter en mettant en place leur propre méthodologie, ou en adoptant une solution préexistante. C'est par exemple le cas d'organisations des institutions politiques, militaires et de la santé :

- c'est le cas de l'organisation politique internationale *Intergovernmental Panel on Climate Change (IPCC)*⁸ qui doit comprendre et rapporter avec exactitude ce que lui communiquent les scientifiques, pour éclairer les décisions dans le cadre de la crise climatique planétaire ;
- c'est le cas de forces armées qui doivent recueillir, gérer et diffuser des renseignements en interne (les informations sur leurs méthodes sont en quantité ou en accès limité ; mais les évoquer avec un aperçu nous a semblé suffisamment intéressant) ;
- et c'est le cas d'autorités de la santé comme la Haute Autorité de Santé (HAS) et le Conseil National de l'Ordre des Médecins qui doivent s'assurer

⁸Groupe Intergouvernemental sur le Changement Climatique (GIEC).

de la qualité et de la compréhension des informations relatives à la santé mises à la disposition du public.

Les éléments présentés à continuation abordent des problématiques et des solutions retenues par ces organisations pour fonctionner avec l'incertitude. Ces pratiques suggèrent, entre autres, que caractériser l'incertitude ou certains types d'incertitudes, la représenter et la réutiliser scientifiquement ou techniquement, a mené à l'utilisation d'échelles terminologiques, parfois sur plusieurs dimensions concomitantes telles que, par exemple, la probabilité et la confiance.

2.3.1 La communication scientifique pour l'aide à la décision

[Lipshitz et Strauss, 1997, p. 149-152 et 158-161], mettent en relief la diversité des conceptualisations de l'incertitude dans sa prise en charge au cours des processus de décision. Il considèrent en particulier l'incertitude comme « *a sense of doubt that blocks or delays action* » ; ce doute étant lié à des manques de compréhension, de différenciation des alternatives, et d'information. Pour représenter comment les décideurs procèdent face aux différents types d'incertitudes, ils font l'hypothèse d'une méthode heuristique d'adaptation aux incertitudes « principalement déterminée par la nature ou qualité de l'incertitude », RAWFS : *Reduction, Assumption-based reasoning, Weighing pros and cons, Forestalling, et Suppression*.

L'IPCC, organisation internationale sous l'égide de l'OMM⁹ et du PNUE¹⁰, doit transmettre avec « le degré de précision approprié » l'information scientifique sur le changement climatique, notamment dans des rapports écrits où les mêmes expressions employées par les rédacteurs pour exprimer des niveaux d'incertitude scientifiques doivent renvoyer aux mêmes valeurs terminologiques. De plus, ces incertitudes doivent pouvoir être traduites fidèlement dans les langues très diverses des usagers de cette information au niveau mondial. [IPCC, 2005] fournit une typologie des incertitudes à l'attention des auteurs principaux de rapports présentant des données scientifiques sur le changement de climat (voir tableau¹¹ 2.1).

⁹Organisation météorologique mondiale (*World Meteorological Organization (WMO)*).

¹⁰Programme des Nations Unies pour l'Environnement (*United Nations Environment Programme (UNEP)*).

¹¹Note de traduction du tableau : Le *bootstrap* « à pour objectif d'estimer les incertitudes sur des quantités statistiques » par tirage de nouveaux échantillons de même taille avec remise. (Source principale : « Glossaire » in *CHOCS - Revue scientifique et technique de la direction des applications militaires, 2017, no 48 « Maîtrise des incertitudes »*, p. 94. CEA - DAM, Institut supérieur des études nucléaires de défense (ISENDé), Bruyères-le-Châtel, France.).

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

Type	Exemples de sources	Approches typiques
Imprédictibilité	Prévisions de comportement humain ne se prêtant pas facilement à une prédiction (p. ex. : évolution de systèmes politiques). Composants chaotiques de systèmes complexes.	Utilisation de scénarios portant sur une étendue plausible, posant clairement des hypothèses, des limites envisagées et des jugements subjectifs. Utilisation d'ensembles de simulations.
Incertitude structurelle	Modèles inadéquats, cadres conceptuels incomplets ou concurrents, manque d'accord sur une modélisation, ambiguïté de définitions ou de limites de système, processus significatifs ou relations erronément spécifiés ou non considérés.	Spécifier clairement les hypothèses et les définitions de système, comparer les modèles avec des observations dans un périmètre de conditions, évaluer la maturité des bases scientifiques et à quel point la compréhension repose sur des concepts fondamentaux testés dans d'autres domaines.
Incertitude sur valeur	Données manquantes, inexactes ou non-représentatives ; résolution spatiale ou temporelle inappropriée ; paramètres de modèle peu connus ou changeants.	Analyse des propriétés statistiques d'ensembles de valeurs (observations, résultats d'ensemble modèle, etc.) ; tests statistiques de ré-échantillonnage par <i>bootstrap</i> et hiérarchiques ; comparaison de modèles avec observations.

Tableau 2.1: Typologies de l'incertitude (*A simple typology of uncertainties*) selon l'IPCC. (Traduction).

Pour transmettre l'information d'incertitude de manière non équivoque et consistante, [IPCC, 2005] et [Mastrandrea *et al.*, 2010] 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 linguistiques : l'échelle terminologique qui en résulte (voir tableau 2.2) indique aux auteurs du rapport quelle expression exacte employer pour exprimer la quantité de la mesure d'incertitude d'un résultat scientifique. [Mastrandrea *et al.*, 2010] précisent que

quand l'information présente est suffisante, il est préférable de préciser entièrement la distribution de probabilité ou un intervalle de probabilité (p. ex. : 90-95 %) sans utiliser les termes¹².

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

Tableau 2.2: Échelle de probabilité (*Likelihood Scale*) de l'IPCC, recommandée pour l'usage d'un langage calibré dans la rédaction de son cinquième rapport (AR5). Les niveaux grisés sont ceux dont les termes sont signalés comme moins utilisés lors de la rédaction du rapport précédent (AR4).

Lorsqu'il s'agit d'exprimer la qualité de confiance en la validité d'un résultat scientifique, l'IPCC demande aux auteurs d'employer les expressions qui correspondent à la synthèse des dimensions d'une double échelle (voir tableau 2.3, p. 15). Le croisement des valeurs (termes) des deux dimensions est synthétisé ensuite en un niveau de confiance exprimable par le terme 'confidence' accompagné du qualificatif d'un des cinq niveaux possibles : 'very low', 'low', 'medium', 'high' et 'very high' (voir tableau 2.4, p. 16).

Termes	Dimensions d'évaluation de la validité
'limited', 'medium', 'robust'	'type', 'amount', 'quality', 'consistency of evidence'
'low', 'medium', 'high'	'degree of agreement'

Tableau 2.3: Double échelle de confiance de l'IPCC (*confidence scale*) : dimensions à croiser pour évaluer la validité d'un résultat.

¹²Traduction de « *When there is sufficient information, it is preferable to specify the full probability distribution or a probability range (e.g., 90-95 %) without using the terms* ». [Mastrandrea et al., 2010]

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

High agreement Limited evidence	High agreement Medium evidence	High agreement Robust evidence
Medium agreement Limited evidence	Medium agreement Medium evidence	Medium agreement Robust evidence
Low agreement Limited evidence	Low agreement Medium evidence	Low agreement Robust evidence

Tableau 2.4: Double échelle de confiance de l'IPCC (*confidence scale*) : croisement des dimensions de la confiance dans la validité d'un résultat. Tableau de synthèse repris de [Mastrandrea *et al.*, 2010]

Selon ces critères terminologiques, l'incertitude d'un résultat scientifique probable à 91 %, avec une confiance 'medium evidence' et 'low agreement', pourra donc être exprimée par « *very likely with a low confidence* ». Des précautions à prendre en compte dans l'usage de ce langage calibré sont indiquées : par exemple, [Mastrandrea *et al.*, 2010] précisent « transmettez l'incertitude avec soin, en utilisant un langage calibré pour les résultats principaux, et fournissez des indications contrôlables sur l'origine de vos évaluations des preuves et de la confiance décrites dans votre chapitre »¹³. D'autres points délicats sur le langage en général sont abordés, par exemple : « *a 10% chance of dying is interpreted more negatively than a 90 % chance of surviving* ». Les auteurs de l'IPCC continuent à suivre ces recommandations, reprises dans [Mastrandrea *et al.*, 2011] : ils le montrent par exemple dans le rapport spécial sur le réchauffement climatique (*Global Warming*), [IPCC, 2018, p. 91], publié en 2019.

Des institutions internationales œuvrant sur d'autres sujets préconisent aussi ce type d'échelles de langage calibré pour communiquer l'incertitude scientifique dans l'aide à la décision. Par exemple, le Comité scientifique de l'Autorité européenne de sécurité des aliments (AESA)¹⁴ propose aussi les options « *inconclusive* » et « *unknown* » pour une probabilité subjective de 0-100 % [Benford *et al.*, 2018, p. 23-24].

¹³Traduction de « *communicate uncertainty carefully, using calibrated language for key findings, and provide traceable accounts describing your evaluations of evidence and agreement in your chapter* ».

¹⁴European Food Safety Authority - EFSA.

2.3.2 La gestion et diffusion du renseignement

Des organisations de recherche liées à des services d'état de renseignement étudient comment prendre en compte et transmettre l'incertitude dans la chaîne de traitement des informations de renseignement. Par exemple, la RAND¹⁵ Corporation mentionne, dans son rapport [Brown et Shuford, 1973], une Charte des valeurs numériques des degrés d'incertitude (*Chart of numerical values of levels of uncertainty*) de Sherman Kent (voir tableau 2.5) qui ressemble à l'une des pratiques adoptées dans le domaine de la communication scientifique abordé précédemment. Dans son rapport [Davis et al., 2016, p. xi, xv-xvi et xx], la RAND Corporation fait aussi état de particularités de l'incertitude, au sens de celle traitée pour la fusion d'informations dans le domaine du renseignement, considérée en liaison avec d'autres catégories d'informations qualitatives, subjectives, floues, ambiguës, contradictoires et trompeuses. Par exemple : l'incertitude due à des limites de la connaissance se distingue de l'incertitude due à des processus aléatoires, dont un attribut change peut-être de manière aléatoire ; et l'évaluation de la crédibilité d'un rapport (information) peut elle-même être incertaine.

Probability (percent)	Verbal Equivalent
100	It is certain that ...
85-99	It is almost certain that ...
60-84	It is probable that ...
40-59	The chances are about even that ...
15-39	It is probable that ... not ...
1-14	It is almost certain that ... not ...
0	It is impossible that ...

Tableau 2.5: Charte des valeurs numériques des degrés d'incertitude (*Chart of numerical values of levels of uncertainty*) de Sherman Kent, pour faire correspondre des phrases à des intervalles de probabilité. [Brown et Shuford, 1973] cité par [Auger et Roy, 2008, p. 1865-1866].

Une autre pratique, liée à la réduction de l'incertitude, est l'utilisation de l'intelligence de collectifs humains dite *crowdsourcing*, notamment à partir de sources publiques ou scientifiques disponibles en ligne. C'est notamment ce que fait l'Intelligence Advanced Research Projects Activity (IARPA) dans plusieurs

¹⁵Acronyme de *Research AND Development*.

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

projets¹⁶. Par exemple, la IARPA annonçait en 2017 son lancement du logiciel *Crowdsourcing Evidence, Argumentation, Thinking and Evaluation - CREATE*¹⁷ pouvant réduire l'incertitude d'informations élaborées par des agences de renseignement, et pouvant aider les analystes à comprendre et à communiquer, entre autres, des degrés d'incertitude.

La prise en compte de l'incertitude dans la chaîne de recueil et traitement des renseignements est normalisée par des organisations militaires. À ce propos, [Dragos, 2013, partie VI] souligne l'importance de l'aspect composite des produits finaux de l'information :

« L'incertitude [...] devient un enjeu majeur dès lors qu'elle a un impact sur des tâches d'aide à la décision. Dans le domaine militaire, les données digitalisées sont une ressource très utile pour atteindre différents objectifs de fusion d'information de haut niveau »¹⁸.

Le [CoFAT, 2001, chap. 1 partie 24 et chap. 2 partie 21] précise qu'un renseignement est recueilli et structuré afin de répondre à des questions organisées pour être utilisables : faits ou indices (qui, quoi, où, quand, combien), conditions de recueil (par qui, auprès de qui, quand, comment), et commentaire séparé de l'information elle-même, pour mieux la comprendre où l'exploiter. Il signale notamment que le résultat d'évaluation d'un renseignement donne lieu à une *cotation* via une double échelle terminologique, par :

- Un coefficient de qualité de sa source, lequel caractérise le moyen humain ou matériel qui transmet un fait (acuité, spécialisation, expérience, etc.), sur six degrés : 'complètement sûre', 'habituellement sûre', 'assez sûre', 'habituellement pas sûre', 'pas sûre' et 'sûreté ne pouvant être appréciée'.
- Un coefficient de probabilité d'exactitude de l'information recueillie, lequel caractérise la précision et les conditions de recueil et transmission (lieu, heure, conditions particulières, etc.), sur six degrés : 'confirmé par d'autres

¹⁶Forecasting Science & Technology (ForeST), Foresight and Understanding from Scientific Exposition (FUSE), Open Source Indicators (OSI), Aggregative Contingent Estimation (ACE), Hybrid Forecasting Competition (HFC), ...

¹⁷Source : ODNI (08/02/2017), « IARPA Launches "CREATE" Program To Improve Reasoning Through Crowdsourcing ». Communiqué n° 4-17 de l'Office of the Director of National Intelligence, Washington, EUA. Consulté le 19/07/2019 sur : <https://www.dni.gov>

¹⁸Traduction de « Uncertainty is [...] becoming a crucial issue when impacting decision support tasks. In the military field, soft data are a valuable resource to achieve different goals of high-level information fusion ».

sources', 'exactitude probable', 'exactitude possible', 'exactitude douteuse', 'peu probable' et 'exactitude ne pouvant être appréciée'.

L'Accord de normalisation 2511 de l'OTAN, [OTAN, 2003], précise que le degré de confiance dans l'information de chaque rapport est noté par une évaluation combinée à partir de deux échelles indépendantes (leurs degrés sont identiques aux deux échelles françaises précédentes) : une échelle de *fiabilité de la source* dont les degrés sont notés de *A* à *F*, et une échelle de *crédibilité de la source* dont les degrés sont notés de *1* à *6*. Cela permet de noter chaque source du document indépendamment avec un code à deux symboles, par exemple 'C2' avec une seule source à la fois 'Fairly reliable' et 'Probably true' (ou en séparant par des virgules, par exemple 'B1,C3,A2', pour plusieurs sources).

[Capet et Revault d'Allonnes, 2013, parties 4.2.1, 4.4.1.1 et 4.4.4] signalent que la cotation des renseignements par cette double échelle française similaire à celle de l'OTAN est critiquable sur divers aspects, dont entre autres : déterminer sur quoi porte cette cotation renvoie à la problématique de granularité de l'information (donnée brute ou renseignement enrichi) pour les opérations de regroupement et de fusion qui suivent la cotation ; et le degré de confiance dans la source est supposé immuable, ce qui ne prend pas en compte le fait que les compétences de cette source varient en fonction des thèmes abordés.

2.3.3 La communication scientifique auprès du public général

Dans le but de faciliter l'accès rapide et efficace à des informations environnementales fiables, notamment aux scientifiques, décideurs et grand public qui doivent faire face aux défis actuels résultant du changement climatique, [Giuliani, 2016, p. 1-3 et 21] fournit des recommandations « afin de suggérer quelques directions de recherche ». Sa recommandation n° 2 *gérer et communiquer à propos de la qualité des données* (« important challenge qui n'est actuellement pas encore résolu ») propose que l'évaluation quantitative et qualitative de la qualité des données prenne en compte l'« incertitude associée aux données et/ou modèles », en gérant et communiquant mieux sur la qualité des données. Cette recommandation évoque nommément le développement en cours d'un label sur les données et services environnementaux : le projet de ce label en particulier semble maintenant à l'arrêt¹⁹, mais, concernant le traitement de l'incertitude, il est utile de faire un parallèle entre le traitement des données du changement climatique et le traitement

¹⁹Le site de ce label, dernièrement consulté le 21/02/2022 sur <http://www.geolabel.info/home.htm> n'est manifestement plus maintenu, et nous n'avons pas eu accès à des cas

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

des données de la santé. Dans ce sens, les éléments consultables²⁰ sur la gestion et les dimensions proposées pour ce label de qualité des données environnementales sont suffisants pour remarquer de fortes similitudes avec des labels de la qualité des données de la santé, dont l'usage effectif est au contraire documenté et soutenu, ce qui fait de ces derniers des exemples pratiques et réalistes :

En juin 2004, le Conseil de l'Union européenne²¹ a confirmé la nécessité d'élaborer des standards de qualité et d'indiquer une direction claire aux états membres en ce qui concerne l'information de santé sur internet :

« pour décider raisonnablement des informations en matière de santé auxquelles il fera confiance ou des produits ou services qu'il utilisera, tout visiteur d'un site doit connaître les normes que ce dernier applique pour concevoir son contenu ».

Dans le but de participer à la gestion de l'incertitude des informations de santé dirigées vers le grand public sur internet, la certification de sites web de la santé est un service proposé au niveau mondial par plusieurs organisations, [Haute Autorité de Santé, 2007], dont la *Fondation Health On the Net (HON)* avec son système des *principes* de certification dit *HONcode*. Pour les spécialistes des sciences de l'information [Accart et Rivier, 2012, p. 133], « L'entreprise qui paraît avoir été le plus loin dans la crédibilisation des ressources est *Health On the Net* » et « Cette labellisation explicite, établie sur des critères rigoureux [...] est à l'information ce que sont les normes Iso 9000 pour la gouvernance des organisations ».

HON, [Rey, 2011, p. 11-15], a pour but principal d'aider à trouver de l'information médicale de qualité sur internet. C'est une association à but non lucratif basée à Genève (Suisse), reconnue d'utilité publique et liée par sa trajectoire aux *Hôpitaux Universitaires de Genève (HUG)*. C'est une organisation internationale par ses membres et son personnel, et par son travail : elle est reconnue comme ONG avec le statut consultatif spécial²² auprès du Conseil économique et social des

d'utilisation avérés du label dont les informations les plus récentes que nous avons pu consulter datent de 2015.

²⁰Page <http://www.geolabel.info/facets.htm>, consultée dernièrement le 21/02/2022.

²¹Communiqué de presse de la 2586^e session du Conseil de l'Union Européenne, du 2 juin 2004, point 14. Disponible sur : http://europa.eu/rapid/press-release_PRES-04-163_fr.htm. (Consulté le 10/08/2018).

²²Depuis 2002. *Statut consultatif spécial auprès de l'ECOSOC*. Disponible sur

Nations Unies (ECOSOC) ; l'Union européenne et la Commission Européenne lui ont décerné le prix Européen « *eHealth award* » ; et elle a le statut de liaison *Catégorie A* de la commission technique informatique de la santé (ISO / TC 215) pour contribuer à l'élaboration de normes, développées en particulier dans le domaine de la structure de données, et dans celui du contenu sémantique et de la sécurité.

Le HONcode a été créé pour être à la fois un standard éthique et de qualité. En obligeant les sites fournisseurs d'information de la santé à informer sur leurs informations (*sic*), la certification facilite la vérification de la qualité par les autorités, et par le public lui-même quand il cherche de l'information certaine sur la santé.

C'est l'expertise humaine qui est la base du travail de vérification et revérification périodique du respect de standards et de bonnes pratiques du HONcode concernant l'incertitude de l'information de milliers de sites web de la santé, en diverses langues. Ces *réviseurs* de niveau universitaire sont formés aussi en interne ; et c'est notamment sur des critères de formation interne que HON est accréditée selon la norme ISO 65. Bien que cette certification HONcode ne soit pas automatisée, certaines fonctions automatiques de vérification sur des points précis sont effectuées par des ordinateurs.

L'intérêt et le succès mondial de cette certification HONcode de sites web d'information médicale, dans de nombreuses langues, est basé sur huit principes de levée d'ambiguïté sur la qualité de l'information (voir tableau 2.6). Les webmasters des sites candidats doivent respecter ces principes pour obtenir et conserver un certificat personnalisé, révisé périodiquement, dont les détails personnalisés sont consultables sur le site de HON²³, et pour obtenir l'autorisation d'apposer sur leur propre site un sceau de certification cliquable.

Les autorités publiques ou collégiales en charge de la lutte contre le réchauffement et les dérèglements climatiques en viendront-elles à prescrire une telle labélisation intégrant l'incertitude pour certaines données environnementales ; et quel rôle pourra y jouer la classification automatisable des incertitudes ? La question est fondée, puisque la certification des sites web d'information de santé en France, via le système de HON, a été prescrite en 2007 par la Haute Autorité de Santé (HAS)²⁴ ; et le Conseil National de l'Ordre des Médecins a conseillé offi-

<http://esango.un.org/civilsociety/displayAdvancedSearch.do?method=search&sessionCheck=false>. (consulté le 10/08/2018).

²³Site www.healthonnet.org.

²⁴La HAS a opté pour la certification HONcode, [Nabarette et Caniard, 2010, p. 52-55], pour

2.3. PRISE EN COMPTE DE L'INCERTITUDE PAR DES ORGANISATIONS

ciellement aux praticiens de demander la certification de leur site web auprès de HON²⁵.

<i>Autorité</i>	Le site indique la qualification des rédacteurs.
<i>Complémentarité</i>	Le site complète et ne remplace pas la relation patient-médecin ; il présente sa mission et son audience.
<i>Confidentialité</i>	Le site préserve la confidentialité des informations personnelles soumises par ses visiteurs.
<i>Attribution</i>	Le site cite les sources des informations publiées et date les pages de santé.
<i>Justification</i>	Le site justifie toute affirmation sur les bienfaits ou les inconvénients de produits ou traitements.
<i>Professionnalisme</i>	Le site rend l'information la plus accessible possible, identifie le webmestre, et fournit une adresse de contact.
<i>Transparence du financement</i>	Le site présente ses sources de financements.
<i>Honnêteté publicité / édition</i>	Le site sépare les politiques publicitaire et éditoriale. La publicité est identifiée.

Tableau 2.6: Les huit principes de confiance du HONcode (version ici simplifiée).

établir une procédure de certification des sites informatiques dédiés à la santé suivant l'article L. 161-38 de la loi n° 2004-810 du 13 août 2004.

²⁵Conseil National de l'Ordre des Médecins (CNOM) (2010), « Charte de conformité ordinale applicable aux sites web des médecins », p. 2. CNOM, Paris.

Chapitre 3

Le concept d'incertitude en science

3.1 Définitions de l'incertitude

Le philosophe [Kant, 1781, traduction p. 635-636] liait implicitement l'incertitude à la subjectivité :

« La croyance, ou la valeur subjective du jugement, par rapport à la conviction (qui a en même temps une valeur objective), présente les trois degrés suivants : l'*opinion*, la *foi* et la *science*. L'*opinion* est une croyance qui a conscience d'être insuffisante *aussi bien* subjectivement *qu'*objectivement. Si la croyance n'est que subjectivement suffisante et si elle est en même temps tenue pour objectivement insuffisante, elle s'appelle *foi*. Enfin, la croyance suffisante aussi bien subjectivement qu'objectivement s'appelle *science*. La suffisance subjective s'appelle *conviction* (pour moi-même) et la suffisance objective, *certitude* (pour tout le monde). Je ne m'arrêterai pas à éclaircir des concepts si clairs ».

La recherche scientifique vise une objectivité dans le raisonnement et les résultats obtenus. Elle se base, entre autres, sur la validation des produits de la recherche par les pairs. Dans les articles scientifiques, les modalités épistémiques permettent aux chercheurs d'exprimer une forme de subjectivité qui fait partie intégrante des savoirs : l'*incertitude sur les faits*. Il est donc nécessaire de tenir compte de cette subjectivité des connaissances dans les échanges entre scientifiques afin de construire un dialogue productif cherchant à dégager une connaissance objective.

3.1. DÉFINITIONS DE L'INCERTITUDE

L'incertitude sur les faits joue un rôle important en science, mais aussi au-delà de la sphère scientifique, par exemple parce qu'elle est associée au risque dont la gestion, selon [Motet, 2010, p. 33-36], « n'a pas pour but d'éliminer l'incertitude. Elle est propre à nos activités et est sans doute même indispensable. Reconnaître l'existence de l'incertitude comme attribut inhérent à nos actions, à nos décisions, etc. est essentiel ». La norme ISO 31000:2009(fr) « Management du risque » définit le concept de risque comme l'« effet de l'incertitude sur l'atteinte des objectifs ». [Motet, 2010] signale aussi que, dans la culture de l'ingénierie des risques, une démarche scientifique doit intégrer « la prise en compte du degré de confiance reconnu aux données ou son évaluation sur les résultats tirés de l'exploitation desdites données ». Cette considération sur le degré de confiance dans les données se retrouve dans des pratiques de prise en compte de l'incertitude par des organisations (voir § 2.3 page 10).

3.1.1 Définition générale de l'incertitude

D'un point de vue étymologique, le terme *incertitude* apparaît dans la langue française au XV^e siècle comme antonyme et dérivé de *certitude* apparut au XIV^e siècle [Académie française, 1935], lui-même emprunté au bas latin *certitudo* (« caractère de ce qui est certain, vrai, certitude morale, conviction »¹) dérivé de *certus* [Dauzat et al., 1964] (c'est-à-dire *certain*).

Les dictionnaires et encyclopédies généralistes proposent des définitions du terme *incertitude* qui sont relativement courtes, par exemple : « connaissance sûre et claire de quelque chose »².

Une recherche du sens d'*incertitude* peut s'étoffer diachroniquement avec une palette de définitions, par exemple³ : « caractère de ce qui n'est pas certain, assuré » (1531) ; « état d'une personne indécise, irrésolue » (1538) ; « caractère de ce qui est imprévisible » (1559) ; « état d'une personne qui ne sait pas ce qu'elle doit croire » (1580) ; « chose incertaine, mal connue, qui prête au doute » (1636).

Le principe d'incertitude, énoncé par le physicien Heisenberg (1927) exprime la propriété physique quantique selon laquelle la certitude dans la mesure de la position d'un électron est proportionnellement liée à l'incertitude dans la mesure de sa vitesse, et inversement ; à tel point que lorsque la position devient parfaite-

¹CNRTL, consulté le 24/01/2018.

²Traduction de : Enciclopedia Espasa, 1995. Espasa Calpe, Madrid, Espagne. ISBN : 84-239-6171-0.

³Source : CNRTL, consulté le 24/01/2018.

ment connue, alors la connaissance de la vitesse devient complètement inconnue, [Dougherty, 2016, p. 69-70]. À l'époque actuelle, de nouveaux champs de recherche sont entrain de s'approprier de nouveaux concepts de l'incertitude. Par exemple, l'enseignement de l'intelligence artificielle (IA) [Higuera (de la), 2019, p. 5] doit prendre en compte le caractère imparfait ou aléatoire des données (*Big Data*), et l'indéterminisme des modèles. De plus, la prise en compte de l'incertitude implique de considérer les erreurs et imperfections de celui qui transmet le savoir.

La diversité d'emplois du concept de l'incertitude est étendue transversalement en science, par exemple (voir aussi des exemples de définitions en annexes A.2, p. 213 ; A.3, p. 215; et A.4, p. 217) : le recours au terme *incertitude* est proposé par l'économiste [Knight, 1921, § I.I.26 et I.I.27] pour désigner ce qui n'est pas quantifiable et pas mesurable⁴ ; même si des théories supposent que la nature de l'incertitude peut être influencée par les agents économiques agissant sur sa source, son degré et ses effets, [Viviani, 1994, p. 119-127]. Le « principe d'incertitude biologique » est proposé par le sociologue [Morin, 1980, p. 277-281 et 379-380] pour caractériser la situation de l'individu à la fois unique et continu dans l'espèce et l'univers. Décrite par les physiciens [Dossantos-Uzarralde *et al.*, 2017, p. 84-85], « l'incertitude épistémique » est liée à une connaissance insuffisante pour la modélisation suffisamment fiable et la qualité des prédictions, de paramètres quantitatifs d'un phénomène étudié, et réductible par acquisition d'informations issues d'expériences dédiées ou par amélioration du modèle. Le tableau 3.1 (p. 26) résume une palette de prises en compte du concept d'incertitude dans des champs scientifiques différenciés comme le sont les sciences économiques, les sciences physiques, les sciences biomédicales, et les sciences humaines et sociales.

3.1.2 Sources de l'incertitude en science

L'incertitude est partie constituante de la démarche scientifique : 1) dans la réflexion des scientifiques tournée vers l'inconnu ou l'indéfini ; 2) à travers la méthode scientifique dont les résultats doivent être communiqués de manière à être transparents, reproductibles, réfutables ; et 3) éventuellement dans les usages extra scientifiques qui sont faits des productions scientifiques, notamment politiques,

⁴P. ex. : « We shall accordingly restrict the term "uncertainty" to cases of the non-quantitative type. [...] The crucial character of the distinction between measurable risk and unmeasurable uncertainty will become apparent in this discussion ».

3.1. DÉFINITIONS DE L'INCERTITUDE

Champs	Nom de notion	Approximation de la définition
Economie	'Incertitude'.	Ce qui est non-quantifiable ou non-mesurable , par opposition au 'risque' qui est mesurable. [Knight, 1921, § I.I.26].
Physique	'Incertitude de mesure' accompagnant la valeur numérique d'un résultat de mesure.	La marge d'erreur des instruments de mesure est elle-même mesurée. L'incertitude connue sert à qualifier la qualité de la mesure. Les composantes individuelles de l'incertitude sont combinées en une seule incertitude globale. Permetts que ceux qui utiliseront le résultat puissent estimer sa fiabilité. [JCGM, 2008] et [JCGM, 2012].
Biomédical	'Niveau de preuve scientifique'.	Evaluation 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, etc. [Haute Autorité de Santé, 2013].
SHS	'imprécision', 'indétermination', 'incomplétude', 'ambiguïté' et 'imprévisibilité'.	L'incertitude est liée à la complexité des objets d'étude sociaux et humains, à l'influence du contexte et des 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 <i>et al.</i> , 2015].

Tableau 3.1: Diversité de définitions de l'incertitude : exemples de selon des champs scientifiques larges et différenciés.

économiques, ingénieriaux, industriels et commerciaux, pédagogiques et journalistiques.

Le propos n'est pas ici d'aborder exhaustivement le phénomène de l'incertitude, ni de le rechercher dans la totalité des sciences, mais d'en retenir un ensemble de caractéristiques suffisamment englobant, à travers un éventail de champs scientifiques, pour faire avancer notre réflexion sur les manifestations et prises en compte l'incertitude dans les sciences en général à travers leurs productions textuelles.

3.1.2.a L'incertitude dans les sciences exactes

L'incertitude scientifique fait l'actualité scientifique et l'actualité en général.

À un extrême, 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 ou laisser se développer des catastrophes, notamment industrielles, sanitaires et écosystémiques, jusqu'à l'échelle planétaire comme les pollutions, l'acidification océanique, l'extinction de masse des espèces et le réchauffement-dérèglement climatique.

À un autre extrême, la physique apporte un exemple de l'utilité universelle de l'incertitude et de ses conséquences, 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), est 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 à travers le monde). Des ajustements fondamentaux peuvent donc se produire grâce à la correcte transmission et intégration de l'information sur l'incertitude scientifique.

L'incertitude en physique La *science physique* (ou les sciences physiques : physique classique, chimie, astronomie, ...) est celle qui étudie (par l'expérimentation, par l'élaboration de concepts et par la modélisation de phénomènes na-

3.1. DÉFINITIONS DE L'INCERTITUDE

turels, ...) les propriétés fondamentales (grandeurs, constituants fondamentaux, interactions, transformations, ...) des objets réels constituants de l'Univers (matière, ondes, énergie, espace, temps, ...). La physique est une science dite exacte qui utilise ses propres concepts d'*incertitude* pour ses mesures. Il s'agit d'incertitude sur des grandeurs identifiées : ces grandeurs doivent être mesurées de manière fiable pour valider des phénomènes physiques, tels que des poids ou des vitesses. Les mesures peuvent être faites par des instruments techniques dont la marge d'erreur est elle-même mesurée. L'*incertitude* sert ici à qualifier la qualité de la mesure.

LE GUM (*Guide pour l'expression de l'incertitude de mesure*) [JCGM, 2008, p. v et vi] « établit les règles générales pour l'évaluation et l'expression de l'incertitude de mesure » « pour exprimer l'incertitude de mesure et pour combiner les composantes individuelles de l'incertitude en une seule incertitude globale » ; et pour que ses recommandations soient utilisables en normalisation, dans l'étalonnage, dans l'accréditation des laboratoires et dans les services de métrologie.

Le GUM [JCGM, 2008, p. vii] avertit que : « Lorsqu'on rend compte du résultat d'un mesurage d'une grandeur physique, il faut obligatoirement donner une indication quantitative sur la qualité du résultat pour que ceux qui l'utiliseront puissent estimer sa fiabilité ». C'est-à-dire qu'il faut une procédure pour évaluer et exprimer l'incertitude d'une mesure. Il ajoute que « un consensus universel sur l'évaluation et l'expression de l'incertitude de mesure permettrait la compréhension aisée et l'interprétation correcte d'un vaste spectre de résultats de mesure en science, ingénierie, commerce, industrie et réglementation ».

Le GUM [JCGM, 2008, p. viii] précise que toute description détaillée de l'incertitude devrait comprendre ses composantes et indiquer pour chacune la méthode utilisée (statistique ou autre, voir ci-après) pour lui attribuer une valeur numérique ; et [JCGM, 2008, p. vii] la méthode idéale d'évaluation et d'expression de l'incertitude devrait être :

- *universelle*, pour s'appliquer à tout mesurage et tout type de donnée d'entrée ;

et la grandeur qui l'exprime devrait être :

- *logique* (« *internally consistent* »), pour être déductible de ses composantes,
- et *transférable*, pour être réutilisable dans l'évaluation de l'incertitude d'une autre mesure.

Selon le VIM (Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés) [JCGM, 2012, p. 26], l'évaluation d'une composante de l'incertitude, par d'autres moyens qu'une analyse statistique des valeurs mesurées, peut être fondée sur des informations :

- associées à des valeurs publiées faisant autorité, ou à la valeur d'un matériau de référence certifié,
- obtenues à partir d'un certificat d'étalonnage, ou à partir de la classe d'exactitude d'un instrument de mesure vérifié,
- concernant la dérive,
- ou obtenues à partir de limites déduites de l'expérience personnelle.

Le GUM [JCGM, 2008, p. vi] souligne l'importance pour les laboratoires de métrologie nationaux de « combiner les composantes individuelles de l'incertitude en une seule incertitude globale », même si [JCGM, 2008, p. 73] « Dans la plupart des situations pratiques de mesure, le calcul d'un intervalle correspondant à un niveau de confiance spécifié, au mieux, ne peut qu'être approximatif comme, en fait, l'estimation de la plupart des composantes individuelles de l'incertitude dans ces situations. Même si l'on obtient un écart-type expérimental de la moyenne à partir d'un nombre d'observations répétées aussi élevé que 30 pour une grandeur décrite par une loi normale, cet écart-type a lui-même une incertitude d'environ 13 pour-cent ».

[Robert-Schwartz et Treiner, 2002] décrivent l'incertitude des mesures en physique en signalant l'opposition *valeur exacte* vs *valeur approchée* : « On ne peut en général pas parler de *valeur exacte* de la grandeur à mesurer, sauf dans certains cas, par exemple s'il s'agit d'une constante mathématique ». Ils citent [Cessac et Treherne, 1967] : « Le nombre a , résultant de la mesure d'une grandeur A , n'est qu'une valeur approchée de A ». Lors de la mesure d'une grandeur, la *valeur approchée* est la norme (p. ex. : une taille ou une température) et la *valeur exacte* est donc l'exception.

Les causes de l'approximation des valeurs sont, d'une part, des erreurs systématiques produites par les méthodes ou instruments imparfaits, et, d'autre part, des erreurs accidentelles causées par les imperfections de l'opérateur. Concernant les instruments, ceux-ci se caractérisent par un temps de réponse, une sensibilité et une exactitude [JCGM, 2008]. L'exactitude est elle-même dépendante de la justesse et de la fidélité ou la reproductibilité des indications de l'appareil.

3.1. DÉFINITIONS DE L'INCERTITUDE

La détermination des influences au cours du mesurage est importante pour définir l'incertitude des mesures puisque « une connaissance incomplète des grandeurs d'influence et de leurs effets peut souvent contribuer d'une manière significative à l'incertitude sur le résultat d'un mesurage » [JCGM, 2008, p. 53]. Les valeurs et incertitudes des grandeurs déterminées au cours du mesurage dépendent de facteurs multiples. Elles peuvent être obtenues :

« par exemple, à partir d'une observation unique, ou à partir d'observations répétées, ou par un jugement fondé sur l'expérience. Elles peuvent impliquer la détermination de corrections pour les lectures d'instruments et de corrections dues aux grandeurs d'influence telles que la température ambiante, la pression atmosphérique ou l'humidité ». [JCGM, 2008, p. 9]

La recherche de la valeur numérique comporte les particularités suivantes :

- La marge d'erreur des instruments de mesure est elle-même mesurée.
- L'incertitude connue sert à qualifier la qualité de la mesure.
- Il faut « combiner les composantes individuelles de l'incertitude en une seule incertitude globale », [JCGM, 2008, p. vi],
- pour « donner une indication quantitative sur la qualité du résultat pour que ceux qui l'utiliseront puissent estimer sa fiabilité », [JCGM, 2008, p. vii].

Les concepts, méthodes, outils et résultats de la physique sont utiles, voire essentiels, pour d'autres disciplines scientifiques et pour l'évolution technologique. La transmission et l'interprétation correcte des résultats de mesures physiques conditionne les sciences et applications dans les domaines les plus divers, de l'informatique à la médecine en passant par la météorologie spatiale.

3.1.2.b L'incertitude dans les sciences de l'environnement

Avec pour but que les experts scientifiques puissent mieux échanger et rendre compte de la qualité de la connaissance dans des processus d'évaluation environnementale (voir aussi des exemples de définitions en annexe A.4, p. 217), [Sluijs (van der) *et al.*, 2006a, p. 2 et 7-10] proposent une typologie de l'incertitude ramifiée sur six dimensions reprises et synthétisées ci-dessous :

- la *localisation* de l'incertitude. Cette dimension indique où l'incertitude se manifeste dans la problématique traitée, à travers les catégories :
 - *contexte* : cadre déterminant la problématique, ses limites et son environnement (représentation écologique, technologique, économique, sociale et politique).
 - *données* : mesures et calibration, et données d'observation et d'enquête utilisées.
 - *modèle* : modèles allant des concepts mentaux jusqu'aux modèles mathématiques, étant informatisés ou non, et pouvant inclure des structures, des techniques, des paramètres et des entrées de données.
 - *jugement expert* : contributions à l'évaluation (récits, scénarios, orientations et recommandations) issues de la réflexion, qui dépassent le contexte, le modèle et les données.
 - *résultats* : conséquences, indicateurs, propositions ou affirmations qui sont d'intérêt pour la problématique.
- le *niveau* de l'incertitude. Cette dimension indique si ce qui caractérise l'incertitude relève plutôt d'une des classes suivantes :
 - *incertitude statistique* : incertitude associée à des probabilités ou des intervalles.
 - *incertitude de scénario* : incertitude associée à une palette d'événements ou résultats alternatifs envisagés de manière non probabiliste.
 - *ignorance reconnue* : incertitude associée à ce qui est connu mais ne peut être évalué.
 - *ignorance inconnue (unknown unknowns)* : incertitude associée à ce qui ne peut être ni connu ni évalué, et dont l'incidence sur la problématique reste par définition une extrémité théorique qui ne peut être prise en compte au moment de l'évaluation.
- la *nature* de l'incertitude. Cette dimension indique ce qui est à l'origine de l'incertitude :
 - *incertitude issue de la connaissance imparfaite* : incertitude associée à des incomplétudes de la connaissance ou à des failles épistémiques.

3.1. DÉFINITIONS DE L'INCERTITUDE

- *incertitude issue de la variabilité* : incertitude associée à une indétermination ou variabilité.
- la *qualification du socle de connaissances*. Cette dimension indique les forces et les faiblesses de l'évaluation (empiriques, théoriques, méthodologiques ...) à l'aide des classes de soutien de fiabilité suivantes :
 - *faible* : ce qui est affirmé repose sur un nombre significatif de connaissances incertaines, et doit être reconsidéré avec des suggestions de réévaluation.
 - *acceptable*.
 - *fort*.
- l'*influence des valeurs sur les choix*. Cette dimension indique des biais de l'arbitraire qui peuvent avoir une incidence préoccupante sur l'évaluation qu'ils façonnent, parmi lesquels sont suggérés :
 - *le problème est cadré* : la variété des points de vue possibles sur la problématique n'est peut-être pas objectivement prise en compte dans la perspective adoptée.
 - *la connaissance et information est sélectionnée et appliquée* : le choix des données ou des modèles utilisés est peut-être partisan.
 - *les explications et conclusions sont exprimées et formulées* : le choix de ce qui est exposé est peut-être tendancieux.

L'incertitude en hydrologie Selon [Bourgin, 2014, p. 9-11], en hydrologie « les incertitudes peuvent être caractérisées selon leur type et leurs sources », en particulier :

- deux types :
 - l'incertitude de type « réductible (ou épistémologique) » résultante « d'un manque de connaissance ».
 - L'incertitude de type « irréductible (ou structurelle, inhérente, stochastique) » résultante « d'une variabilité naturelle ».
- et quatre sources :

- l'incertitude liée aux données : provenant d'erreurs sur les mesures (ex. : des mesures de précipitations et débits) et d'erreurs du traitement des données.
- l'incertitude liée à l'estimation des paramètres : provenant de la difficulté à estimer et spécifier les (jeux de) paramètres à partir d'autres informations (ex. : des mesures d'un débit).
- l'incertitude liée au modèle : provenant des limites de capacités à représenter la dynamique d'un système (simplification de sa structure, résolution, numérisation) « même en l'absence d'erreur dans les données ».
- l'incertitude liée aux forçages⁵ futurs : provenant de l'imprévisibilité de processus significativement influents sur l'exactitude de prévisions.

L'incertitude en océanographie Selon [Brankart, 2014, p. 1 et 5, § I, p. 68, 93-96 et 109-110], en océanographie l'incertitude est inévitable, et elle doit être traitée en tant qu'élément essentiel et central à la fois des systèmes océaniques réels et de leur description : il faut la comprendre, la simuler et l'explorer. L'incertitude agit dans tous les composants et à toutes les échelles du système. Cependant, malgré l'avantage socio-économique des prévisions probabilistes en météorologie, l'incorporation optimale de l'incertitude dans les méthodes et théories n'est pas encore résolue :

- l'incertitude en observation de données : elle correspond à des erreurs de mesure. Elle résulte de l'instrument utilisé pour l'acquisition de données physiques ou biologiques. La relation entre les données et le système étudié, ainsi que la nature de l'incertitude de ces données, doivent être précisés.
- L'incertitude en modélisation : elle correspond à une incertitude de représentativité du modèle d'océan. Elle résulte « de ce que le modèle ne résout pas » (les prévisions sont probabilistes).
- L'incertitude en assimilation de données : les données à assimiler peuvent correspondre par exemple à l'altitude, la température, la salinité, etc. Leur

⁵Le forçage climatique ou météorologique est un processus externe, d'origine naturelle (ex. : solaire ou volcanique) ou anthropique (ex. : gaz à effet de serre), pouvant altérer significativement la dynamique du système considéré. Une des sources consultées le 27/12/2019 : <http://www.climat-en-questions.fr/reponse/fonctionnement-climat/forçage-climatique-par-david-saint-martin-olivier-boucher>.

3.1. DÉFINITIONS DE L'INCERTITUDE

assimilation correspond au développement (adéquation et pertinence scientifique) et à l'application (implémentation machine) pratiques d'un modèle opérationnel : elle est tributaire des équipes de chercheurs et des technologies, et elle résulte des algorithmes.

3.1.2.c L'incertitude dans les sciences humaines et Sociales

En Sciences Humaines et Sociales (SHS), [Fusco *et al.*, 2015, partie 2], se dévoile une incertitude « paradigme à part entière du processus de connaissance ». Ses concepts sont d'une diversité importante, par exemple : 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 par conséquent de « faire science avec l'incertitude » [Fusco *et al.*, 2015, partie 3] : l'évaluer, la mesurer, la réduire, et l'intégrer dans les processus, les résultats et la communication scientifique. Le lien entre l'intégration des incertitudes diverses et leur communication pour une prise en compte sujette à interprétations est par exemple évident dans l'étude scientifique de la gestion de crise, où [Dautun *et al.*, 2006, p. 3] indiquent que les « incertitudes sont [...] présentes à tous les niveaux », et que parmi les principales se trouvent : les causes de l'évènement, la source de danger, et les conséquences humaines, matérielles ou environnementales.

L'incertitude en ethnographie [Rinaudo, 2015] examine trois choix épistémologiques en lien avec l'incertitude, chacun ayant pour but de produire et interpréter des données dans la relation d'enquête ethnographique :

- « écarter toute possibilité d'incertitude » : la démarche implique une neutralité de l'enquêteur pendant les interactions de l'enquête ; et plus concrètement, celui-ci est censé mettre en œuvre une « capacité à neutraliser sa propre personnalité ». S'y ajoute la négation du fait que le sujet enquêté produit ses propres interprétations. Ce choix épistémologique a fait l'objet de critiques sévères et est considéré comme irréaliste.
- « tenter de [...] réduire ou de [...] neutraliser » l'incertitude : la démarche implique que l'incertitude peut être connue et doit être limitée grâce à une rigueur méthodologique, notamment par la critique du modèle et sa

souplesse, et par l'évaluation des résultats. S'y ajoute la reconnaissance que les enquêteurs et les enquêtés sont aussi des sujets et sont donc dotés d'un point de vue subjectif. Sont prises en considération l'influence de l'environnement, la documentation personnelle des enquêtés, la relation entre enquêteur et enquêté et sa dynamique, et les possibles difficultés de terrain de l'enquêteur. Les distorsions et effets produits par la relation sociale d'enquête sont contrôlés pour les maîtriser, par exemple : l'effet que les questions ont sur l'enquêté et ses réponses et sur l'analyse, pour choisir des questions appropriées à la fois à l'enquêté et à l'analyse ; ou encore l'effet de pouvoir social dissymétrique. Se dégager l'idée que bien prendre en compte la subjectivité conduit à une certaine objectivité.

- « tirer parti de l'incertitude pour produire de la connaissance » : la démarche implique que les effets délibérés et non délibérés de la subjectivité puissent produire des résultats désirables, par exemple un nouveau discours de la part de l'enquêté, et des résultats qui éclairent sur les données recueillies lors de l'enquête. Un entretien avec des résultats imprévus peut se révéler utile à la recherche, pertinent pour envisager de nouveaux questionnements, et mettre en relief des données différenciées. Se dégage l'idée d'une maturité méthodologique prônant de tirer parti des réalités établies et préparant à l'utilisation des possibles avantages de l'incertitude dans le respect de la déontologie.

L'incertitude en géographie En géographie, l'incertitude dépend de la question, de l'actualisation des données et du mode de représentation des résultats intermédiaires et finaux, ce qui ne sont pas forcément des particularités très distinguantes parmi de nombreuses autres disciplines scientifiques ; mais des auteurs s'attachent particulièrement à des aspects de la géographie pour étudier l'incertitude.

C'est par exemple le cas de l'analyse de [Funtowicz et Ravetz, 1990, p. 83-93] à propos des cartes. Ils en mettent en exergue les aspects qualitatifs imprécis ou vagues, mais globaux (*Gestalt*), avant de rappeler la « dialectique » de la communication des cartes avec des aspects quantitatifs simplificateurs ou conventionnels « inévitables », par exemple lors d'une réduction d'échelle ou lors de l'utilisation de bases de données. Ils font des cartes en général des témoignages de l'avancée de la science, et les comparent à des théories scientifiques pour le meilleur et

3.1. DÉFINITIONS DE L'INCERTITUDE

pour le pire : la « certitude apparente du savoir »⁶. Parmi les incertitudes que transmettent les cartes, ils notent entre autres celles dues à :

- la *granularité* : les limites imposées par l'échelle et par la densification de l'information ;
- le *design* : l'économie de détails, car ceux-ci ont un coût, et la spécialisation des fonctions de la carte pour des usages généralistes ou particuliers ;
- la *projection* : aire couverte, distorsions, centrage sur une région ;
- les *choix politiques* : frontières, choix de nommages et noms absents, embellissement des situations économiques ou d'équipements représentés ;
- la *mise à jour des données* : obsolescence générale ou partielle des détails des régions représentées, erreurs dans les données (rhétorique, symboliques, de précision par la mesure et par les choix de différents standards de calibration, quantitatives) ;
- la *qualité des logiciels* : différences courantes entre les qualités scientifique vs commerciale des éditeurs ;
- l'*ignorance* : des parties peuvent rester en blanc par manque d'information, et des parties peuvent être remplacées ou illustrées par l'imagination et par la superstition ou les dogmes ;
- la *gestion de l'incertitude (management of uncertainty)* ;
- les *considérations esthétiques et ergonomiques* : par exemple, des cartes de réseaux de métros distorsionnent la topographie (distances et positions relatives des stations et des correspondances) ;
- l'*utilisation de mêmes appellations pour des choses différentes* : dans les données, instruments, théories, méthodes et résultats.

De plus, si l'on considère l'incertitude en géographie d'une manière plus générale, elle est liée à un ensemble de causes qui peuvent s'illustrer par leur hétérogénéité et leur multidimensionnalité. Par exemple, [Fusco, 2013] :

⁶ « Paradoxically, in a way analogous to scientific theories, they have gained in apparent certainty of knowledge as they have lost in immediacy of experience », p. 84.

- le *traitement des données spatiales* : choix des échelles spatiales et temporelles et de leurs changements (composition, combinaisons, et désagrégation), et choix des données produites (diversité des sources, contexte, échantillonnage). → Pollution aérienne, chômage, etc.
- la *définition des objets géographiques* : définition (ambiguïté et manque de consensus, applicabilité), conceptualisation (limites et continuités). → Régions, lieux, paysages, etc.
- les *schémas d'appréhension des phénomènes spatiaux* : théorisation (conceptualisation consistante, validation, adaptation, application). → Pertinence d'un schéma « centre-périphérie » pour expliquer comment fonctionne un espace, etc.
- la *complexité des systèmes spatiaux* : connaissance d'états et systèmes spatiaux et de processus spatiotemporels (reconstitutions et prospective incertaines). → Réalité de la gentrification d'un quartier, etc.
- la *modélisation spatiale* : introduction et propagation de l'incertitude des règles (agents, algorithmes, équations...) à travers les modèles imparfaits d'une simulation (descriptions, expérimentations virtuelles ...) et propagation de cette incertitude formelle à celle des résultats. → Diffusion d'une maladie dans l'espace, etc.
- la *représentation de la connaissance spatiale* : représentation cartographique statique ou dynamique (par exemple par des choix sémiologiques graphiques) et autres représentations de la connaissance sur l'espace, communication spatiale. → Limites des zones inondables incertaines, probabilités qu'un lieu fasse partie de groupes de lieux, etc.
- la *subjectivité dans les phénomènes spatiaux* : conception et perception des espaces en relation avec des approches subjectives (sensibilité, cognition, culture et intentionnalité des personnes). → Gêne acoustique en ville, structuration de propriétés foncières et durabilité des inégalités sociales, etc.
- le *processus décisionnel sur l'espace* : aide à la décision autour de l'aménagement du territoire (études d'impacts, scénarios). → Impact d'une nouvelle carte scolaire avec stratégies familiales de contournement.

3.2 Les modalités de l'incertitude : un point de vue linguistique

[Sauri *et al.*, 2006, p. 1 et 2] remarquent un éventail de six sortes de modalités dans le discours : des degrés de possibilité, des croyances, des évidentialités, des attentes, des tentatives et des commandements. Ils mentionnent le lien avec les inférences, qui ne sont pas les mêmes lorsqu'elles se basent, dans un cas, sur des événements non réalisés ou seulement peut-être réalisés, ou bien, dans un autre cas, sur des événements considérés réels. Leur description de catégories lexicales de marqueurs de la modalité dans la langue anglaise comporte :

- des prédicats verbaux : *'claim'*, *'suggest'*, *'think'*, *'believe'*, *'suppose'*, ...
- des prédicats nominaux : *'promise'*, *'hope'*, *'love'*, *'request'*, ...
- des prédicats adjectivaux : *'ready'*, *'eager'*, *'able'*, ...
- des auxiliaires modaux, dont ceux de :
 - possibilité : *could*, *may*.
 - obligation : *must*, *have to*.
 - nécessité : *need to*.
- des modificateurs adverbiaux de :
 - possibilité : *probably*, *perhaps*.
 - Fréquence : *usually*, *always*.

Il est des expressions de la modalité qui sont, de manière générale, fréquemment mises en relief dans la recherche en linguistique, par exemple :

- Les expressions considérées à partir des verbes modaux : composées des verbes modaux *'may'*, *'might'*, *'can'*, *'could'*, *'shall'*, *'should'*, *'would'*, *'will'*, *'must'*, ... accolés à des verbes courants tels que *'be'*, *'have'*, ou moins courants, mais plus de spécialités des textes étudiés, par exemple dans des articles scientifiques, *'result'*, *'indicate'*, *'suggest'*, *'cause'*, etc., et changeables sous l'effet de modificateurs tels que, par exemple dans des articles scientifiques, *'often'*, *'sometimes'*, *'potentially'*, *'mostly'*, *'around'* et *'approximately'*.

- Les expressions considérées à partir des degrés de probabilité : par exemple celles décrites par [Mosteller et Youtz, 1990, tableau 2, p. 6 ; et p. 8-9]⁷ (voir aussi l'explication et la liste des expressions p. 48), composées de termes de base tels que '*always*', '*certain*', '*frequent*', '*likely*', '*probable*', '*often*', '*possible*', '*chance*', ou '*usually*', et changeables sous l'effet de modificateurs et préfixes ou suffixes tels que '*not*', '*very*', '*low*', '*high*', '*poor*', '*moderate*', '*in-*', '*im-*' et '*un-*'.
- Les expressions considérées à partir de l'influence de la négation sur les modalités : par exemple, parmi celles incluant préfixes et suffixes citées par [Benamara et al., 2012, p. 10] qui ne limitent pas leur étude aux textes scientifiques, se trouvent '*any*', '*anything*', '*ever*', '*maybe*', '*certainly*', '*il-*' et '*-less*'.

Concernant l'anglais, [Kratzer, 2012, p. 28] signale par exemple ce rôle, d'expressions traditionnellement centrales dans la recherche, pour « *necessarily*, *possibly*, *must*, *can*, *should* » ou « *may* ». Cependant, ces expressions et les nombreuses autres de la modalité expriment leurs subtilités ; et il apparaît dans un grand nombre de ces occasions que, en linguistique comme dans d'autres disciplines, l'observateur influence l'observé. Pour débiter par un raccourci temporaire, nous constatons qu'il est difficile d'exprimer, et même de traduire fidèlement, les modalités⁸ :

- Les *modalités* sont comparables à l'âme⁹ de la phrase. L'imprécision de leur définition est indépendante de la langue.
- Les *marqueurs de modalité* sont les éléments qui renvoient à l'attitude de l'énonciateur, et la faculté de celui-ci pour les exprimer est dépendante de la langue utilisée. Par exemple, '*semble-t-il*', '*possibly*', '*a lo mejor*', '*probabilmente*' et '*vielleicht*' peuvent être d'une part employés comme équivalents, mais sont d'autre part différemment nuancés dans leurs langues respectives¹⁰.

⁷[Mosteller et Youtz, 1990, p. 10] mettent en avant 52 expressions sur 300 trouvées dans 40 articles scientifiques traitant des expressions de probabilités de sens qualitatifs.

⁸Seules les modalités textuelles seront abordées ici ; à la différence d'autres modalités, telles que celles de la voix traitées par la reconnaissance prosodique automatique, par exemple mise en œuvre par [Sztahó et al., 2009].

⁹L'expression « âme de la phrase » est employée par [Bally, 1965].

¹⁰Pour des expressions simples exprimant la probabilité, des différences entre langues sont qualifiées globalement de « légères » par [Mosteller et Youtz, 1990, p. 2 et 9].

- Les *contextes des modalités* influencent aussi l'attitude du lecteur. Par exemple, [Mosteller et Youtz, 1990, p. 3 et 5] notent que les probabilités sont perçues plus modestes lorsqu'elles portent sur des événements ordinaires¹¹.

Pour [Le Querler, 1996, p. 49], « Les théories des modalités en linguistique constituent une véritable « nébuleuse »¹². Pareillement, repris par [Vold, 2008, p. 42], [Perkins, 1983] exprime que « il n'y a aucun moyen facile pour découvrir ce qu'est en fait la modalité »¹³, et [Herslund, 1989] appelle la notion de modalité « *this elusive notion* ». Par ailleurs, des usages du vocabulaire employé pour cerner les modalités peuvent comporter des ambiguïtés, ce que signalent par exemple [Halliday et Matthiessen, 2014, p. 142] en particulier pour le mot anglais *modal* pouvant signifier soit *mood* soit *modality*, expressions pour lesquelles ils tiennent à réserver des sens distincts, comme pour d'autres expressions qu'ils emploient à leur manière autour des modalités.

Pour [Desclés et Jackiewicz, 2006, p. 45-46], l'énonciateur utilise les modalités pour exprimer ses représentations de la causalité : « Quand un énoncé causal est modifié par une modalité épistémique¹⁴, il rend compte d'un état de connaissance ou d'un degré de certitude de l'énonciateur par rapport à la réalité du lien causal prédiqué ». Dû à la « présence significative des éléments modaux », ils considèrent deux constituants de l'expression de causalité, par exemple dans la phrase « L'ingestion de vin aurait un effet protecteur sur l'organisme », qui formeraient une double dimension langagière (voir schéma 3.1, p. 41, basé sur le leur), dont les composantes sont étroitement liées, voire même indissociables et intriquées dans une relation complexe. Ces deux constituants sont :

- la dimension de la *relation causale*, efficiente, entre deux situations ;
- et la dimension de la *relation subjective*, entre la relation causale-effective et l'énonciateur qui la prend en charge. Cette dimension subjective pouvant influencer ce que contient la relation entre les deux situations, et guider l'interprétation qu'en fait l'énonciateur en présence de marques de polysémie.

¹¹Dans l'article auquel les auteurs se réfèrent en particulier pour cet exemple, il est précisé, [Pepper et Prytulak, 1974, p. 97], que les sujets qui étaient testés dans l'étude correspondante étaient un groupe de 33 universitaires.

¹²Terme que Nicole Le Querler reprend de [Meunier, 1981], en le citant.

¹³Traduction de « *it is by no means easy to find out what modality actually is* ».

¹⁴Voir § 3.2.2 *Les modalités épistémiques*, p. 48.

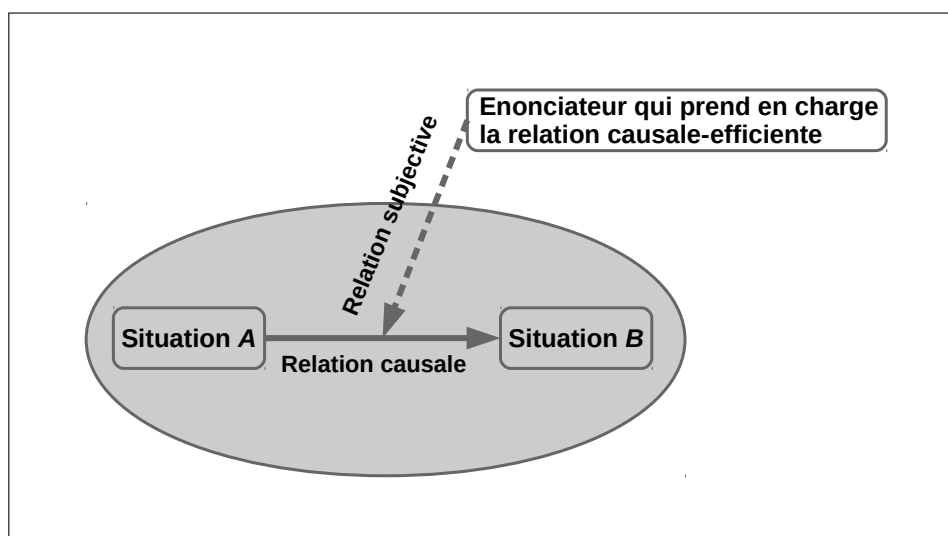


Schéma 3.1: Double dimension de la causalité dans la langue. Adaptation du schéma de [Desclés et Jackiewicz, 2006, p. 45, fig. 3].

Pour le francophone [Bally, 1965, § 26, 28, 48 et 227]¹⁵, la phrase explicite comprend deux parties complémentaires :

- Le *dictum* : « corrélatif du procès qui constitue la représentation », p. ex. la pluie, une guérison.
- Le *modus* : « expression de la modalité, corrélatrice à l'expression du sujet pensant » qui se compose d'un *verbe modal* (verbe qui contient l'assertion d'un jugement de fait ou de valeur : p. ex. « croire, se réjouir, souhaiter ») et d'un sujet modal (son sujet).

Le dictum et le modus peuvent être fondus l'un avec l'autre par l'emploi des modes verbaux, ce qui se reconnaît dans des sous-entendus. Par exemple :

« *Sortez !* = "*Je veux que* vous sortiez" ».

[Le Querler, 1996, p. 67-70] précise que « La portée (ou incidence ou *scope*) des modalités est une composante essentielle de leur signification ». La portée peut être :

- Syntaxiquement *intra-prédicative* sous la forme sujet–MODALITÉ–verbe :

¹⁵Nous notons que ce point de vue de Charles Bally concernant les modalités est une référence chez la plupart des auteurs francophones que nous avons consulté.

3.2. LES MODALITÉS DE L'INCERTITUDE : UN POINT DE VUE LINGUISTIQUE

- Sémantiquement *intra-prédicative*.
P. ex. : 'Il *peut* venir' (\cong 'Il *a la possibilité de* venir')¹⁶.
- Sémantiquement *extra-prédicative*.
P. ex. : 'Il est *peut-être* venu' (\cong '*Peut-être qu'*il est venu')
- syntaxiquement *extra-prédicative* sous la forme MODALITÉ → [sujet-verbe] :
 - Sémantiquement *extra-prédicative*.
P. ex. : sur l'énoncé '*Peut-être qu'*il est venu'
 - P. ex. : sur l'énonciation '*Franchement*, il n'est pas bien malin'

Selon les besoins des auteurs, d'autres caractéristiques des modalités peuvent être mises en avant, par exemple, des polarités, des degrés ou des intervalles de validité estimée (plus ou moins floue et élastique).

Les anglophones [Halliday et Matthiessen, 2014, p. 144, 148 et 176-178] expliquent que la modalité est un ensemble de « degrés intermédiaires », tels que '*sometimes*' ou '*maybe*', de divers types (voir tableaux 3.2, p. 44, et 3.3, p. 45), entre deux pôles inverses ; et qu'elle sert à analyser et interpréter (*construes*) une « zone » d'incertitude « qui s'étend entre 'oui' et 'non' » (ces extrêmes ou leurs avatars constituant, eux, à part, la polarité) et qui permet d'exprimer une évaluation de la validité de ce qui est affirmé par une déclaration, un questionnement, une imposition ou une proposition^{17 18}. Ils utilisent le terme *zone* et non pas *dimension*, et précisent ensuite que « il y a plus d'une route » entre ce *oui* et ce *non*. Poursuivant ce raisonnement, ils décrivent deux trajectoires d'incertitude possibles avec, entre autres, les expressions et exemples suivants :

- Les *propositions* : probabilités relatives entre '*it is*' et '*it isn't*'.
→ '*it must be*', '*it will be*', '*it may be*'.
P. ex. : '*that will be John, he'll sit there all day*'

¹⁶[Le Querler, 1996, p. 6] définit le signe « \cong » pour signifier « l'énoncé qui suit est sémantiquement équivalent au précédent ».

¹⁷« *stated, questioned, commanded or offered* ».

¹⁸Alors que notre perspective est l'automatisation, il peut être remarqué que [Halliday, 1995a, p. 253], s'exprimant sur ce qu'il considère être une série de malentendus « désastreux pour la linguistique », affirme une précaution quant à l'utilisation d'un certain type de logique pour comprendre la langue (*language*) : il affirme qu'il faut reconnaître que c'est dans la langue que se construit la connaissance, et que la logique formelle n'en est pas la clé « à cause des limitations de la logique conventionnelle ». Il avertit aussi, [Halliday, 1995b, p. 198], que les linguistes ont tendance à trop simplifier la langue en idéalisant les données.

Elles sont déclinées en deux sous-trajectoires formant des échelles de probabilité (dont une, autre concept, dit elle-même *de degrés de probabilité*) avec des compléments adverbiaux modaux (*modal adjuncts*) de probabilité ou habitude (voir la table 3.3, p. 45), plus une trajectoire combinée avec un opérateur modal (*finite modal operator*¹⁹) (voir la table 3.2, p. 44) dans le groupe verbal :

- degrés de probabilité → 'possibly', 'probably', 'certainly'²⁰.
Ex. : 'that's **probably** John'.
- degrés d'habitude → 'sometimes', 'usually', 'always'²¹.
Ex. : 'he **usually** sits there all day'.
- degrés de probabilité et habitude combinés avec des opérateurs modaux.
Ex.²² : 'that'll **probably** be John, he'll **usually** sit there all day'.

- Les *prescriptions (proposals)* : options discrétionnaires entre 'do!' et 'don't!' → 'you must do', 'you should do', 'you may do'.

Elles sont divisées en deux sous-trajectoires formant deux échelles²³, d'obligation pour l'imposition (*command*) et d'inclination pour l'offre, dont les modalités peuvent être exprimées l'une et l'autre avec des opérateurs modaux (voir tableau 3.2, p. 44) et des Prédicateurs²⁴ (*Predicators*) :

- degrés d'obligation → 'allowed to', 'supposed to', 'required to'
Ex. avec un opérateur modal : 'you **should** know that'.
Ex. avec un prédicateur verbe passif : 'you're **supposed to** know that'.
- degrés d'inclination → 'willing to', 'anxious to', 'determined to'
Ex. avec un opérateur modal : 'I'll **help** them'.
Ex. avec un prédicateur adjectif : 'I'm **anxious** to help them'.

¹⁹Une forme verbale *finie* montre ou comporte le temps, la personne et le nombre, par opposition à un infinitif. [Halliday et Matthiessen, 2014, p. 140] utilisent aussi en particulier le terme *fini* lorsqu'il s'agit de la modalité exprimée par des verbes modaux : « *The Finite element is one of a small number of verbal operators expressing tense (e.g. is, has) or modality (e.g. can, must)* ».

²⁰Ils rappellent que 'that's *certainly* John' est moins certain que 'that's John'.

²¹Ils rappellent que 'it *always* rains in summer' est moins invariable que 'it rains in summer'.

²²[Halliday et Matthiessen, 2014, p. 177] considèrent 'will' comme un opérateur modal, ce que montre cet exemple, bien qu'il figure comme opérateur temporel du futur dans leur table des opérateurs verbaux finis (voir table 3.2, p. 44)

²³Exemple de diversification de termes dans la famille des *modalités* : les auteurs nomment ces deux échelles '*modulation*' (en anglais) pour les distinguer de la modalité en général qu'ils nomment *modalization*.

²⁴Les auteurs emploient le terme 'Predicator' avec une majuscule, gardée pour signifier que c'est leur concept.

3.2. LES MODALITÉS DE L'INCERTITUDE : UN POINT DE VUE LINGUISTIQUE

Types d'opérateurs	Positifs	Négatifs
Temporaux au futur	will, shall, would, should	won't, shan't, wouldn't, shouldn't
Modaux bas	can, may, could, might, (dare)	needn't, doesn't need to, didn't need to, have to
Modaux médians	will, would, should, is to, was to	won't, wouldn't, shouldn't, (isn't to, wasn't to)
Modaux hauts	must, ought to, need, has to, had to	mustn't, oughtn't to, can't, couldn't, (mayn't, mightn't, hasn't to, hadn't to)

Tableau 3.2: Opérateurs verbaux finis (*finite verbal operators*) en langue anglaise selon [Halliday et Matthiessen, 2014, p. 145], extrait adapté : opérateurs temporeux du futur, et opérateurs modaux.

[Halliday et Matthiessen, 2014, p. 181-182](voir le tableau 3.4 p. 45) notent par ailleurs que, dans le cas de fortes probabilités, l'expression de la subjectivité dépend, d'une part, de la source de la conviction (ex. : 'We are certain' vs 'It is certain'), et, d'autre part, de l'opposition implicite vs explicite (ex. : 'must be certain' vs 'are certain').

La *modalité* est un concept historiquement et encore actuellement instable en linguistique. Les chercheurs en linguistique utilisent souvent ce concept avec précaution, d'autant plus lorsqu'il s'agit de le définir : chaque auteur est amené à préciser les différents types des modalités, ainsi que leurs sens et leurs relations, dans le cadre où il les emploie. Cette instabilité conceptuelle est tour à tour la cause et la conséquence de la multitude de définitions autour de la modalité.

Dans les sections suivantes, nous donnerons quelques définitions de la modalité et expliquerons la relation entre modalité et subjectivité.

3.2.1 Approche des définitions des modalités

L'une des caractéristiques de la notion de *modalité* (en anglais *mood*, ou aussi *modality* ou *mode*) est qu'elle joue des rôles parfois instables ou multiples : ses sens peuvent se chevaucher dans ses définitions (dont exemples de définitions en annexe A.1, p. 211) ; et les structures et articulations qui organisent ses types (voir exemples des schémas 3.4, p. 51, et 3.5, p. 51) ne sont pas consensuelles.

Type	Signification	Exemples
probability	how likely?	probably, possibly, certainly, perhaps, maybe
usuality	how often?	usually, sometimes, always, (n)ever, often, seldom
typicality	how typical?	occasionally, generally, regularly, for the most part
obviousness	how obvious?	of course, surely, obviously, clearly
opinion	I think	in my opinion, personally, to my mind
admission	I admit	frankly, to be honest, to tell you the truth
persuasion	I assure you	honestly, really, believe me, seriously
entreaty	I request you	please, kindly
presumption	I presume	evidently, apparently, no doubt, presumably
desirability	how desirable?	(un)fortunately, to my delight/distress, regrettably, hopefully
reservation	how reliable?	at first, tentatively, provisionally, looking back on it
validation	how valid?	broadly speaking, in general, on the whole, strictly speaking, in principle
evaluation	how sensible?	(un)wisely, understandably, mistakenly, foolishly
prediction	how expected?	to my surprise, surprisingly, as expected, by chance

Tableau 3.3: Compléments adverbiaux modaux en langue anglaise (*modal adjuncts*), selon [Halliday et Matthiessen, 2014, tableau 3-5, p. 109].

	Subjectif	Objectif
Implicite	<i>must</i>	<i>certainly</i>
Explicite	<i>I'm certain that ...</i>	<i>it is certain that ...</i>

Tableau 3.4: Matrice de Halliday et Matthiessen pour illustrer les dimensions implicite-explicite et subjectif-objectif des modalités. Source : [Halliday et Matthiessen, 2014, matrice p. 181].

Il existe des définitions de la modalité qui s'éloignent du concept dont nous voulons faire usage. Pour aborder la notion de modalité de manière pratique avant de la préciser, en voici donc une première définition pour suivre notre propos : *la modalité est un élément de la langue, par lequel l'énonciateur exprime un degré*

3.2. LES MODALITÉS DE L'INCERTITUDE : UN POINT DE VUE LINGUISTIQUE

ou une variation d'une assertion ; par exemple, à propos d'une vérité, ou à propos de l'état d'une chose ou d'un être.

D'après [Barreau, 2013, p. 20-23], les mégariques²⁵ et les stoïciens employaient déjà « des notions modales (possible, impossible, nécessaire) » pour définir « les opérations logiques en fonction des valeurs de vérité », et, d'autre part :

« On a vu se créer des logiques déontiques, des logiques juridiques, et même des logiques épistémiques qui font intervenir d'ordinaire des opérateurs modaux ».

Parmi les différents types de modalités, ces notions se retrouvent dans les *modalités* qui correspondent au sens classique logico-linguistique : les modalités aléthiques²⁶, qui, déjà chez Aristote selon [Rendsvig et Symons, 2019]²⁷, servent à signifier dans un énoncé le possible, le contingent, le nécessaire ou l'impossible.

Des figures géométriques, notamment le triangle²⁸, le carré, l'hexagone²⁹ et le cube³⁰ peuvent être utilisées pour représenter sur leurs sommets des propositions logiques, et pour représenter par les liens entre ces sommets des correspondances et oppositions de ces propositions. Le traditionnellement nommé *carré d'Aristote*³¹ (voir schéma 3.2, p. 47) relève de ce procédé³² pour représenter en un diagramme les relations entre des propositions logiques simples affirmatives et négatives en langage naturel, telles que décrites par Aristote, selon par exemple [Correia, 2017, p. 1-5]. Les quatre propositions logiques sur les sommets du carré d'Aristote sont une manière d'exprimer les interrelations des modalités aléthiques.

[Vinzerich, 2007, p. 49-51] reprend le carré d'Aristote³³ et en donne une version moderne, issue elle-même de l'évolution du carré d'Aristote en *carré logique*

²⁵De l'école de philosophie fondée par Euclide à Mégare (ville de Grèce) vers -400.

²⁶Dites aussi modalités *ontiques*, et, selon [Gardies, 1979, p. 65], dites aussi modalités *temporelles* chez les mégariques.

²⁷« While Aristotle addressed the four alethic modes of possibility, necessity, impossibility, and contingency [...] ».

²⁸[Le Querler, 1996, p. 48].

²⁹Blanché Robert (1966), « Structures intellectuelles : Essai sur l'organisation systématique des concepts », par ex. p. 56. Vrin, Paris. Cité par [Gardies, 1979, p. 21].

³⁰[Ciucci et al., 2015, p. 5-7 (p. 123-125 du recueil)].

³¹Dit aussi *carré d'opposition d'Aristote* ou *carré des oppositions d'Aristote*.

³²Procédé repris entre autres dans le *carré d'Apulée*, traditionnel aussi, qui s'inspire de celui d'Aristote.

³³A la suite de, par exemple dans sa bibliographie, [Gardies, 1979, p. 17-21] et [Le Querler, 1996, p. 36-39 et 46-48].

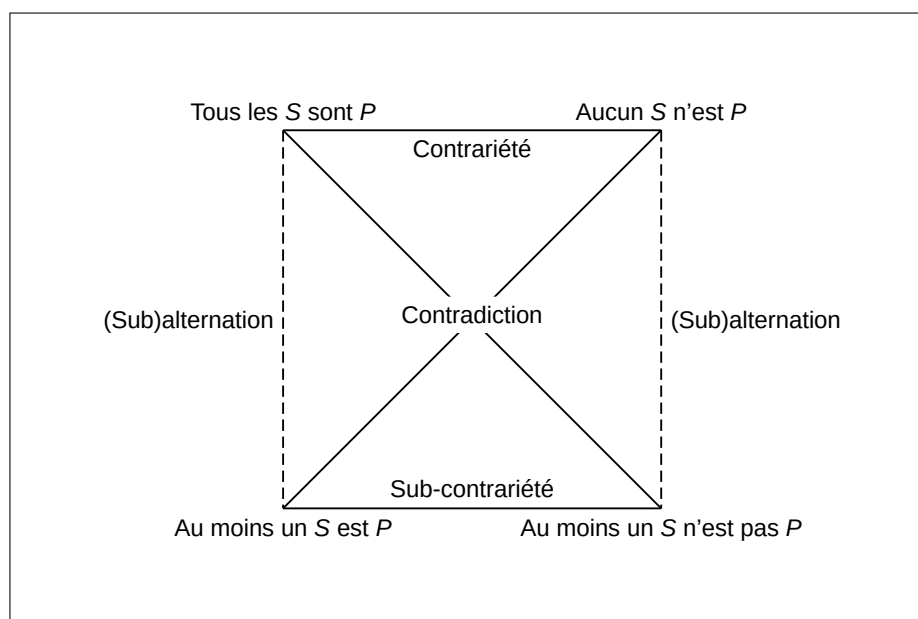


Schéma 3.2: Carré d'Aristote selon [Correia, 2017, p. 2], traduction. Les lignes verticales discontinues représentent des liens de relations subalternes qui, [Correia, 2017, p. 2, nbp 3], n'ont pas été mentionnées par Aristote.

d'Apulée, afin d'exprimer les interrelations des modalités d'énoncé logiques aléthiques (voir schéma 3.3, p. 48).

Engel³⁴ signale que la difficulté des notions modales qui traditionnellement « ont été rejetées aux frontières de la logique classique du vrai et du faux » est contrebalancée par leur intérêt. Par rapport à cette modalité aléthique classique, il distingue parmi les modalités de la logique les modalités non-aléthiques, dont celles déontiques (« il est permis que », « il est obligatoire que ») et épistémiques (« croire que », « savoir que ») :

« On appelle également « modalité », au sens large, toute modification quelconque du sens d'une proposition, soit par adjonction d'adverbes, soit par subordination de cette proposition à certains verbes ou formes verbales ».

³⁴Engel Pascal (Consulté le 29/01/2018), « MODALITÉS, logique » in Encyclopaedia Universalis, en ligne. Article encyclopédique. URL : <https://www.universalis.fr/encyclopedie/modalites-logique/>

3.2. LES MODALITÉS DE L'INCERTITUDE : UN POINT DE VUE LINGUISTIQUE

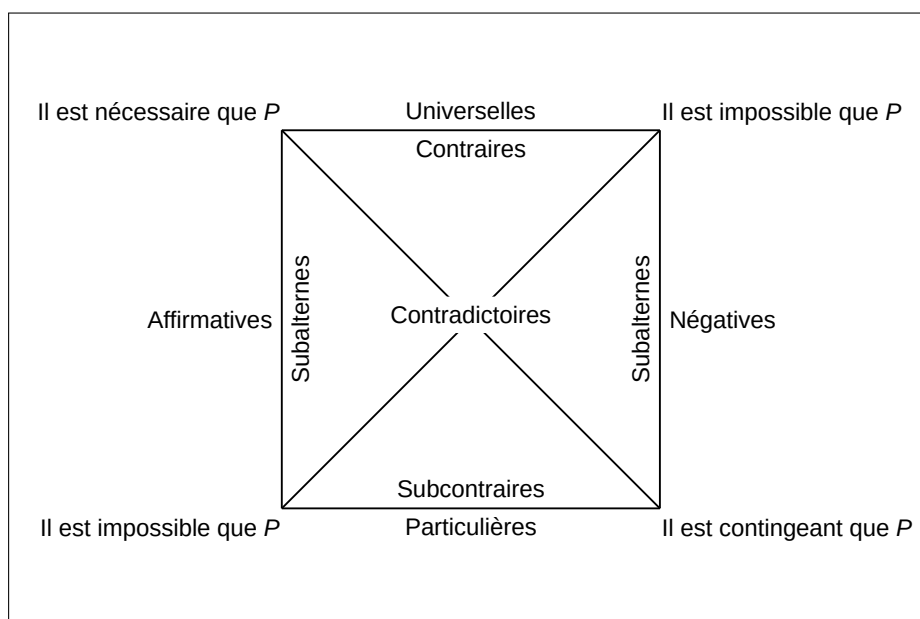


Schéma 3.3: Carré des modalités aléthiques selon [Vinzerich, 2007], s'appuyant sur [Gardies, 1979] et [Le Querler, 1996].

3.2.2 Les modalités épistémiques

Dans une analyse commençant par l'examen des résultats de 20 études précédentes faites par différents auteurs auprès de scientifiques ou de personnels médicaux, [Mosteller et Youtz, 1990] ont comparé la sémantique d'expressions de la probabilité en langue anglaise. L'analyse des réponses à un questionnaire d'estimations de probabilités qu'ils ont administré ensuite à un panel d'auteurs scientifiques leur a permis de montrer que 52 expressions qu'ils ont sélectionnées à partir des 20 études précédentes peuvent être mises chacune en correspondance avec un pourcentage de probabilité suffisamment stable, à l'exception de l'expression '*possible*' qui peut être interprétée de différentes manières. Ils ont regroupé ces expressions, [Mosteller et Youtz, 1990, tableau 2], par probabilité décroissante dans de petites séries lexicales : *always*, *almost always* / *certain*, *almost certain* / *very frequent*, *frequent*, *not infrequent*, *infrequent*, *very infrequent* / *very high probability*, *high probability*, *moderate probability*, *low probability*, *very low probability* / *very likely*, *likely*, *unlikely*, *very unlikely* / *very probable*, *probable*, *improbable*, *very improbable* / *very often*, *often*, *more often than not*, *as often as not*, *less often than not*, *not often*, *not very often* / *possible*, *impossible* / *high chance*, *better than even chance*, *even chance*, *less than an even chance*, *poor*

chance, low chance / liable to happen, might happen / usually, unusually / sometimes, once in a while, not unreasonable, occasionally, now and then / seldom, very seldom / rarely, very rarely / almost never, never.

Pour [Vold, 2008, p. 10 et 12] :

« Le recours à la modalité épistémique, ou au marquage d'incertitude, constitue une stratégie d'atténuation fréquemment employée dans le discours scientifique et jugée essentielle pour les auteurs de textes scientifiques spécialisés ».

Elle remarque que les modalisateurs épistémiques sont des marques de subjectivité qui :

- Expriment « le jugement du locuteur par rapport à la valeur de vérité d'un énoncé ».
- Peuvent transmettre « une prédiction ou une évaluation de points de vue soutenus par d'autres chercheurs ».
- Peuvent être employés par l'auteur pour « des raisons de persuasion et de négociation ».

[Le Querler, 1996, p. 61-67] analyse la modalité comme « expression de l'attitude du locuteur par rapport au contenu propositionnel de son énoncé ». Elle définit différents types de modalités et les organise entre elles et les différencie les unes par rapport aux autres. Partant du sujet énonciateur (locuteur ou sujet de la phrase), elle considère la ramification des modalités en trois types :

- Modalités *objectives* (= *implicatives*) lorsqu'elles subordonnent le contenu de la proposition à une autre proposition, et ne dépendent alors ni du jugement, ni de l'appréciation, ni de la volonté de l'énonciateur. Ce sont les *modalités implicatives* au sens logique (« p implique q »), et aussi celles au sens large : « rapports de condition (« Pour grandir, il faut [...] »), de conséquence (« Il est si grand que [...] »), de but, et d'opposition (« Si tu bouges, [...] ») ».
- Modalités *intersubjectives* lorsqu'elles établissent un rapport entre l'énonciateur et une autre personne.

3.2. LES MODALITÉS DE L'INCERTITUDE : UN POINT DE VUE LINGUISTIQUE

- Modalités *subjectives* lorsqu'elles portent sur le rapport entre l'énonciateur et le contenu de sa proposition.

Elle subdivise ces modalités subjectives en modalités appréciatives et modalités épistémiques, exprimables par des marqueurs lexicaux, morphologiques ou intonatifs :

- Les modalités *appréciatives* : pour exprimer les appréciations personnelles sur le contenu de la proposition, telles qu'une approbation ou une indignation ('j'apprécie que', *heureusement*, 'hélas !').
- Les modalités *épistémiques* : pour les assertions des degrés de certitude pouvant dépendre d'un verbe (« tel que *savoir*, *douter*, *ignorer*, etc. »), et pouvant être modifiées par un adverbe modal épistémique (« '*peut-être*', '*sans doute*', *probablement*, *certainement*... »), une proposition en incise ('j'en doute'), certains emplois de *pouvoir* et *devoir*, (ainsi que par une intonation ou une mimique).

« On appelle modalités logiques les diverses manières d'envisager le prédicat de la phrase comme vrai, contingent (ou nécessaire), probable (ou possible) » et « On donne parfois le nom de certitude à la modalité logique du nécessaire (ex. : Il doit pleuvoir demain) », [Dubois *et al.*, 2007]. Pour situer les modalités épistémiques, voici un double schéma (voir schéma 3.4, p. 51, et schéma 3.5, p. 51) illustrant une façon³⁵ d'envisager leurs rapports structurels avec d'autres modalités, leurs dimensions et des exemples. C'est un choix de distinction des modalités voulu pratique, et qui peut se relativiser parmi d'autres façons de présenter les modalités : [Fintel (von), 2006, p. 21], par exemple, présente aussi une découpe de distinction des significations modales en aléthiques, déontiques et épistémiques ; mais il y adjoint, selon ses termes, celles boulétiques (*bouletic* ou *boulomaic*) qui expriment le possible ou nécessaire au vu d'un ensemble de conditions particulières, celles circonstancielles (*circumstantial*) qui expriment le possible ou nécessaire au vu du désir personnel, et celles téléologiques (*teleological*) qui expriment le possible ou nécessaire pour atteindre un but particulier, tout en signalant qu'il existe un foisonnement de la nomenclature des modalités (*taxonomic exuberance*) qui va bien au-delà.

Le sens des modalités à l'intérieur de cette organisation des modalités logiques (celle des schémas 3.4, p. 51, et 3.5, p. 51) peut être précisé étymologiquement

³⁵ Au départ inspirée de : Büyükgüzel Safinaz (2011), *Modalité et subjectivité : regard et positionnement du locuteur* in *Synergies Turquie* num. 4 - 2011, p. 139-151.



Schéma 3.4: Une structure des modalités linguistiques : dimensions.

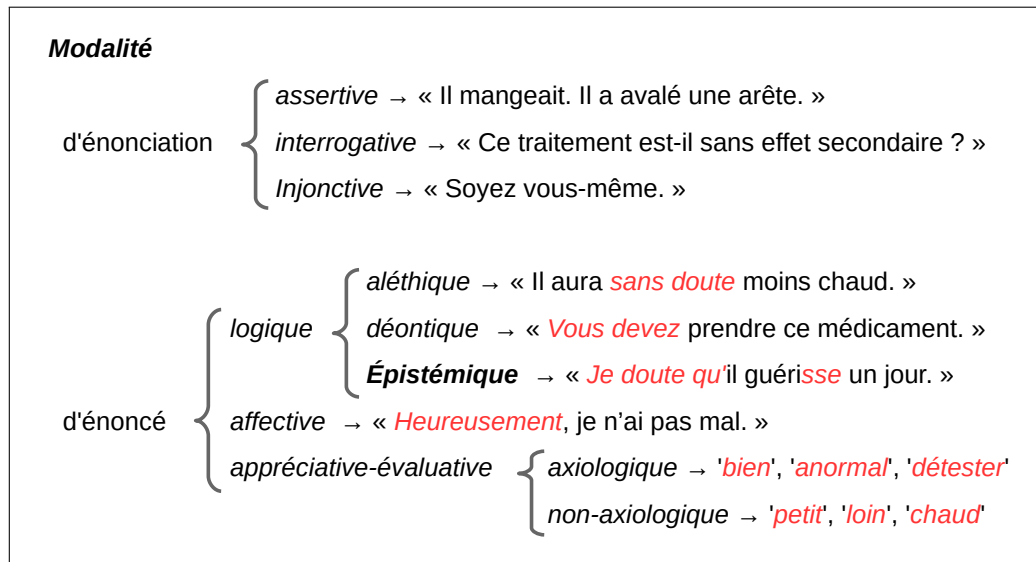


Schéma 3.5: Une structure des modalités linguistiques : exemples.

par l'origine grecque³⁶ de l'adjectif qui les qualifie : *aléthique* vient de ἀλήθεια [alèthéia] qui signifie *vérité, réalité* ; et *déontique* vient de δέον [déon] qui signifie *ce qu'il faut (besoin, nécessité), ce qui convient (ce qui est opportun)*, et à aussi donné 'déontologie'. Quant à l'adjectif *épistémique* qui qualifie les modalités que nous allons utiliser dans nos travaux, il a pour origine ἐπιστήμη [épistèmhè] qui signifie *science, application de l'esprit, étude* ; mais pour éviter une possible ambiguïté, il faut préciser que l'emploi du terme 'épistémique' ne relève pas ici du domaine de l'épistémologie (domaine qui étudie la science « ou plutôt » les sciences, et les activités scientifiques, et d'abord les « disciplines qui suivent, par choix délibéré, des méthodes scientifiques », [Barreau, 2013, p. 3, 96 et 120]) : il relève du domaine de la linguistique, où l'expression *modalité épistémique* correspond à ce qu'exprime l'énonciateur à propos de l'incertitude sur les faits qu'il rapporte.

Même si nous ne nous en inspirons pas, nous remarquons que, s'agissant de concepts servant *in fine* à manier l'incertitude, des structures différentes peuvent aussi se dégager dans d'autres disciplines proches du traitement automatique des langues. Par exemple, en intelligence artificielle, [Smith *et al.*, 2018, p. 562-563] conceptualisent une *incertitude épistémique* (relative à un manque de données, de connaissances, de compréhension) en la distinguant d'une *incertitude aléatoire* (relative au caractère stochastique des données).

3.3 Typologie conceptuelle de l'incertitude

Les concepts utilisés pour étudier l'incertitude se recoupent souvent et leurs relations peuvent être nombreuses. Par exemple, [Morante et Sporleder, 2012, p. 2] mentionnent que la modalité est liée aux concepts de *hedging*, et aussi à ceux de *subjectivité*, *incertitude*, *évidentialité*, *croyance engagée (committed belief)*, et *factualité* ; et [Cruz Díaz et Maña López, 2019, p. 27] mentionne que le langage spéculatif est lié aux concepts de *spéculation*, mais aussi aux mêmes autres de *subjectivité*, *incertitude*, *évidentialité*, *croyance engagée* et *factualité*. Ces concepts à travers ou autour de l'incertitude varient aussi selon les langues naturelles choisies pour les nommer. Se limiter à explorer le français et l'anglais ne mène pas à l'exhaustivité dans le dénombrement de ces concepts, même pour ces deux seules langues. Selon les champs scientifiques, en français et en anglais,

³⁶Bailly A. (1901), *Abrégé du dictionnaire grec français*, p. 32-33, 188, 345. Hachette, Paris. Dépôt légal n° 8167.3.1984.

d'autres notions ou termes que ceux déjà présentés sont utilisés dans les discours sur l'incertitude sans qu'ils nous aient paru suffisamment opportuns pour nous y attarder spécialement ici (flou, équivoque, contradiction, imprécision, inconsistance, ignorance, questionnement, paradoxe, complication, indétermination, relativité, niveau d'évaluation, risque, etc.). Cependant, d'autres concepts sont utilisés avec un sens spécialisé et se sont révélés ici d'intérêt ou inspirants, tels que les *déictiques*, *hedging*, *boosting*, etc. Parmi ces concepts spécialisés en linguistique et en traitement des langues, en plus de ceux déjà mis en relief, ceux à continuation ont aussi retenu particulièrement notre attention en relation avec notre recherche : pour considérer plus et mieux la variété des expressions de l'incertitude, et pour la réflexion sur les relations entre les expressions de l'incertitude en vue de notre ontologie de l'incertitude.

3.3.1 Le *hedging*

Le *hedging* est un concept issu de la langue anglaise et est difficilement traduisible par une expression équivalente en langue française. La forme anglaise (*hedging*, *hedge*) est reprise dans des textes d'études linguistiques de langues diverses, dont notamment ceux de langue française.

L'expression *hedging* exprime une sorte de détour précautionneux du discours, lequel détour permet de ménager l'effet de l'affirmation qu'il accompagne, afin de se prémunir contre des réactions ou conséquences indésirables que cette affirmation pourrait provoquer ou induire. Par exemple, [Lakoff, 1973, p. 471-477] présente des expressions associés au phénomène du *hedging*, tels que '*pseudo-*', '*-ish*', '*for the most part*', '*more of a ... than anything else*', '*loosely speaking*' et '*technically*' ; et des contextes de phrases courtes telles que '*Loosely speaking, a whale is a fish*' et '*A whale is technically a mammal*' ; et [Novák, 2007, p. 18] présente, entre autres, des hedges de « focalisation » (*narrowing linguistic hedges*) tels que '*very*', '*extremely*', '*significantly*', et des hedges de « relâchement » (*widening linguistic hedges*) tels que '*more or less*', '*roughly*', '*very roughly*', lesquels sont inclus parmi les modificateurs linguistiques (*linguistic modifiers*) dans des recherches en logique floue, par exemple par [Kupka et Rusnok, 2019, p. 143].

D'un point de vue de la langue générale, [Atkins *et al.*, 1993] traduisent le terme *hedge* par, entre autres, les expressions suivantes : *une sauvegarde* ou *une couverture contre l'inflation*, *expliquer avec des détours*, *entourer d'une haie*, *couvrir* (un pari ou un risque), et *esquiver* (une question).

3.3. TYPOLOGIE CONCEPTUELLE DE L'INCERTITUDE

D'un point de vue linguistique, selon [Hyland, 1996b, p. 433-435], les *hedges* expriment la tentative et la possibilité dans le discours. Leurs marqueurs textuels incluent par exemple '*I think*', '*sort of*', '*maybe*' et '*possibly*'. Les déclarations scientifiques autres que celles affirmant une vérité catégorique sont des déclarations avec du hedging ou bien non-factuelles (*hedged or non-factive statement*) :

« Là où existe une incertitude sur le statut d'évidence d'une présomption entre des données et des hypothèses, des déclarations requièrent de faire varier des degrés de hedging »³⁷.

Dépendant du contexte socio-normatif scientifique, l'usage des *hedges* est une attitude qui :

« permet aux auteurs d'anticiper une possible opposition aux déclarations en exprimant des affirmations avec précision, prudence, et déférence diplomatique envers les avis de collègues »³⁸.

Il rappelle que des recherches sur les *hedges* les ont aussi associées à la transmission du vague ou à une prise de distance entre l'orateur et son discours ; et il souligne que sur ce sujet la recherche linguistique a fait la liaison avec les auxiliaires modaux comme outils épistémiques. Il ajoute que :

« En science, les *hedges* jouent un rôle crucial pour gagner la ratification d'affirmations de la part d'un groupe de pairs puissant en permettant aux auteurs de présenter des déclarations avec les précision, prudence et humilité appropriées. Les *hedges* aident à négocier la perspective à partir de laquelle des conclusions peuvent être acceptées »³⁹.

[Hyland, 1996b, p. 452] rappelle la nature dynamique et interactive de l'écrit scientifique, situé socialement et structuré pour atteindre des objectifs. Il insiste sur le rôle essentiel des *hedges* :

³⁷Traduction de « *Where there is uncertainty about the evidential status of the assumptions between data and hypotheses, claims require varying degrees of hedging* ».

³⁸Traduction de « *Hedges allow writers to anticipate possible opposition to claims by expressing statements with precision, caution, and diplomatic deference to the views of colleagues* ».

³⁹Traduction de « *In science, hedges play a critical role in gaining ratification for claims from a powerful peer group by allowing writers to present statements with appropriate accuracy, caution, and humility. Hedges help negotiate the perspective from which conclusions can be accepted* ».

« Les hedges sont abondants en science et jouent un rôle crucial dans l'écriture universitaire plus généralement. Ils constituent un élément essentiel de l'argumentation en présentant de nouvelles affirmations pour leur ratification, et figurent parmi les caractéristiques principales qui donnent leur forme aux articles de recherche en tant que véhicule principal de la connaissance nouvelle »⁴⁰.

[Hyland, 1996a, p. 251-276] précise que « Les Hedges expriment la tentative et la possibilité dans la communication, et leur utilisation appropriée dans le discours scientifique est vitale »⁴¹. Il propose, entre autres, un ensemble des formes principales et fonctions principales que prend le *hedging* dans un corpus d'articles de recherche en biologie cellulaire et moléculaire en langue anglaise. Ce corpus est étudié en partie à partir d'expressions de hedging identifiées par d'autres auteurs tels que Perkins, Kennedy et Holmes dans des textes de chercheurs : certaines de ces expressions sont identifiées positivement par ces auteurs, car elles apparaissent dans les textes étudiés, telles que '*appear*', '*assume*' et '*suggest*' ; et d'autres sont identifiées négativement, car elles n'y apparaissent pas, telles que '*infrequently*' et '*on occasion*'. Sur la base des données qu'il obtient, Hyland indique que presque la moitié des hedges apparaissent groupés par deux ou plus. Il remarque un phénomène de simultanéité des fonctions des hedges, qui peuvent se recouvrir avec une ambiguïté, dont les auteurs peuvent se servir ; et il nomme en particulier deux des fonctions des hedges dans les textes scientifiques :

- une fonction « à cause des croyances » (*content-motivated*) : adapter le contenu du texte pour qu'il soit en adéquation avec les croyances sur la réalité.
 - Exemples basés sur l'exactitude : « ... was generally confined to... », « ... was almost completely prevented by... », « ... might serve to... », « ... is a probable candidate for... ».
 - Exemples basés sur la limitation de l'engagement du rédacteur : It seems that ... », « These data indicate that ... », « ... the evidence strongly favours the conclusion that ... ».

⁴⁰Traduction de « *Hedges are abundant in science and play a critical role in academic writing more generally. They constitute an essential element of argumentation in presenting new claims for ratification, and are among the primary features which shape the research article as the principal vehicle for new knowledge* ».

⁴¹Traduction de « *Hedges express tentativeness and possibility in communication, and their appropriate use in scientific discourse is vital* ».

3.3. TYPOLOGIE CONCEPTUELLE DE L'INCERTITUDE

- Une fonction « à cause du lecteur » (*reader-motivated*) : rendre le contenu du texte acceptable dans la relation avec le lecteur. Par exemple par déférence (politesse, absence d'opposition explicite à la communauté ...) et par coopération (présentation d'alternatives, ouverture à des vérifications ...) : « In our hands, ... », « I believe that the ... », « I suggest here that the ... », « My analogy is with a ... ».

D'un point de vue lexical (*lexical hedges*) [Hyland, 1996a, p. 260-271] signale que 77 % de toutes les utilisations de hedges dans son propre corpus d'articles de recherche (26 articles de biologie cellulaire et moléculaire) correspondent à 'would' ou 'may' ou 'could'. Il remarque des particularités du langage des auteurs scientifiques, comme le fait que 'would' est souvent préféré à 'will' pour exprimer des probabilités au présent, et que 'could' est souvent préféré à 'must' pour exprimer une possibilité de tentative. Il distingue cinq principales formes lexicales de hedging⁴² :

- les verbes modaux : 'could', 'may', 'might', 'would', 'should', 'will' et 'must'. Il note que 'shall' et 'ought' sont absents de son corpus d'articles de recherche, et d'autres corpus qu'il compare.
- les verbes lexicaux épistémiques : leurs occurrences, dans son corpus d'articles de recherche, sont constituées à 56 % de 'indicate', 'suggest', 'appear' et 'propose', les deux premiers étant aussi nettement plus représentés dans d'autres corpus, alors que 'guess', 'reckon' et 'think' en sont absents. Il met en avant deux sub-catégories :
 - les verbes de jugement (sur la vérité d'une proposition) : 'suggest', 'propose', 'believe', 'speculate'. 'calculate', 'infer', 'indicate', 'imply', où, pour des expressions complexes, 'leads us to the conclusion that' et 'my interpretation is that'.
 - les verbes évidentiels : spéculatifs ou déductifs tels que 'speculate' et 'suggest', de discernement et perception tels que 'appear' et 'seem', et de recherche tels que 'investigate' et 'attempt' ou 'gain insight'.
- les adjectifs épistémiques : les plus fréquents, dans son corpus d'articles de recherche, sont 'likely' et 'unlikely', 'possible', 'most' et 'consistent with'.

⁴²Il indique plusieurs exemples de ces hedging en contexte, faisant parfois partie d'expressions complexes telles que « [...] would favor the hypothesis that [...] ».

- les adverbes épistémiques : 36 formes ont été identifiées, dont les plus courantes '*apparently*', '*probably*', '*essentially*', '*relatively*' et '*generally*'. Les catégories les plus mises en avant, qui peuvent se recouper, sont celles d'atténuation tels que '*usually*' et '*quite*' et '*almost*', de disjonction tels que '*probably*' et '*generally*', de diminution tels que '*partially*', de minimisation tels que '*rarely*', de rapprochement tels que '*virtually*' et '*almost*', de jugement tels que '*possibly*', de discernement et perception tels que '*apparently*' et '*evidently*', de scepticisme tels que '*allegedly*' et '*supposedly*', de contour (*contrast*, *limitation*) de réalité tels que '*intuitively*' et '*essentially*' et '*presumably*', et numérique tels que '*about*', '*some*', '*around*' et '*approximately*'.

D'un point de vue du discours [Hyland, 1996a, p. 271-274] signale aussi des hedges associés à trois stratégies du discours, soit 15 % des hedges de son corpus, lesquels font référence à :

- des faiblesses ou limites expérimentales : usage de conditionnel avec *if* tel que dans '*These results suggest that if [...], it would [...]*' et '*If correct, [...]*', manque de support tel que dans '*We have not been able to determine precisely whether [...]*' et '*So it is difficult to conclude whether [...]*'.
- des limitations du modèle, de la théorie ou de la méthodologie : '*We are aware of the concerns expressed in the literature (réf.) concerning [...]*', '*Viewed in this way [...]*', '*The procedure only identifies [...]*'.
- des manques ou inadéquations des connaissances : '*[...] do not know whether [...]*' ou '*One cannot exclude [...]*' ou '*Could such a putative [...] have a [...] significance?*'

[Hyland, 1996a, p. 274] remarque une tendance à associer des hedges dans une même phrase, soit 43 % des occurrences dans son corpus avec principalement '*suggest*', '*would*', '*possible*' et '*indicate*' : '*would probably suggest*', '*indicate (that ...) may*', '*could possibly*'.

Parmi les motifs divers qui gouvernent la sélection de hedges par les auteurs pour des usages scientifiques, [Hyland, 1996a, p. 276-277] souligne quatre fonctions génériques contextuelles⁴³ parmi celles qui ont pu être déterminées⁴⁴ (voir tableau 3.5 p. 58) :

⁴³Nous en retenons ici un sens qui pourrait aller au delà de celui qu'en donne l'auteur.

⁴⁴L'auteur remarque que ce qui motive certains hedges a « un haut degré d'indétermination ».

3.3. TYPOLOGIE CONCEPTUELLE DE L'INCERTITUDE

- usages liés directement aux faits :
 - fonctions de *vérification* : l'auteur reconnaît des incertitudes sur l'exactitude ou la vérité d'une affirmation.
 - fonctions de *spécification* : l'auteur apporte des précisions et nuances sur la pertinence de sa présentation des faits.
- usages liés à l'auteur :
 - fonctions d'*agentivité* : l'auteur affirme ou infirme son implication.
- usages liés indirectement au lecteur :
 - fonctions de *coopération* : l'auteur tente d'impliquer le lecteur, en proposant le partage d'arguments et références, et en mettant en avant son ouverture à d'autres possibilités.

Hedges liés aux faits	Verbes modaux épistémiques, adverbes épistémiques (de disjonction de contenu ou bien de style, de diminution), adjectifs épistémiques, indicateurs de manques ou inadéquations des connaissances.
Hedges liés à l'auteur	Verbes lexicaux épistémiques (de jugement, d'évaluation), expressions impersonnelles (voix passive ...), moyen out tournure épistémique thématique, attribution de source, référence impersonnelle (à une méthode, à un model ou à des conditions expérimentales).
Hedges liés au lecteur	Verbes lexicaux épistémiques (de jugement, de déduction), attribution personnelle, référence personnelle (à une méthode ou à un modèle), conditionnels, implication du lecteur par des questions directes et des références à la vérifiabilité ou à des objectifs partagés.

Tableau 3.5: Catégories d'orientations de hedges selon [Hyland, 1996a, tableau 8 p. 277] en fonction de leur usage par les auteurs scientifiques. Adaptation du tableau d'origine.

[Hyland, 1996a, p. 278] nuance ces classifications en insistant avec des remarques sur la variabilité et le caractère « fonctionnellement indéterminé » des

expressions de hedging en anglais, et sur le fait que « toute recherche d'une caractéristique de surface fonctionnelle et rationnelle doit admettre que [leur] catégorisation est un plus ou moins, plutôt qu'une évaluation tout ou rien ».

3.3.2 Le *boosting* vs l'atténuation

Alors que le *hedging* manifeste un désengagement du locuteur, le *boosting* manifeste un engagement du locuteur. Partant du concept de *mitigation* qu'elle décrit comme « une stratégie utilisée pour réduire l'effet négatif anticipé d'un acte de parole », [Holmes, 1984, p. 345-348 et 359] décrit les « deux stratégies de base pour modifier la force illocutoire » : le *boosting*⁴⁵ comme stratégie d'amplification, et son opposé et complémentaire l'*atténuation* comme stratégie d'affaiblissement. Elle souligne que les expressions employées à ces fins, *boosters* et *downtoners* (ces derniers incluant les *disclaimers* et des *hedges*), peuvent être sémantiquement négatives ou positives. Parmi les catégories de ces expressions, elle nomme entre autres celle exprimant des degrés de croyances, parmi ses exemples : « *I guess it's probably a day-school* » pour le *boosting*, et « *I'm quite sure it's a day-school* » pour l'atténuation.

[Holmes, 1984, p. 348] indique que le locuteur a deux motifs principaux de modification de la force illocutoire dans le contenu propositionnel :

- transmettre la signification modale de son attitude sur le contenu propositionnel, c'est-à-dire exprimer un niveau de son incertitude ou conviction « concernant la validité ou vérité de la proposition », ou concernant l'acceptation ou non d'une responsabilité sur la validité de la proposition. Parmi ses exemples : « *I'm not at all sure [...]* » et « *Certainly [...]* ».
- transmettre des affects ou son attitude sur l'interlocuteur.

[Holmes, 1984, p. 351-362] cite des mécanismes de *boosting* et d'atténuation en particulier que voici, avec quelques-uns de ses exemples⁴⁶ parmi ceux qui correspondent à ce que nous avons trouvé par la suite dans notre corpus scientifique :

- des procédés de *boosting* liés à :
 - la prosodie : « *It's ridiculous* ».
 - la syntaxe : « *Isn't that great!* ».

⁴⁵Le terme *boosting* est gardé pour exprimer le concept en langue française.

⁴⁶Une partie de ses exemples ici rapportés sont attribués d'autres auteurs.

- le lexique :
 - * orienté locuteur : « *in my opinion* » ;
 - * orienté auditeur : « *as you know* » ;
 - * orienté contenu : « *indubitably* ».
- le discours : « *I would emphasize* ».
- des procédés d'atténuation liés à :
 - la prosodie (« such as intonation contour, stress, volume and pitch »).
 - la syntaxe :
 - * les questions : *tags* tel que « *isn't it?* » ;
 - * la double négation : '*not un-...*' tel que dans « *not unlikely* ».
 - le lexique :
 - * orienté locuteur : « *in my opinion* », « *unless I heard it incorrectly* » ;
 - * orienté auditeur : « *you know* », « *if you like* », « *perhaps you could* » ;
 - * orienté contenu et vers les autres : « *reportedly* », « *claim* », « *strictly speaking* », « *rather* ».
 - le discours : « *incidentally* », « *that reminds me* ».

3.3.3 Les opinions et sentiments

La *fouille d'opinion et analyse sentiments* (*Opinion Mining and Sentiment Analysis*) est un champ à part entière du traitement automatique des langues où la recherche foisonne. Les résultats de la recherche dans ce domaine sont en demande croissante de la part d'acteurs des secteurs gouvernementaux, des médias et des affaires, [Sokolova et Lapalme, 2008, p. 320]. Ces résultats ont parfois des implications liées à des enjeux sociétaux qui peuvent être particulièrement importants, par exemple d'un point de vue politique, comme le commentent [Jackiewicz et al., 2010, § 2.3 Impact économique et politique], ou d'un point de vue industriel, comme le commente [Cardey, 2013, chap. 3.3 Sense mining, p. 137].

Contrairement à ce que pourrait éventuellement laisser croire une première approche de ce champ, il aborde des aspects des langues plus diversifiés que la

mesure sémantique d'une 'polarité' ou 'orientation' ou 'préférence' d'un énonciateur, où se centrent cependant et à juste titre de nombreux chercheurs, à l'instar par exemple de [Bhati, 2019] pour l'analyse de sentiment. Ce champ comprend aussi, par exemple, l'analyse du langage figuratif, celui-ci étant entendu comme un détournement figuré ou imagé du sens premier, présent dans l'humour, l'ironie et le sarcasme, ou la métaphore et les jeux de mots, [Benamara *et al.*, 2017, p. 1-4] ; et, autre exemple, l'analyse d'information subjective, celle-ci étant présente dans les « opinions, sentiments, émotions, jugements d'évaluation, etc. », [Fraisie et Paroubek, 2015, p. 1].

[Fraisie et Paroubek, 2015, p. 1-3] considèrent par exemple les concepts de l'*intellectualité* vs l'*affectivité* comme extrêmes d'un axe principal où situer d'autres axes de dimensions perpendiculaires. Ils semblent se rapprocher ainsi des concepts de *dictum* et *modus* de [Bally, 1965] (voir p. 41). Ils constatent que les concepts des émotions, des sentiments et des opinions ne sont pas entièrement séparés conceptuellement. Ils remarquent les avantages comparatifs principaux des modèles de l'information subjective qui utilisant l'étiquetage et ceux utilisant des dimensions (voir tableau 3.6 p. 61). Leur propre modèle propose de faire se rejoindre les étiquetages et les dimensions, en s'appuyant sur la polarité, et en utilisant comme étiquettes des expressions subjectives opposées qui représentent deux à deux les valeurs polaires extrêmes 1 ou 0 de leur dimension. Parmi les neuf dimensions polaires qu'ils utilisent sur l'axe principal *intellectualité-affectivité*, situées vers l'extrême *intellectualité* se trouvent notamment les deux dimensions constitutives d'une catégorie nommée *intellectif*, où l'*intellectualité* tend vers 1 alors que l'*affectivité* tend vers 0, à savoir : ce sont les dimensions *accord-désaccord* et *valorisation-dévalorisation*. Chaque étiquette est elle-même rattachée à un ensemble de classes sémantique, par exemple pour l'étiquette 'dévalorisation' : les classes *dévalorisation*, *désintérêt* et *dépréciation*.

Base de modélisation	Avantages	Inconvénients
Par étiquetage	<i>Compréhension aisée</i>	<i>Problèmes d'interprétation sémantique pour les utilisateurs</i>
Sur des dimensions continues	<i>Redimensionnabilité</i>	<i>difficulté accrue de mise en oeuvre</i>

Tableau 3.6: Avantages comparatifs des types de modélisation des expressions subjectives, reprenant [Fraisie et Paroubek, 2015, p. 1].

[Sokolova et Lapalme, 2008, p. 320-322] proposent de catégoriser les verbes d'opinion suivant plusieurs catégories sémantiques, pour des tâches d'apprentissage automatique (*machine learning*) d'opinions dans des textes de débats politiques et d'avis de consommateurs en anglais. Ces catégories peuvent éventuellement éclairer la nature sémantique des expressions plus complexes d'incertitude qui sont prises en compte dans les règles de notre système. Il s'agit notamment des catégories de verbes non-modaux *cognition* pour des verbes tels que 'believe', 'consider', 'hope', 'think' et 'know' ; *perception* pour des verbes tels que 'feel', 'hear', 'see', 'smell' et 'taste' ; et *attitude* pour des verbes tels que 'enjoy', 'fear', 'like', 'love' et 'hate'. Leurs catégories de verbes modaux sont *primary modal* pour les verbes 'can', 'may', 'will', 'shall', 'have to' et 'must' « signifiant directement la possibilité ou permission ou nécessité d'une action » ; et *secondary modal* 'could', 'might', 'should' et 'would' pour les verbes signifiant « plus poliment, indirectement ou conditionnellement » une tentative ou intention plus hypothétique.

3.3.4 L'ambiguïté

En linguistique, les ambiguïtés textuelles peuvent être *lexicales* lorsqu'un mot réunit plusieurs sens alternatifs, ou bien *structurelles* lorsque c'est une suite de mots qui correspond à plusieurs sens alternatifs, [Trask, 1999, p. 7-8]. C'est le cas par exemple de l'ambiguïté structurelle dans 'Young penguins and seals are impacted by the ecosystem change', où l'adjectif 'young' peut faire référence soit à 'penguins' soit à 'penguins and seals'.

[Fuchs, 1996, p. 49-55], bien que délimitant ses recherches à la langue française, propose entre autres une caractérisation de l'ambiguïté selon trois types de récepteurs :

- l'humain, qui ne parvient pas forcément à une lecture qu'il pourrait relier à sa connaissance du monde : il échoue à traiter des équivoques.
- Le linguiste, qui peut se doter des moyens pour réduire les ambiguïtés, tels que des modèles, des représentations et des niveaux d'analyse : il peut déceler les ambiguïtés effectives.
- La machine, à la fois trop rigide et trop puissante : d'une part elle butte sur des ambiguïtés virtuelles qui peuvent être irrésolubles par manque de connaissance à certains niveaux de traitement, et par manque d'interactions

entre ces niveaux ; et d'autre part ses règles peuvent s'avérer inadaptées et décrire des ambiguïtés qui n'ont pas lieu d'être.

L'ambiguïté en TALN

[Yvon, 2007] décrit les difficultés du TALN à travers deux enjeux où des progrès doivent être accomplis : l'*ambiguïté* et l'*implicite*. Le problème de l'ambiguïté concerne d'un côté la reconnaissance, interprétation et compréhension des entités linguistiques ; et d'un autre côté les choix à faire pour limiter les ambiguïtés lors de la génération de phrases. Le problème de l'implicite relève du contexte d'interaction et de la connaissance du monde ; le besoin de précision des non-dits augmentant d'autant plus qu'une communication naturelle doit s'engager avec un humain. Yvon insiste sur le fait que l'ambiguïté caractérise les divers niveaux du langage naturel ; et il remarque, ce qui peut se traduire par une réduction d'ambiguïté en amont, qu'aborder des sous-domaines du langage naturel peut-être une stratégie efficace de réduction des ambiguïtés :

« Le langage naturel est ambigu, et ce à quelque niveau qu'on l'appréhende. Cette ambiguïté, loin d'être marginale, est un de ses traits caractéristiques. [...] Dès lors, en effet, que l'on restreint le cadre des textes analysés à un sous-domaine particulier [...], il devient possible d'une part d'ignorer un grand nombre d'ambiguïtés [...] ; et d'autre part de représenter formellement un grand nombre des connaissances nécessaires à la compréhension des énoncés du domaine considéré ».

[Jean *et al.*, 2017, p. 1 et 3] considèrent que l'ambiguïté linguistique constitue un premier niveau d'incertitude du langage naturel, auquel « l'imprécision due aux formulations » peut être ajoutée par des expressions telles que 'il semble que'. Ils font correspondre ce niveau de l'ambiguïté aux différentes correspondances possibles entre « des mots ou des suites de mots » et leurs significations, et en situent la source principalement dans la polysémie et l'homonymie. Ils indiquent que ce sont principalement les méthodes de reconnaissance d'entités nommées (REN) qui sont utilisées pour traiter l'ambiguïté. Le deuxième niveau qu'ils nomment « L'imprécision (le flou) » est celui qui correspond à notre recherche : « il exprime un savoir incomplet sur des faits ou des événements », par exemple avec le mot 'plusieurs' ; et il correspond à des « marqueurs de certitude » qui peuvent être dissociés par exemple en tenant compte de degrés d'incertitude.

[Vernier *et al.*, 2009, p. 3 et 6], dans une recherche visant à catégoriser la subjectivité de textes journalistiques et parlementaires en français, se servent de termes *axiologiques*. Ces termes sont liés à l'expression « de préférence et de rupture de l'indifférence » dans les jugements de valeur liés à l'expérience humaine, tels que par exemple le nom 'élégance' ou le verbe 'plaire' ou leurs contraires. La dimension de « valeur axiologique » sur laquelle, parmi d'autres dimensions, ils s'appuient pour catégoriser des segments de phrases, se compose des degrés alternatifs 'positive', 'négative' et 'ambigüe'. [Vernier, 2011, p. 42] précise que le terme *axiologie* est lié aux champs de la morale et de l'éthique, de la pragmatique (par exemple sur l'axe des termes 'important' vs 'dérisoire'), du cognitif et de l'intellectuel, de l'esthétique, ainsi que de l'hédonique et de l'affectif, ce qui est éloigné du champs de l'épistémique que nous avons retenu. Cependant, il indique des « grandes catégories » de marqueurs axiologiques qui, plus que les marqueurs qu'elles regroupent, peuvent rejoindre notre recherche, à savoir : marqueurs lexicaux « en liaison parfois avec les registres de langue », suffixes, termes objectifs (plutôt termes non-objectifs, pour notre recherche), et figures de style « en particulier la métaphore et la comparaison ».

3.3.5 L'évidentialité

L'*évidentialité* (*evidentiality*), dite aussi *médiativité* ou *médiation* ou *médiatif*⁴⁷, est l'expression de la source d'acquisition ou du mode d'accès à une connaissance. [Aikhenvald, 2004, p. 3, 333] précise : « sans nécessairement » impliquer un degré de certitude ou de vérité. Cependant, nombre d'auteurs utilisent une prise en compte de cette implication.

[Mélac, 2014, p. 31] signale que « il existe peu de consensus sur l'étendue et le domaine d'application de cette notion, même parmi les auteurs de référence ». [Aikhenvald, 2004, p. 5] remarque une confusion terminologique due à la non-prise en compte de la polysémie des termes anglais 'evidence' et 'evidential'. [Nuyts, 2017, p. 58] commente que des dimensions maintenant « plus ou moins traditionnelles » y ont été graduellement ajoutées (voir ses catégories, ci-après).

⁴⁷[Dendale et Tasmowski, 2001, p. 340-341] expliquent d'une part que le terme *médiatif* a été préféré par des auteurs s'exprimant en langue française ; mais que d'autre part, en langue anglaise, la signification du terme *mediativity* peut être distinguée de celle du terme *evidentiality* tout en posant « quelques problèmes dans la recherche actuelle » sur ces deux concepts. De plus, [Mélac, 2014, p. 34] signale que c'est le terme *évidentialité* qui est le plus en usage en français dans les travaux tournés vers la langue anglaise. C'est donc le terme *évidentialité* que nous avons choisi d'utiliser.

Dans un sens large, pour [Chafe, 1986, p. 262] et par la suite d'autres auteurs tels que [Mushin, 2001, p. 23-33], l'évidentialité concerne l'attitude envers la connaissance. Selon l'ouverture plus ou moins large du concept par les auteurs, les (*marqueurs*) *évidentiels* peuvent exprimer notamment, avec des degrés de précision de l'information et avec éventuellement des degrés de distanciation de l'énonciateur : le rattachement à un raisonnement (ce qui rejoint plus particulièrement notre méthodologie) tel qu'une inférence logique, la simple supposition liée à la connaissance générale ou à l'intuition, l'attribution à un auteur cité ou non cité, la perception à travers le ressenti général ou des sens en particulier tels que la vue ou l'audition ou l'odorat, l'évocation mémorielle, etc.

Des marqueurs évidentiels en langue anglaise sont par exemple, [Kronning, 2003, table 1 p. 133], 'I see' pour la perception visuelle, 'I feel' pour la perception non visuelle, 'must' pour l'inférence et 'have been said to' pour l'information empruntée. [Willett, 1988, chap. 2 p. 56-63, et annexes A et B p. 96-97] présente une synthèse structurée de catégories qui correspondent à des marqueurs évidentiels étudiés dans 38 langues, dont nous reprenons les notions quasi à l'identique ici :

- **évidence directe :**

- attestée :

- * évidence visuelle → vu.
 - * évidence auditive → entendu.
 - * évidence sensorielle → ressentit physiquement.

- **évidence indirecte :**

- rapportée :

- * évidence de deuxième main → témoignage reçu d'un témoin direct.
 - * évidence de troisième main → témoignage reçu d'un témoin indirect.
 - * évidence à partir de folklore → fondée sur l'histoire orale.

- inférée :

- * inférence à partir de résultats → observable, issue de la perception ou d'une causalité.

3.3. TYPOLOGIE CONCEPTUELLE DE L'INCERTITUDE

- * inférence à partir de raisonnement → construction mentale, basée sur la logique, l'expérience, l'intuition, un rêve, ou autre construction mentale.

Le nom, le nombre et la hiérarchie des catégories de marqueurs de l'évidentialité sont fréquemment exprimés avec des originalités ou des divergences selon les auteurs, et selon les langues étudiées. Cependant, un tronc commun de catégorisations plus communes se remarque. [Nuyts, 2017, p. 66-68], s'appuyant sur [Willett, 1988], présente les « catégories évidentielles classiques » avec ses exemples correspondants en langue anglaise :

- (Evidence directe) *Experimenté* : ex. : « A: John's had a severe car accident today. – B: How do you know? – A: I've *seen* it happen myself! »
- (Evidence indirecte) *Inféré* : ex. : « John *appears* to be home. [His car is in front of the door and the lights are on.] »
- (Evidence indirecte) *Oui-dire* : ex. : « I *hear* you've got a tuition waiver at Stanford, congratulations! »

[Chafe, 1986, p. 262] trouve en langue anglaise une proportion de 64 marqueurs évidentiels pour 1000 mots dans un corpus de textes de publications universitaires, et il remarque des différences de fréquences des types de marqueurs par rapport à un corpus de conversations. [Chafe, 1986, p. 264-265] explique que, lorsque les énonciateurs utilisent des expressions de degré de confiance en évaluant la connaissance, des différences apparaissent en anglais entre d'une part des publications universitaires écrites et d'autre part des expressions utilisées dans des conversations orales, ses exemples : aux adverbes 'maybe', 'probably' et 'certainly' déjà présents dans les conversations orales s'ajoutent 'possibly', 'undoubtedly', 'surely' et 'perhaps' à l'écrit.

Cependant, des marqueurs peuvent aussi être employés avec des sens non-évidentiels. C'est ce que montrent [Grossmann et Tutin, 2010, p. 288-290] à partir d'un corpus de textes scientifiques de langue française, avec des exemples d'emploi du verbe *voir* dans les sens d'*examiner*, de *juger* et d'*auxiliaire être*, dont l'emploi en langue anglaise est parfois équivalent : 'Voyons les sèmes que le contexte...'/ 'Let us look at the semantic features that...', 'Les cotisations du système par répartition sont souvent vues comme des prélèvements obligatoires...'/ 'Contributions paid into a redistribution system are often seen as obligatory deductions...', 'La théorie des opérations énonciatives se voit intimement

associée à une définition du langage...’/’The theory of enunciative operations is intimately associated with a definition of language...’, etc.

Les marqueurs évidentiels ont pour fonction d’indiquer comment la connaissance est acquise et comment en catégoriser la source, [Aikhenvald, 2004, p. 365]. [Kronning, 2003, p. 145-146] évoque d’autres fonctions possibles de l’évidentialité, notamment, d’une part reprenant [Desclés et Guentchéva, 2000], la fonction de désengagement vis-à-vis de faits énoncés ; et d’autre part la fonction de permettre l’évaluation d’un énoncé. A propos de la fonction de permettre l’évaluation, il donne en exemple l’utilisation du passé composé dans ‘il a pleuré’, dont la finalité est que celui qui interprète l’énoncé puisse comprendre que cet énoncé est la conséquence d’un raisonnement abductif, et puisse reconstruire ce raisonnement « Si l’on pleure, les yeux deviennent rouges. Or, le concierge a les yeux rouges. Donc, il a pleuré », en examiner la véracité des prémisses, et considérer éventuellement d’autres causes vraisemblables.

L’évidentialité peut être, selon les cas, unie avec les modalités. La forte ressemblance entre les deux concepts fait que certains auteurs les traitent comme une même catégorie, [López Ferrero, 2005, p. 123]. Cependant, [Dendale et Tasmowski, 2001, p. 341-343], les opinions divergent sur les relations entre ces concepts. [Saussure (de), 2012, p. 133-137] montre que le rapprochement avec les modalités épistémiques, bien que pas toujours évident, apparaît de manière logique dans certains cas ; et il désigne « quatre types généraux de relation » entre la modalité et l’évidentialité : absence de relation, disjonction, inclusion et chevauchement.

3.3.6 L’éllicitation

L’éllicitation, [Dias *et al.*, 2018, p. 1], est ce qui rend possible ou facilite « l’expression quantitative d’un jugement subjectif » à propos de faits ou de valeurs. C’est un « processus structuré » sur des « connaissances qui ne peuvent être facilement formalisées autrement », [Sluijs (van der) *et al.*, 2006a, p. 12].

Dans le sens d’*Expert Elicitation* qui est celui considéré ici, l’éllicitation est définie et mise en œuvre dans un cadre scientifique. Dans la bibliographie consultée, les méthodes décrites nécessitent que les jugements utilisés soient ceux d’experts humains, et elles font une place importante à leurs échanges notamment en langue naturelle, en particulier : dans la forme des questions aux experts, dans le matériel de prise de notes fourni à chaque expert, dans le recueil d’informations complémentaires auprès des experts, et dans les échanges entre ces experts.

3.3. TYPOLOGIE CONCEPTUELLE DE L'INCERTITUDE

[Gosling, 2018, p. 61-64 et 78-80 et 82] insiste sur le fait que le processus de « capture » du jugement des experts et les résultats de l'élicitation doivent être transparents et défendables, notamment sur les raisonnements des experts qui doivent pouvoir être révisés par ces derniers. Il précise que c'est l'ensemble de formulaires d'enregistrement et guidage utilisés par l'équipe de facilitateurs de l'élicitation qui est l'outil le plus important de l'infrastructure logicielle Sheffield Elicitation Framework (SHELF). Cette infrastructure étant « largement appliquée en économie de la santé et en médecine », mais aussi appliquée par exemple dans les sciences environnementales dû à leur augmentation des analyses de l'incertitude. Parmi les extensions et adaptations de SHELF, il signale l'intérêt pour « capturer plus d'information qualitative sur les incertitudes qui ne peuvent être quantifiées ».

L'élicitation aide à traiter des problèmes d'incertitude sur des grandeurs ou des quantités dont la mesure ne peut être précise, et ses domaines d'application sont diversifiés. Par exemple, [Bastin *et al.*, 2011] commentent que l'expertise sous forme d'élicitation, « bien étudiée en statistiques et en psychologie », peut être utilisée pour acquérir des entrées ou des paramètres sur des quantités incertaines pour la construction de modèles environnementaux. D'autres commentent son utilisation, par exemple, [Dias *et al.*, 2018, p. 5-7], pour la prise de décisions dans des domaines tels que le lancement de campagnes de vaccination lors d'épidémies, la communication sur les risques de tremblements de terre, et la localisation d'aéroports ou de dépôts de déchets radioactifs ; ainsi que, [Hora, 2016, p. 498], pour l'estimation de la relation risques-bénéfices liée à l'exposition à certains types de radiations d'usage médical, ou liée à la réduction de vitesse routière ; et aussi, [Usher et Strachan, 2013, p. 1], pour son utilisation dans les techniques de mesure d'incertitude pour la modélisation des politiques sur l'énergie.

Le Modèle Classique d'élicitation, pour l'analyse d'incertitude quantitative, [Quigley *et al.*, 2018, p. 15-19, 23 et 32-33], utilise l'agrégation mathématique de mesures dont les valeurs subjectives sont issues des « jugements de multiples experts », en pondérant ces valeurs par la précision (calibration à un endroit sur l'échelle de valeurs employée, qui doit être corrigée pour chaque expert) et l'informativité (précision, avec le choix par l'expert d'intervalles les moins larges possibles) statistiques attribuées à chaque expert ; et, de plus, la dimension qualitative de chaque jugement doit être recueillie à partir d'explications de chaque expert sur son raisonnement et les fondements de son raisonnement.

À l'interface des jugements et des valeurs numériques, certaines tech-

niques d'élicitation font usage d'expressions préétablies en langue naturelle pour représenter et sélectionner des intervalles ou des niveaux, ce qui rejoint ou s'apparente aux techniques d'organisations qui utilisent des échelles d'incertitude (voir la prise en compte de l'incertitude par des organisations, p. 10) : par exemple, pour l'élicitation de probabilités, [Nunes *et al.*, 2018, p. 6-8] décrivent l'usage d'expressions en langue anglaise qui sont mises en correspondance avec des valeurs numériques sur des échelles de probabilité : 'certain'-100, 'probable'-85, 'expected'-75, 'fifty-fifty'-50, 'uncertain'-25, 'improbable'-15 et 'impossible'-0; ce qui « permet aux experts du domaine de penser en termes de proportions visuelles plutôt qu'en termes de nombres précis ». [Nunes *et al.*, 2018, p. 15] décrivent aussi une échelle de 9 niveaux de probabilités comparées par paires d'événements, soit du niveau 1 à 9 : 'Equal likely', 'Weak or slight', 'Moderate more likely', 'Moderate plus', 'Strong more likely', 'Strong plus', 'Very strong more likely', 'Very, very strong' et 'Extremely more likely'.

[Hora, 1996, p. 217-218] souligne le fait que dans le processus d'élicitation une distinction est parfois faite et est parfois nécessaire entre les incertitudes, lesquelles sont alors considérées soit épistémiques soit aléatoires :

- l'*incertitude épistémique* relève d'un « manque de connaissances » : celle-ci est résoluble « en principe » par l'étude, et peut éventuellement être réduite grâce au jugement d'experts ;
- l'*incertitude aléatoire* relève d'une « variation imprévisible » : celle-là est irréductible.

Il signale cependant que ces deux types ne sont généralement pas nettement distincts, et il présente un exemple de jugement d'élicitation en hydrologie pour montrer que cette distinction *épistémique* vs *aléatoire* dépend du choix de l'échelle de ce qui est considéré dans le système étudié. Ceci est une indication que les expressions textuelles de ces incertitudes nommées *aléatoires* peuvent être correctement prises en compte sous forme de catégorie ontologique différenciée par notre système, notamment à la catégorie '*i*' qui est la moins spécifique de l'ontologie de l'incertitude (voir schéma de l'arbre des incertitudes 6.3, p. 135).

3.3.7 La prise en charge

Selon [Mélac, 2014, p. 56-60], « La prise en charge se décline en cinq types : totale, réduite, personnelle, nulle et négative ». Suivant [Desclés, 2009, p. 29-30,

34-35, 39 dont nbp 25, et 40-44], il y a toujours une prise en charge de l'énoncé, de son contenu, via les choix d'un énonciateur : par l'engagement ou le désengagement personnel ou reporté sur autrui, ou par des jugements épistémiques ou aléthiques. « *l'assertion est la prise en charge de la vérité* d'une proposition [...], ce qui revient à prendre en charge à la fois un contenu propositionnel et à affirmer sa vérité ». Il considère que les « opérations énonciatives », qui laissent des traces telles que, ses exemples, 'pour être plus exact' ou 'en explicitant ce que je veux dire', « mettent plutôt en jeu des opérateurs d'ajustements dialogiques qui ont pour opérands des segments linguistiques déjà énoncés ou des conditions contextuelles extra-linguistiques ». Il distingue la « simple prise en charge d'une relation prédicative [...] par un énoncé déclaratif » de son assertion qui « peut être indiquée par des marqueurs explicites comme *j'affirme, c'est certain que, il est vrai que, ...* ». Par exemple, il distingue les simples déclarations 'Il pleut' ou 'la conjecture de Fermat n'est pas un théorème' des assertions 'J'affirme qu'il pleut' ou 'tu as tort lorsque tu dis que la conjecture de Fermat n'est pas un théorème' ; bien qu'il qualifie la « distinction *déclaration/assertion* » de « dans certains cas [...] pas évidente ». Il présente une « sorte d'ontologie linguistique » d'assertions, dont celle de type 'Assertion soumise à condition', comme dans '..., à condition toutefois que...'. Il aussi indique qu'il existe des énonciations dites *médiatives* ou *médiatisées* qui « prennent en charge la plausibilité d'une proposition », en constatant certains indices déclencheurs d'un raisonnement abductif qui « remonte vers l'une des causes possibles » et conduit vers une hypothèse encore négociable : « Il aurait donc plu [abduction : « puisque la chaussée est mouillée »] ».

3.3.8 La spéculation

[Cruz Díaz et Maña López, 2019, p. 27] relie le concept de *langue spéculative* à entre autres la subjectivité, à l'évidentialité, à l'incertitude, à la croyance engagée (*committed belief*) et au factuel, dans le cadre de la modalité qu'ils rattachent aux expressions de croyance, d'attitude et d'obligation qu'utilise un locuteur dans ses déclarations.

[Diab *et al.*, 2009, p. 69] relie la *croyance engagée* aux croyances et intentions, à ce que le texte veut faire croire sur les états cognitifs de personnes, dont le locuteur. Ils donnent des exemples de catégorie 'non-committed belief' avec des expressions comportant des auxiliaires modaux (« *GM may lay off workers* ») et des discours rapportés (« *The politician claimed that ...* »), et de catégorie 'non applicable' telle que pour des propositions qui n'ont pas valeur de vérité avec des

expressions comportant du désir (« *Some wish GM would lay off workers* »), des questions (« *Many wonder if ...* ») et des demandes (« *GM is required to lay off workers* »). [Prabhakaran *et al.*, 2010, p. 1014] associent la croyance engagée aux propositions du texte auxquelles l'auteur croit.

3.4 La notion de subjectivité

La pratique scientifique implique l'usage de leur liberté par les scientifiques.

Du point de vue épistémologique, [Kuhn, 1970, p. 186] remarque que l'avancement de la science peut s'appuyer à la fois sur des normes communes et sur la variation individuelle de l'activité scientifique dans l'usage de ces normes :

« les valeurs partagées peuvent être des déterminants importants du comportement de groupe, bien que tous les membres du groupe ne les appliquent pas de la même manière. [...] la variabilité individuelle dans l'application de valeurs partagées peut servir à des fonctions essentielles à la science »⁴⁸.

Ce point de vue épistémologique se prolonge chez [Barreau, 2013, p. 12-13] lorsqu'il affirme ce qui permet la continuité historique de la science et la croissance des sciences :

« les schémas de la connaissance commune ne cessent pas d'habiter l'esprit des scientifiques [...] dans leur travail de recherche », et « Il est bien certain qu'un esprit inventif ne doit pas sans cesse faire la police de son esprit ».

En ce sens, la subjectivité est donc constituante de l'activité scientifique, et ne se limite pas à la perception et compréhension de leurs objets de recherche par les scientifiques.

La subjectivité est aussi présente dans la communication scientifique, entre autres dans les textes des documents qui recueillent et transmettent les savoirs, réflexions et avancées des domaines scientifiques. C'est le cas lorsque l'auteur d'un texte scientifique cherche à persuader du bien fondé de son raisonnement par

⁴⁸Traduction de « shared values can be important determinants of group behavior even though the members of the group do not all apply them in the same way. [...] individual variability in the application of shared values may serve functions essential to science »

3.4. LA NOTION DE SUBJECTIVITÉ

des choix linguistiques. Sur ce point, [Latour et Woolgar, 1986, p. 86-88] insistent sur le fait que l'activité normale d'un laboratoire scientifique passe par une essentielle remise en scène scientifique des informations⁴⁹. Ces choix relèvent de la rhétorique décrite par Aristote : choix des moyens d'influence sur l'émotion (*pathos*), par exemple sur l'aspect utile de sa recherche ; choix des moyens agissant sur la perception de sa crédibilité (*ethos*), par exemple par la connexion bibliographique ; et choix des moyens mettant en avant l'évidence de sa logique (*logos*), par exemple avec l'emploi d'une notation mathématique.

[Bachimont, 2007, p. 9-13 et 243-245] souligne que l'utilisation des machines comme médiatrices pour traiter automatiquement des connaissances n'élimine pas cette subjectivité : l'utilisateur de leurs connaissances doit les interpréter culturellement « en abordant le réel à travers la médiation des représentations construites par les outils [...] » ; « une improvisation du sens, où le lecteur s'approprie de manière singulière l'inscription qu'il consulte [et qui est] ainsi à la fois un objet matériel et un objet intentionnel » où « dans la plupart des cas la relation entre le sens de l'expression à sa forme reste arbitraire », « car seuls les êtres humains peuvent assumer une interprétation dans toute sa complexité ».

3.4.1 Définition générale de la subjectivité

Dans cette section, nous nous intéressons à la notion de subjectivité au sens général pour nous rapprocher ensuite de ses sens en sciences.

L'adjectif *subjectif* a pour origine étymologique le latin *subjectus*⁵⁰, c'est-à-dire *placé dessous*. Du point de vue étymologique, les premières traces du mot *subjectivité* renvoient vers ses emplois comme dérivé de *subjectif*, peut-être sous l'influence de l'allemand *Subjektivität*⁵¹ dans l'ouvrage [Villers (de), 1801, p. ex. : p. 113 et 277] sur la philosophie de Kant :

« Je multiplie à dessin les exemples, pour offrir au lecteur, sous toutes les formes, cette distinction nécessaire de la subjectivité et de l'objectivité dans la connaissance que nous prenons des choses »,

« Le premier résultat de nos recherches transcendantales est donc la subjectivité et l'à *priorité* pure de l'espace et du temps ; ils sont les

⁴⁹Notamment : « A laboratory is constantly performing operations on statements; adding modalities, citing, enhancing, diminishing, borrowing, and proposing new combinations ».

⁵⁰Nouveau Larousse Universel, édition 1948.

⁵¹CNRTL, consulté le 22/01/2018.

deux formes originaires et virtuelles de notre sensibilité ; ils sont les produits de notre *sensorium*, comme les couleurs sont les produits de l'œil, les sons celui de l'oreille, les saveurs celui du palais ».

Du point de vue de la langue courante, entre autres dans les cultures anglaise et française, la subjectivité existe par l'intériorité du sujet, en contraste avec la réalité extérieure :

« The quality of being based on or influenced by personal feelings, tastes, or opinions. [...] The quality of existing in someone's mind rather than the external world. » ⁵².

« Caractère de ce qui est subjectif, par opposition à objectivité. / État de quelqu'un qui considère la réalité à travers ses seuls états de conscience » ⁵³

L'Académie française [[Académie française, 1935](#)] place le terme de *subjectivité* dans le champ de la philosophie pour signifier le « Caractère de ce qui est subjectif ». Le [Nouveau Larousse Universel] ⁵⁴ décrit la subjectivité comme :

« ce qui se rapporte au sujet pensant, par opposition à objectif qui se rapporte à l'objet pensé. Grandeur subjective ou psychologique, inexprimable avec précision ».

Dans les sciences, la notion de *subjectivité* est souvent abordée par ses liens avec le concept d'*incertitude* ; lequel est reprécisé par le champ scientifique qui l'utilise. Par ailleurs, dans les différentes sciences le sens de *subjectif* et *subjectivité* peut varier. Par exemple, en médecine les symptômes subjectifs sont ceux qui ne sont perçus que par les malades. Concernant la prévention médicale, [[Deccache, 1999](#), p. 32] indique que la « subjectivité de la santé » est « une "construction sociale" et non seulement un produit "scientifique" objectif ».

3.4.2 La notion de subjectivité en philosophie

Au-delà de l'apparition du mot *subjectivité* retracée chez des philosophes (cité ci-dessus), la notion de subjectivité est abordée en philosophie depuis plusieurs siècles. Ainsi, selon [[Assanzadeh et al., 2011](#), p. 10] :

⁵²Oxford Dictionaries, version anglaise, consulté le 24/21/2018

⁵³Larousse « Dictionnaire de français », édition en ligne, consulté le 24/21/2018.

⁵⁴Nouveau Larousse Universel, édition 1948.

3.4. LA NOTION DE SUBJECTIVITÉ

« Carnéade (219 av. J.C., 128 av. J.C.), philosophe sceptique, remet en cause la connaissance acquise même par la raison. Car, il y a des représentations fausses qui ne nous permettent pas une connaissance certaine. Cette critique de la certitude a comme conséquence une suspension de jugement en ne croyant à rien (Cicero). Ainsi donc émerge la notion de subjectivité avec la mise en évidence du rôle de nos représentations dans nos jugements ».

La philosophie conçoit la notion de *subjectivité* en opposition à son antonyme *objectivité*, pour exprimer⁵⁵ une « Qualité (inconsciente ou intérieure) de ce qui appartient seulement au sujet pensant » ou pour exprimer des « réalités subjectives » avec « Le moi, la conscience ». L'opposée *objectivité* y est⁵⁶ la qualité de ce qui a rapport avec l'objet, ou de ce qui est dans l'objet : réalité objective. Est *objectif* (adjectif) ce qui est indépendant de toute préférence individuelle : récit objectif.

3.4.3 La notion de subjectivité en linguistique

L'une des définitions de la *subjectivité* (voir aussi des définitions en annexe A.5, p. 218), celle la plus communément rencontrée et acceptée dans les sources (articles et ouvrages consultés) en linguistique et sciences sociales, peut se résumer à : *ce qui est en relation avec le sujet énonciateur*.

[Benveniste, 1958], n'imagine pas le langage sans la subjectivité :

« le langage [...] est marqué si profondément par l'expression de la subjectivité qu'on se demande si, autrement construit, il pourrait encore fonctionner et s'appeler langage ».

Pour lui, « le discours est le langage mis en action » et la subjectivité dans le langage est « l'émergence dans l'être d'une propriété fondamentale du langage » : le langage propose des formes linguistiques vides ; et le locuteur se les approprie par les « instances discrètes » qui constituent le discours, provoquant ainsi l'émergence de la subjectivité par les faits des langues particulières.

Il remarque que la subjectivité a plus d'une dimension, dont il souligne celle de l'*ego* dans l'intersubjectivité, « la capacité du locuteur à se poser comme « sujet » », et celle de la temporalité, « organisation langagière de la notion de temps » :

⁵⁵CNRTL, consulté le 23/01/2018.

⁵⁶Inspiré du dictionnaire Nouveau Larousse Universel, édition 1948

- « la condition d'*intersubjectivité* [...] seule rend possible la communication linguistique » : l'*intersubjectivité* est à la fois la *subjectivité* de l'*ego* qui s'exprime et de l'autre, complémentaire, qui l'interprète.
- « le domaine de la *subjectivité* [...] doit s'annexer l'expression de la temporalité. [...] Le temps linguistique est *sui-référentiel*. En dernière analyse la temporalité humaine avec tout son appareil linguistique dévoile la *subjectivité* inhérente à l'exercice même du langage ».

Certains auteurs, [Le Querler, 1996, p. 55], distinguent des *modalités temporelles*, telles que celles exprimant des nuances de 'toujours' et 'parfois', dans leur classement des modalités en linguistique.

Le traitement automatique des aspects de la temporalité constitue un problème d'actualité. [Pustejovsky *et al.*, 2003, p. 1-2, 4, 11] décrivent le système TimeML qui sert à l'annotation d'événements temporels dans un document, et permet de classer des expressions temporelles dont la certitude est limitée à leurs relations logiques le texte. Ils précisent que Time-ML fut conçu pour répondre au problème du raisonnement avec des expressions temporelles partiellement spécifiées (*underspecified temporal expressions* ou *partially determined temporal expressions*, telles que *last week*, *two weeks before*, *monday*, *next month*, *last year*, *two days ago*) qu'ils qualifient de fonctions temporelles. L'interprétation de ces expressions-fonctions temporelles, qui peut être retardée dans le déroulement du texte, permet de les ordonner les unes par rapport aux autres :

« TimeML systematically anchors event predicates to a broad range of temporally denotating expressions; it provides a language for ordering event expressions in text relative to one another [...]; and it provides a semantics for underspecified temporal expressions, thereby allowing for a delayed interpretation ».

Cependant, [Lefeuve *et al.*, 2014, § 2, § 4.1, § 5] soulignent que la norme ISO-TimeML basée sur TimeML⁵⁷, bien qu'elle serve déjà à annoter des corpus de diverses langues, a des défis à résoudre comme par exemple le repérage des entités imbriquées ; et qu'elle montre des insuffisances dans la pratique pour certaines tâches comme par exemple « du fait de la juxtaposition d'annotations

⁵⁷Le dialecte du XML *TimeML* est à la base du cadre d'annotation sémantique du standard terminologique *ISO 24617-1:2009 SemAF-Time (Semantic annotation framework)* ; lequel spécifie ISO-TimeML pour extraire et représenter de l'information temporelle. Références à ce standard international sur : <https://www.iso.org/standard/37331.html>

3.4. LA NOTION DE SUBJECTIVITÉ

de niveaux linguistiques différents »). La norme doit être encore améliorée pour correspondre notamment aux besoins du traitement automatique des langues.

Selon les points de vue envisagés en linguistique, la conception de la subjectivité est liée en particulier à d'autres notions mises en avant. Par exemple, pour [Dubois *et al.*, 2007] :

« On appelle subjectivité la présence du sujet parlant dans son discours ; ainsi, la subjectivité du discours se manifeste par les embrayeurs ». « Les embrayeurs sont une classe de mots dont le sens varie avec la situation [...] Par exemple, je , hier , ici [...] ».

Pour la recherche que nous décrivons ici, la subjectivité est étroitement liée à la notion de *modalité*. Par exemple, [Saurí *et al.*, 2006], qui utilisent la norme TimeML en s'intéressant aux événements dans les textes anglais, emploient le terme *modalité* pour exprimer « le degré d'engagement » de l'énonciateur et le caractère factuel associé aux événements.

Chapitre 4

La prise en compte de l'incertitude dans le domaine du Traitement Automatique des Langues

4.1 Dimensions des expressions de l'incertitude

Dans cette section nous abordons quelques dimensions structurantes envisagées dans le traitement automatique des expressions de l'incertitude.

En traitement automatique des langues, bien que cela ne soit pas exclusivement réservé aux expressions de l'incertitude, ces dernières sont considérées sur des dimensions de degrés (ex. : échelles), polarité (affirmation *vs* négation), portée (*scope*), et diverses classes (valeur ponctuelle, intervalle, temporalité ...), mais pas encore en catégories de raisonnements scientifiques.

4.1.1 Des dimensions des modalités

[Saurí *et al.*, 2006] explorent l'étendue des modalités dans le cadre de la langue anglaise, notamment dans le cadre de l'extraction d'événements. Dans ce cadre, la modalité traitée est ce qui exprime « le degré d'engagement de l'orateur envers les événements auxquels un texte se réfère »¹ et son traitement inclut l'aspect factuel des événements. Ils proposent un axe de véridicité pour la caractérisation des types modaux entre deux extrêmes, depuis *véritablement factuel* (*truly fac-*

¹Traduction de « the speaker's degree of commitment to the events being referred to in a text ». [Saurí *et al.*, 2006, p. 1]

4.1. DIMENSIONS DES EXPRESSIONS DE L'INCERTITUDE

tual) jusqu'à contre-factuel (counter-factual) : 'degré de possibilité', 'croyance', 'évidence', 'attente', 'essai', 'commandement'. (voir tableau 4.1, p. 78).

Type modal	Phrase d'exemple
Degrees of possibility	<i>These results indicate that Pb2+ may inhibit neurite initiation by inappropriately stimulating protein phosphorylation by CaM kinase.</i>
Belief	<i>Chinese analysts believe that the United States will continue to provoke North Korea.</i>
Evidentiality	<i>Subcomandante Marcos said that the Mexican government is not interested in putting an end to the conflict.</i>
Expectation	<i>Hans Blix wants the US to allow UN inspectors back into Iraq to verify any weapons found by coalition forces.</i>
Attempting	<i>George Mallory and Andrew Irvine first attempted to climb Everest in 1924.</i>
Command	<i>John Murtha called for the immediate withdrawal of U.S. troops from Iraq.</i>

Tableau 4.1: Axe de Sauri-Verhagen-Pustejovsky pour caractériser la véridicité d'évènements (veridicality axis) par des types modaux.

4.1.2 La polarité

La polarité dans le discours peut exprimer deux phénomènes aux significations alternatives : soit elle affirme, soit elle nie. Il pourrait sembler a priori que ces deux aspects s'excluraient toujours l'un l'autre, et s'opposeraient dans une dipolarité évidente. Pourtant, la polarité peut se révéler complexe à clarifier et à décider, d'autant plus lorsqu'elle interagit² avec les formes de l'incertitude qui s'expriment à l'aide de concepts intermédiaires et de transition.

Comme le remarquent [Morante et Sporleder, 2012, p. 223-224], la négation permet de changer la valeur de vérité d'une affirmation³. Ils font remonter jusqu'à

²Interagit au sens où l'expression de la polarité et l'expression de l'incertitude exprimées dans la phrase ne sont plus indépendantes l'une de l'autre et produisent leur effet l'une avec l'autre.

³« Negation is a grammatical category that allows the changing of the truth value of a proposition ».

Aristote la première grande contribution à l'analyse de la négation, et ils notent la diversité des expressions que la littérature scientifique emploie pour faire référence aux mots et affixes exprimant la modalité et la négation : 'déclancheurs' (*triggers*), marqueurs ('*markers*'), et 'signaux' ('*signals*' et '*cues*'). Ils n'en signalent pas moins que la modalité et la négation, qui sont en interaction l'une avec l'autre, sont non seulement encore un défi pour la théorie, mais aussi pour le traitement automatique ; lequel traitement automatique a abordé deux des principaux problèmes, principalement au niveau de la phrase ou du prédicat. Ces deux problèmes sont la détection de plusieurs formes de la négation et de la modalité, et la résolution de la portée (*scope*) des signaux de négation et de modalité. Ils avertissent que la présence dans une phrase d'une spéculation ou d'une négation, par exemple '*this brings us to the largest of all mysteries*' ou '*little was known*', n'est pas déterminable par un simple examen lexical. Ceci d'autant plus que cette présence est en relation de dépendance avec d'autres facteurs du discours, par exemple des *entailments*⁴ ; et d'autant plus qu'elle peut être exprimée par des expressions complexes, composées de plusieurs mots, et interagissantes avec des expressions temporelles (*mood* et *tense*).

4.1.3 Schéma des dimensions de l'incertitude

Ici nous présentons un essai d'un schéma structurant des dimensions de l'incertitude dans les traitements automatiques. La partie encadrée est celle où nous nous situons.

- **Exactitude** de l'information
 - **Fiabilité** de la source
 - **Véracité** de l'information (vraie ou fausse \Rightarrow pertinence)
 - **Commentaires** libres à part portant sur la source et l'information
 - **Contradiction** entre des sources ou entre des commentaires de l'information
 - **Validité** de l'information : type d'informations espéré, informations à jour
- **Moyen de diffusion** de l'incertitude

⁴Dépendances de la valeur de vérité d'une affirmation, envers les valeurs de vérité d'autres affirmations.

- **Expressions** d'incertitude dans le texte ou dans un autre média
- **Métadonnées** liées à l'incertitude dans le texte ou dans un autre média
- **Champ d'usage** du fait lié à son incertitude
 - **Générique** : dans une simulation ou un scénario, une théorie ou un parti pris, des hypothèses ou des suppositions, ...
 - **Générique scientifique** : mesure de calibration d'un instrument, définition d'une notion par simplification, ...
 - * format quantitatif : nombre, intervalle, formule à résoudre
 - * format qualitatif : expression linguistique, énigme à résoudre
 - * format mixte : raisonnement complexe
 - **Spécifiques à une organisation** ou à une norme volontaire : échelles contraintes mettant en correspondance des expressions linguistiques avec des valeurs ou intervalles de valeurs probabilistes numériques.
 - **Spécifique à un champ** de la connaissance : scientifique, militaire, ...

4.2 Corpus *ad-hoc* pour le traitement de l'incertitude

Il existe des corpus scientifiques *ad-hoc*, c'est-à-dire préparés pour certains sujets de recherche en particulier, et parfois mis à disposition de la communauté scientifique. Ce qui fait que ces corpus sont préparés pour le traitement automatique de l'incertitude est l'annotation spécifique de leurs textes à cet effet ; et à cette caractéristique s'ajoute le fait que ces corpus peuvent devenir des corpus de référence sur lesquels il est possible d'essayer différents traitements automatiques afin de comparer leurs résultats. C'est le cas des corpus pour la recherche de l'incertitude dans des textes du biologique et du médical⁵, par exemple ceux cités par [Szarvas *et al.*, 2008, p. 39] ou [Cruz Díaz et Maña López, 2019, p. 53-57] : *BioScope*, *Hedge classification corpus*, *Genia Event corpus*, *BioInfer corpus*, *THYME corpus* sur le cancer du cerveau et sur le cancer du colon, *Fact-Bank corpus* qui concerne des actualités ... Ou, autre exemple, le corpus utilisé par [Malhotra *et al.*, 2013] : le corpus *i2b2 NLP Dataset* (dit aussi *n2c2 NLP*

⁵Le traitement de corpus médical est relativement ancien en traitement automatique des langues, [Friedman *et al.*, 1998].

Research Data Set) de 2010 du *n2c2 2010: Relations Challenge*⁶⁷ du *Fourth i2b2/VA Shared-Task and Workshop Challenges in Natural Language Processing for Clinical Data*.

Selon [Szarvas *et al.*, 2008, p. 38], le corpus *BioScope* a été créé et mis à disposition de la communauté scientifique pour gérer la négation et l'incertitude dans les textes biomédicaux. Il intègre des textes annotés au niveau des entités lexicales (*token*) de négation et de spéculation, et annotés au niveau de leurs portées dans la phrase. Ce corpus est divisé en trois sous-ensembles : des textes libres (*free texts*) de radiologie, des articles complets de biologie, et des résumés (*abstracts*) scientifiques de biologie.

S'agissant de corpus pour la recherche sur l'incertitude, le domaine biomédical est visiblement un domaine moins mal doté (en langue anglaise). Le remarquer nous a induits à choisir pour notre travail la création d'un corpus spécialisé dans un autre domaine scientifique : le changement climatique.

4.3 Méthodes linguistiques pour l'annotation de textes écrits

En choisissant d'analyser la notion d'incertitude dans le domaine scientifique, nous nous sommes intéressés plus particulièrement aux méthodes linguistiques du TAL, par opposition aux méthodes statistiques ou d'apprentissage. Notre choix méthodologique a été inspiré par deux méthodes qui sont : l'approche microsystemique de la langue basée sur les travaux de [Cardey, 2013], et la méthode d'Exploration Contextuelle [Desclés *et al.*, 1997, Desclés, 2006]. Dans cette thèse, nous proposons des règles d'annotation qui sont descriptives et qui découlent d'une analyse linguistique des occurrences dans les corpus à des fins d'annotation de l'incertitude dans les articles scientifiques.

⁶Le corpus *National NLP for Clinical Challenge (n2c2) NLP Research Data Set* de l'édition 2010 - *Relation* du challenge i2b2/VA du Department of Biomedical Informatics (DBMI) de la Harvard Medical School ; challenge portant entre autres sur les « *classification of assertions made on medical problems* ».

⁷Contrairement à ce qui pourrait sembler indiqué sur le site du corpus (page <https://portal.dbmi.hms.harvard.edu/projects/n2c2-nlp/> consultée le 16/08/2020) par l'emploi du terme « unstructured », [Malhotra *et al.*, 2013, § 3] décrit ce corpus comme pré-annoté.

4.3.1 L'approche microsystemique

L'approche microsystemique trouve ses origines dans les travaux de [Gentilhomme, 1979] qui définit ainsi la notion de microsysteme :

« Un microsysteme est un systeme finalisé⁸, suffisamment petit (voir prédégénéréscent) pour être traité avec l'approximation requise, en un temps dit réel, compte tenu de la technique, discursive ou expérimentale, disponible », [Gentilhomme, 1979, p. 2] ; « mais suffisamment grand pour rendre compte de sa finalité », [Gentilhomme, 1985].

L'approche de l'analyse linguistique développée par Sylviane Cardey [Cardey, 1987, Cardey, 2003, Cardey, 2013] au Centre Tesnière, constitue une application de la théorie microsystemique aux langues, où chaque langue naturelle est considérée comme un systeme décomposable en microsystemes robustes et complets. Les microsystemes qui composent la langue (lexique, morphologie et syntaxe) sont interdépendants, ils communiquent et interagissent entre eux pour construire le sens (voir schéma 4.1, p. 83). L'idée de travailler par microsystemes permet de décomposer la langue et, sans traiter l'ensemble des phénomènes linguistiques dans leurs complexité, d'identifier précisément le problème qui doit être traité, puis de construire les systemes, et enfin décrire le problème en utilisant ces systemes afin d'apporter la solution nécessaire, [Cardey, 2013, p. 145]. Il s'agit de préciser les besoins d'une tâche d'analyse textuelle, et ensuite de créer des règles qui répondent à ces besoins et sont applicables pour traiter tous les exemples de la langue. La recherche par l'analyse microsystemique se fait donc *en intention*, c'est à dire en fonction de la finalité. L'analyse d'un systeme linguistique en ses composants, aux niveaux de granularité structurels où se situent et fonctionnent et interagissent de manière organisée les règles nécessaires à une finalité déterminée, [Cardey et Greenfield, 2008, p. 6-21], rend la méthode microsystemique particulièrement apte au développement de logiciels dont les règles linguistiques sont compréhensibles pour des humains : ces règles peuvent alors être évaluées directement par leur lecture, et être éventuellement complétées ou réécrites.

L'approche microsystemique a été appliquée pour répondre à un grand nombre de problématiques en TAL. Nous citons ici quelques unes des applications sur le traitement de textes écrits :

⁸ « doté d'une finalité externe », [Gentilhomme, 1979, Compendium, p. 1].

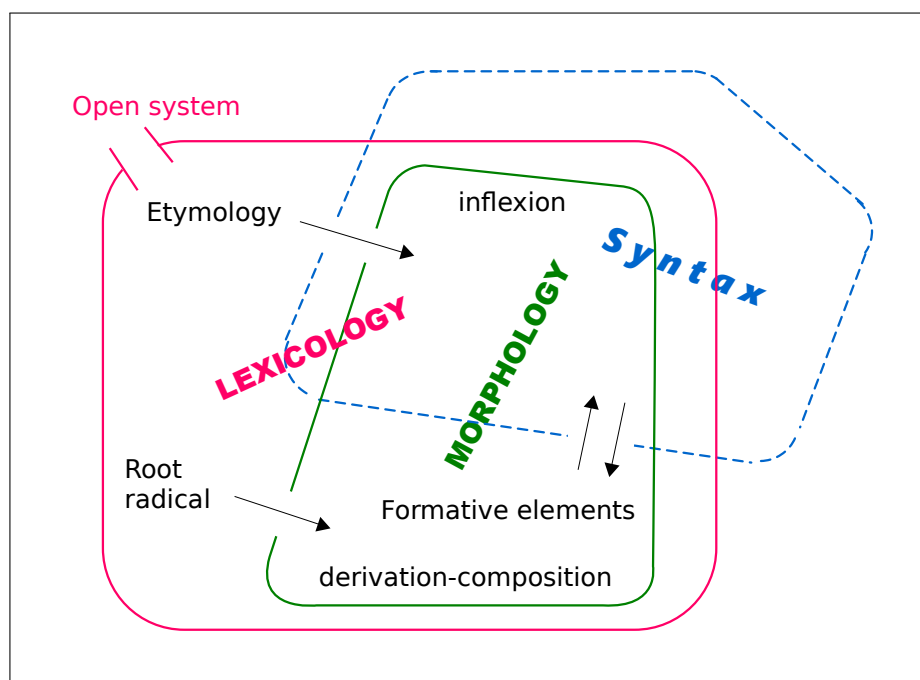


Schéma 4.1: L'approche microsystemique de S. Cardey (2003, 2013) selon [Cardey, 2013, fig. 7 *French lexico-morpho-syntactic system* p. 39].

- Le système Labelgram [Cardey *et al.*, 1997, Cardey et Greenfield, 2003], qui permet de produire l'analyse morphologique pour plusieurs langues, en travaillant avec des règles qui prennent en compte la forme des mots.
- Le projet LiSe (Linguistique et Sécurité) dans le domaine de la sécurité globale [Cardey *et al.*, 2010], qui propose des normes linguistiques pour des applications de sécurité, en proposant des langues contrôlées et en travaillant sur le « Data et Sense Mining » pour la reconnaissance de messages et de leur sens.
- Le dictionnaire MultiCoDiCT (Multilingual Collocation Dictionary Centre Tesnière) [Cardey *et al.*, 2006], qui a été construit en utilisant des algorithmes sur des ensembles et partitions des langues.
- La traduction automatique pour des textes techniques, basée sur les langues contrôlées, et implémentée pour un grand nombre de langues différentes, dont le français, l'arabe, le russe, le chinois et le thaïlandais.

4.3.2 La méthode d'Exploration Contextuelle

La méthode de l'Exploration Contextuelle (EC) [Desclés, 1997, Desclés *et al.*, 1997, Desclés, 2006] repose sur une réflexion linguistique permettant d'annoter des corpus à partir de ressources linguistiques.

En effet, dans ce modèle, [Desclés *et al.*, 1997] distingue le contexte externe du contexte interne :

- le contexte externe correspond à l'ensemble des connaissances sur le domaine cible, sur l'énonciateur ou les interlocuteurs, sur la situation dans laquelle le discours a été produit, etc. ;
- le contexte interne porte sur les connaissances linguistiques identifiables et analysables dans le textes.

L'étude linguistique consiste à déterminer les valeurs sémantiques des différents marqueurs linguistiques d'une notion discursive ou d'une catégorie grammaticale. En attribuant à chaque élément linguistique (valeur aspectuelle, phrase importante dans un texte, etc.) des caractéristiques associées à un ensemble de règles, la méthode d'EC propose une stratégie opératoire permettant de gérer l'indetermination sémantique. Pour cela la méthode d'EC repose sur deux notions clés :

- les *indicateurs* qui représentent les expressions porteuses de sens ;
- les *indices* qui représentent des unités linguistiques contextuelles associées aux indicateurs.

L'identification de la présence d'un indicateur dans un segment textuel déclenche une règle d'EC qui recherche les indices dans le contexte à gauche ou à droite de l'indicateur, et qui décide d'effectuer l'annotation sémantique⁹ (voir schéma 4.2, p. 85). Les indices peuvent être positifs ou négatifs. Selon [Desclés *et al.*, 1997], les règles d'EC permettent de désambiguïser les occurrences des indicateurs en tenant compte du contexte.

⁹Dans d'autres cadres que le nôtre, une règle d'EC peut effectuer d'autres tâches, par exemple, [Desclés, 1997, § 3 *Exemples de systèmes d'exploration contextuelle*] : lever une indétermination sur un aspect sémantique, retenir un segment textuel pour un résumé automatique, localiser des indices pouvant préciser des connaissances.

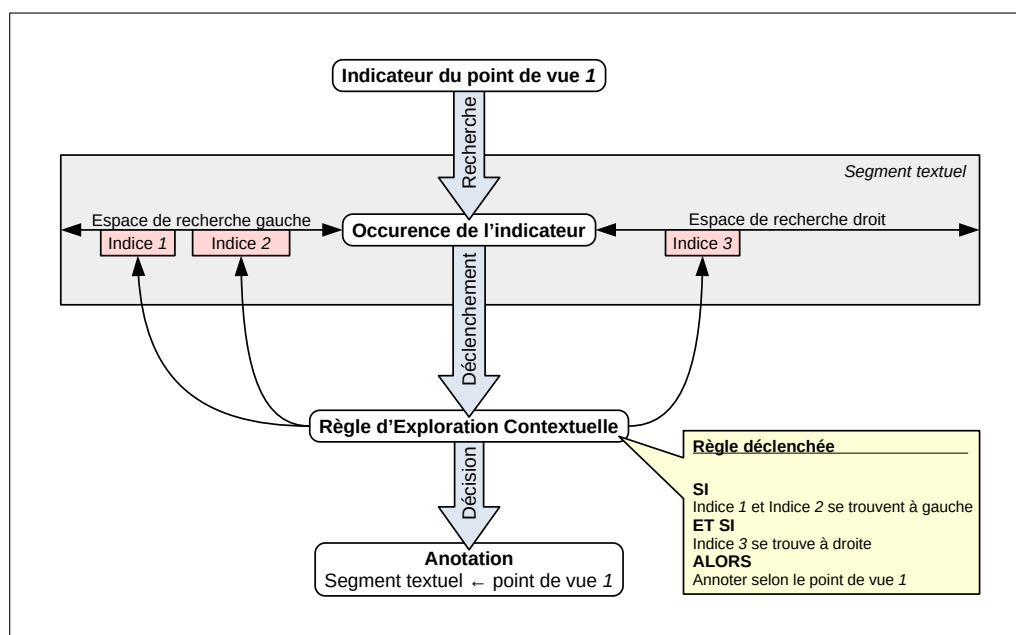


Schéma 4.2: Fonctionnement de la méthode de l'Exploration Contextuelle de J.-P. Desclés (1997, 2006), selon [Atanassova, 2012, fig. 3.1 p. 85].

La méthode d'EC, [Desclés, 1997, § 4], s'appuie sur une « catégorisation morphologique des unités linguistiques », sur des inventaires d'indices et de compléments et alternatives à ces indices, et « ne nécessite pas une analyse syntaxique préalable approfondie ». Elle peut éventuellement aussi signaler un manque d'indices connus liés à un indicateur dans le segment textuel exploré.

Parmi les caractéristiques particulièrement intéressantes des systèmes d'EC, [Desclés, 1997, § 5] mets en avant entre autres les suivantes : ils sont extensibles en complétant les inventaires d'indicateurs et indices, et cette richesse croissante leur permet « de s'approcher d'une solution idéale ».

Les travaux de [Atanassova, 2012, p. 81-85] apportent une description de la méthode d'EC et une application à la recherche d'informations par l'analyse des catégories discursives, notamment les *hypotheses*. Par exemple, le segment textuel « nous supposons que » peut servir à identifier une hypothèse dans une phrase comme « Dans cet article, nous supposons qu'il existe une corrélation entre les deux variables », alors que sa substitution par le segment « nous vérifions que le processus » fait que la phrase n'exprime plus d'hypothèse. Cependant, elle précise que des levées d'ambiguïté sont nécessaires sur les segments avant leur annotation ; et que se sont les indices de leur contexte proche à gauche et à droite

qui servent à préciser le sens dans lequel ces segments sont employés :

« Les unités linguistiques étant généralement polysémiques, l'identification d'un indicateur dans un segment ne suffit pas pour en effectuer l'annotation. En effet, afin de garantir la qualité des annotations et d'éliminer le bruit, l'algorithme doit être capable de tenir compte de différents phénomènes linguistiques, tels que la polysémie et l'homonymie, les usages figés, les usages métaphoriques, etc. ».

Les travaux de [Makkaoui *et al.*, 2012] se proposent d'étudier les phrases spéculatives dans les textes biomédicaux. L'implémentation informatique, BioEx-com, est basée sur l'outil EXCOM [Djioua *et al.*, 2006]. Il s'agit d'une implémentation de la méthode d'EC qui repose sur une recherche hiérarchique de marqueurs linguistiques de surface. Ce travail utilise un corpus BioSpe pour l'annotation des phrases spéculatives, et l'évaluation est réalisée sur le corpus BioScope.

4.4 Marqueurs des expressions de l'incertitude

Par *marqueur*, nous nous référons ici à une expression simple ou complexe qui transmet une qualité remarquable à ce qui est dit dans un texte ou dans un discours. Par exemple dans la phrase « *This may be simply a statistical artefact due to a small sample* »¹⁰, le marqueur d'incertitude '*may be*' transmet la qualité 'incertain' au fait '*is simply a statistical artefact due to a small sample*' sur lequel il porte. Nous partons du point de vue que les marqueurs qui nous intéressent, ceux de l'incertitude, sont détectables dans le texte et différenciables par nos traitements. *Nota bene* : nous ne nous référons aucunement au marqueur de syntagme (*phrase marker*) qui, en linguistique, sert à représenter la structure d'une phrase sous la forme structurée d'un diagramme en arbre (*tree diagram*).

De son côté, [Vold, 2006, p. 65] définit les marqueurs de modalité épistémique comme des « expressions linguistiques qui qualifient la valeur de vérité du contenu d'une proposition (par exemple *peut-être*, *probablement*) »¹¹ ; et par ailleurs, [Dalloux, 2017, p. 1] parle de marqueurs de l'incertitude en tant que « termes qui expriment [...] une incertitude » et qui ont une « influence sémantique [...] sur

¹⁰Phrase extraite du corpus sur le changement climatique : Spillman C.M. et Alves O. (2009), « Dynamical seasonal prediction of summer sea surface temperatures in the Great Barrier Reef » in Coral Reefs, vol. 28, n° 1 p. 197-206.

¹¹Traduction de « *linguistic expressions that qualify the truth value of a propositional content* (for example *perhaps*, *probably*) ».

les parties voisines de la phrases », et il donne avec [Grabar *et al.*, 2018, p. 5] des exemples de marqueurs lexicaux ('probablement' et 'certainement') et d'indices morphologiques verbaux¹² du conditionnel ('indiquerait' et 'proviendrait').

Nous faisons l'hypothèse que les vocabulaires utilisés seuls comme marqueurs de l'incertitude, peuvent aussi se trouver dans des expressions qui en fait ne sont pas des expressions d'incertitude : c'est pourquoi nous les nommons *marqueurs-candidats*. C'est pour palier ce problème que nous employons des expressions linguistiques plus complexes que ce vocabulaire.

Pour [Desclés et Jackiewicz, 2006, p. 35], l'énoncé d'une possible relation causale peut se traduire par des expressions telles que des connecteurs (ex. : 'parce que', 'puisque', 'car', ...), mais aussi par des verbes (ex. : 'provoquer', 'déclencher', 'conduire à', ...).

[Fintel (von), 2006] note la variété des catégories d'expressions qui ont une signification modale, tout en avertissant sur la multiplicité des significations et forces possibles de ces expressions¹³. Il en cite certaines en exemple pour l'anglais, parmi lesquelles des :

auxiliaires modaux : 'must', 'should', 'might', 'may', 'could' ;

verbes semi-modaux : 'have to', 'ought to', 'needs to' ;

adverbes : 'perhaps' ;

noms : '(slight) possibility' ;

adjectifs : 'necessary' ;

conditionnels : 'if' ;

marqueurs d'évidence (*evidentiality*) : 'apparently' ;

marques de modalité (*mood*) : flexions verbales, 'believe' qui apparaît au subjonctif ;

attitude propositionnelle : '[quelqun] suspects that [quelque chose est un fait]',
'Given [quelqun]'s evidence, the [quelqun d'autre] might be guilty' ;

¹²Traduction de : *morphological cues like conditional verbs*.

¹³Parmi ses exemples : 'Considering the evidence before us, it has to be raining', 'Visitors have to leave by six pm', 'You have to go to bed in ten minutes', 'I have to sneeze', 'To get home in time, you have to take a taxi'.

4.4. MARQUEURS DES EXPRESSIONS DE L'INCERTITUDE

expressions illocutoires ¹⁴ (*Expressions of illocutionary force*) : '[quelqun] is, I suspect, guilty' ;

modalités cachées : 'When you have computer trouble, [quelqun] is the person to talk to' pour '[quelqun] is the person one ought to talk to' ;

modalités de nature opaque : 'What Arlo is cooking has garlic in it'.

A l'occasion de l'identification de l'incertitude dans des textes journalistiques, [Rubin *et al.*, 2006] définissent la notion de *certitude* comme l'évaluation par l'auteur sur la vérité d'une affirmation, au moyen d'une gamme graduelle d'expressions à large éventail d'états¹⁵, et remarquent son lien avec d'autres concepts abordés par le traitement automatique des langues : la subjectivité, la modalité, l'évidencialité et le *hedging*. Elles signalent que l'un des types de commentaires épistémiques sont les expressions de certitude comme 'probably', 'perhaps' et 'undoubtedly', et que la mise en évidence de la certitude en anglais se fait par des moyens tels que les adverbies d'attribution (*attributive adverbials*) comme 'supposedly' et 'allegedly' et des verbes déclaratif (*reporting verbs*) comme 'claim' et 'suggest'. En s'appuyant sur ce cadre d'analyse, [Goujon, 2009, p. 120] développe des catégories de formes linguistiques associées aux dimensions ou valeurs 'niveau' haut ou modéré ou bas, 'perspective' du point de vue du rédacteur ou rapporté, 'temps' passé ou présent ou futur, 'réalité' par assertion ou par négation, et 'nom de source' pour un point de vue rapporté. Il en cite des exemples pour le français :

adjectifs :

niveau bas : 'douteux', 'incertain', 'peu probable' ;

niveau modéré : 'présumé', 'supposé' ;

niveau haut : 'vraisemblable', 'probable', 'possible', 'envisageable', 'envisagé'.

verbes :

¹⁴Expressions qui réalisent ou tendent à réaliser, par la parole, l'action dénommée. Par exemple par le moyen de verbes performatifs tels que *ordonner* et *jurer*. Source : [Dubois *et al.*, 2002, p. 240].

¹⁵« *a full spectrum of certainty states ranging from doubt to complete conviction in the truth of a statement* ».

niveau modéré : 'dire', 'déclarer', 'annoncer', 'penser', 'croire', 'douter',
'hésiter'.

expressions :

niveau modéré : 'selon toute vraisemblance', 'sans doute', 'à ce que l'on
dit', 'il se peut que', 'il paraît'.

structures :

perspective du point de vue reporté : 'selon', 'd'après', 'de source(s)',
'« ... »' ;

niveau modéré : 'si' (sauf suivi par un adverbe) ;

temps : futur : 'aller' + infinitif ;

réalité : négation : 'ne' et "n'" (sauf "n'importe").

Pour l'analyse automatique des niveaux de certitude, [Bernhard et Ligozat, 2011, p. 4-5] présentent deux systèmes fonctionnant avec la langue anglaise pour des problèmes médicaux mentionnés dans des comptes-rendus cliniques¹⁶, l'un par apprentissage supervisé et l'autre par règles, pour la graduation des niveaux d'incertitude.

4.5 Métadonnées de l'incertitude

De manière générale, pour l'étude et l'utilisation des données textuelles (ou autres) en traitement automatique des langues, il peut être utile de rendre disponibles ces mêmes données textuelles avec de nouvelles informations que l'on ajoute. Typiquement les données ajoutées, appelées *métadonnées*, se présentent sous forme d'étiquettes. Par exemple, pour indiquer que l'expression « elle » correspond à une partie du discours appelée « pronom », nous pouvons utiliser le format : *<pronom>elle</pronom>*. L'étiquetage (dit aussi *annotation*) permet de signaler et distinguer des segments du texte simples ou complexes, par exemple : un suffixe, ou un mot, ou le titre d'un paragraphe). L'étiquetage permet de délimiter ces segments de texte, et de les qualifier.

L'un des moyens pour mettre en place des métadonnées est d'utiliser le XML (*extensible markup language*) qui est par définition une technologie informatique

¹⁶« [...] afin de savoir si le problème est présent, absent, hypothétique, soumis à certaines conditions, ou encore associé à une personne différente [...] ».

d'étiquetage des données (mais pas seulement, car elle permet aussi le stockage, la transformation, etc., et rend possible ou facilite d'autres technologies). XML permet aux utilisateurs de l'informatique de créer des « dialectes XML » ; c'est-à-dire de créer des vocabulaires d'étiquetage spécialisés pour un usage, par exemple : indiquer à quelle partie du discours appartient chaque expression d'un texte.

Il existe des dialectes XML pour des concepts scientifiques de divers domaines¹⁷, par exemple : AnatML (Anatomical Markup Language), archaeoML (Archaeological Markup Language), BSMML (Bioinformatic Sequence Markup Language), CellML (Cell Markup Language), ChemML et CML (Chemical Markup Language), etc., dont une partie fait l'objet de normes internationales, tel que GML¹⁸, ou s'inspirent de normes internationales, tel que UncertML.

Le dialecte UncertML (Uncertainty Markup Language)¹⁹ apparut vers 2008, et son successeur QualityML, concernent des concepts de l'incertitude scientifique. UncertML est un dialecte XML expérimental créé pour encoder des métadonnées, telles que celles de distributions statistiques, pouvant transmettre des informations sur l'incertitude qui accompagne des données géospatiales (par ex. environnementales et climatiques). UncertML permet par exemple d'indiquer la valeur de la variance dans une distribution qui suit une courbe de Gauss. [Williams *et al.*, 2009, p. 11] précisent que UncertML prend comme référence la norme ISO/IEC GUIDE 98-3: *Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)* et le définissent comme :

« [...] *an XML encoding for the transport and storage of information about uncertain quantities, with emphasis on quantitative representations based on probability theory* ».

L'une des intentions du projet ayant mené à la création de UncertML était que ses métadonnées soient exploitables dans le web sémantique (voir le web sémantique § 5.1, p. 93). Selon [Masó, 2017, § 5.4.2, p. 25] :

« *The problem with UncertML is that it only deals with one way of measuring the data quality: uncertainty. This works well with some*

¹⁷Source principale : Geroimenko (2004), « Dictionary of XML Technologies and the Semantic Web ». ISBN 1-85233-768-0. Springer, Londres, Royaume-Uni

¹⁸Le GML est l'objet de la norme ISO 19136.

¹⁹Sources principales : <http://www.uncertml.org> (le site web du projet a été fermé : la version consultée est celle sauvegardée sur <https://web.archive.org>) et <https://cordis.europa.eu/project/rcn/93798/>, consultées le 6/3/2019.

measurements of accuracy but does not cover other quality measures like consistency, completeness etc. We needed a way to extend the UncertML approach to all other quality indicators and to harmonize them with the measurement definitions that ISO19157 is providing and many more that are used in other domains ».

Chapitre 5

Vers une ontologie de l'incertitude scientifique

Une ontologie informatique est la description formelle d'une collection de (termes-)concepts spécifiques à un domaine de connaissance, de relations explicites que ces concepts ont entre eux, et éventuellement de propriétés qu'ils possèdent. Les ontologies sont, entre autres, une des bases du *web sémantique*¹ et sont, selon [Raj et Raman, 2017, p. 33-38], un constituant potentiel important de l'*internet des objets*. Dans le web sémantique, les constituants de base des ontologies sont des triplets sous la forme *sujet - relation - objet* (p. ex. : 'les gaz à effet de serre' - 'contribuent à la destruction de' - 'la biosphère terrestre').

5.1 Les ontologies informatiques et le web sémantique

Les ontologies informatiques sont des formes d'organisation de connaissances spécifiques à un domaine donné. Elles modélisent un « univers » d'un domaine de connaissance où les objets et leurs relations sont définis. Relier des concepts dans une ontologie permet de les représenter dans des systèmes informatisés pour des calculs, ou dans des graphiques pour des usages humains : une ontologie est représentable informatiquement et visuellement sous la forme d'un graphe où les nœuds représentent les concepts principaux cohérents entre eux, et où les liens

¹L'expression *web sémantique* est parfois écrite avec une majuscule à *Web* comme en anglais (*Semantic Web*), mais, de même qu'*internet* s'écrit sans majuscule en français, nous l'écrivons sans majuscule à *web*.

entre ces nœuds représentent les relations orientées entre ces concepts.

Les concepts des ontologies informatiques permettent de coder ces structures et leurs sens, et éventuellement de les utiliser/calculer et de les visualiser, pour comprendre l'univers qu'elles modélisent et produire des connaissances en y découvrant des interrelations qui n'étaient autrement pas à découvert.

Une ontologie peut être visualisée sous la forme d'une carte conceptuelle, c'est-à-dire sous la forme d'un réseau composé de nœuds qui prennent le nom de l'entité qu'ils représentent (par ex. : le nom 'Jean-François Champolion' ou l'appellation 'les hiéroglyphes égyptiens' ou le titre 'Précis du système hiéroglyphique'), lesquels sont connectés par des lignes prenant le nom de la relation orientée qu'elles représentent (par ex. : 'a déchiffré' ou 'a été écrit par'). Le terme de taxonomie est parfois associé à celui d'ontologie, et, bien que les deux puissent se représenter sous la forme d'un réseau de nœuds connectés, ils désignent des représentations distinctes des connaissances d'un domaine. Dans une taxonomie les nœuds ne représentent que des termes, alors que dans une ontologie les termes sont dans les nœuds. Ces liens permettent d'assembler automatiquement les concepts avec les significations que ces liens représentent ; et permettent même de découvrir algorithmiquement, automatiquement, des liens entre les concepts et entre les entités ou objets que peuvent représenter ces concepts. Les ontologies permettent ainsi de décrire formellement un ensemble articulé de concepts qui composent un univers de connaissances concrètes ou abstraites. Elles permettent de décrire des interrelations entre ces concepts, ou entités ou objets.

Par exemple une ontologie représentant l'univers familial humain sera l'assemblage de concepts définissant qui peuvent être les membres d'une famille humaine (un beau fils, un grand oncle, une mère ...), et décrivant quelles règles produisent les relations de parenté (est mari de la fille, est frère du grand-père, est parent au 1^{er} degré et est de genre féminin ...).

Les « briques » de base pour composer des ontologies sont des *triplets*. Par exemple, dans le domaine familial les triplets sont des enregistrements pouvant prendre la forme '*A*' '*est parent au 1^{er} degré de*' '*B*' ; et ces triplets peuvent être explorés automatiquement (par exemple avec le langage SPARQL) pour en découvrir de nouveaux (créer de nouvelles connaissances) grâce à des règles du type « Si '*A*' '*est du type*' '*féminin*', alors '*A*' '*est mère de*' '*B*' ».

Ces plus petits éléments des ontologies, les triplets, peuvent se décrire sous la forme sujet→relation→objet. Ces "atomes" ontologiques sous forme de triplets permettent d'enregistrer et de présenter des concepts et données liées, par exemple dans le *web sémantique* avec un format de phrase minimaliste de type

'sujet→verbe→complément'. Par exemple, "x"→"est père de"→"y" peut correspondre à :

"Johann Georg Leopold Mozart"→"est père de"→"Wolfgang Amadeus Mozart".

Les ontologies peuvent servir à représenter les connaissances dans les domaines les plus divers, et c'est pourquoi nous avons envisagé de nous appuyer sur une ontologie pour décrire le domaine de l'incertitude dans le discours scientifique : pour nommer et organiser d'un point de vue pratique les concepts de l'incertitude scientifique nécessaires et suffisants pour l'identification des expressions d'incertitude.

Les formes que revêtent les ontologies dépendent de l'usage qui doit en être fait. En traitement automatique des langues, [Couillault *et al.*, 2005, p. 51] :

« les ontologies "réelles" sont [...] le résultat de multiples compromis entre fonctionnalité et complexité. Il s'agit donc d'adapter leur niveau de détail à leurs visées opérationnelles, de concilier volume de concepts à représenter avec le maintien de la cohérence, et de garantir le degré de formalisation requis par les capacités des logiciels qui devront les utiliser ».

Pour être exploitées par des moyens informatiques, les ontologies s'appuient sur des « propriétés de consensus, de cohérence, de précision et d'évolutivité » qui « dépendent directement des propriétés du formalisme de représentation utilisé », [Roche, 2005, p. 57]. La digitalisation des données, informations, documents et objets scientifiques, se fait en particulier dans des formats basés sur des dialectes du XML comme par exemple le langage RDF (Ressource Description Framework) qui permet d'échanger des données sur le web sous forme de graphes, [Gandon *et al.*, 2012]. Les données publiées en RDF peuvent être réutilisées à l'aide d'autres dialectes, manuellement ou avec des applications spécialisées, par exemple : OWL (Ontology Web Language) permet d'utiliser le modèle de triplets de RDF pour décrire une ontologie, et le logiciel Protégé² permet l'édition d'ontologies OWL.

Le web sémantique, selon [Accart et Rivier, 2012, p. 55 et 56], « a pour objectif de faire émerger de l'information pertinente à partir de requêtes même confuses » en améliorant « le partage de données par une description des informations qui permettra à l'ordinateur de les relier entre elles au travers du sens qu'elles

²Protégé est un logiciel open source devenu un standard, sinon *le* standard, universitaire pour l'édition d'ontologies. Il est développé par le Stanford Center for Biomedical Informatics Research. URL : <http://protege.stanford.edu>

véhiculent », et avec la possibilité « de faire communiquer toutes les sources de données entre elles et d'en tirer des déductions logiques ».

Les liens entre les données du World Wide Web³, ont une signification qui peut être représentée sous la forme sujet→relation→objet. Par conséquent, cette signification peut être incluse dans la description de chaque lien entre des données. C'est cette démarche qui a amené Tim Berners-Lee à proposer le concept d'un *web sémantique* (dit aussi *web des données*) où les ordinateurs en réseau peuvent traiter (stocker, échanger, structurer, interroger, découvrir et représenter) la signification des données à travers les liens qui unissent ces données. Au cœur de la construction technologique du web sémantique, nous retrouvons les *ontologies informatiques*.

5.2 Présence de l'incertitude dans des ontologies existantes

Le formalisme structurel des ontologies informatiques peut permettre la découverte automatique de nouvelles connaissances par l'exploration automatique dans une ontologie suffisamment développée, ou dans l'union d'ontologies différentes, ce qui est un domaine de recherche en soi. Ce formalisme structurel des ontologies informatiques facilite aussi la découverte de connaissances ontologiques par l'exploration manuelle d'ontologies à l'aide de moteurs de recherche d'ontologies généralistes comme LOV⁴ et Ontobee⁵, ou spécialisés par exemple en particulier dans le biomédical comme OLS⁶ et OBO Foundry⁷.

La recherche manuelle via des moteurs comme ceux cités a permis de découvrir une ontologie de l'incertitude préexistante : mUnc Vocabulary Specification (voir ci-dessous, p. 98). Cette recherche manuelle a aussi permis de se familiariser quand ce n'était déjà fait avec des termes-concepts d'incertitude « isolés » dans des ontologies qui ne traitent pas spécifiquement l'incertitude ; et elle a suggéré qu'une partie de ces termes-concepts pourraient ontologiquement être connectés

³Ou bien entre les données de l'internet en dehors du World Wide Web ; ou bien entre les données d'autres réseaux d'information isolés d'internet, par exemple des réseaux bancaires ou militaires.

⁴Linked Open Vocabularies (LOV) : <https://lov.linkeddata.es/dataset/lov>.

⁵Ontobee : <http://www.ontobee.org>, au départ spécialisé et devenant généraliste.

⁶Ontology Lookup Service (OLS) : <https://www.ebi.ac.uk/ols/index>.

⁷Open Biological and Biomedical Ontology (OBO) Foundry : <https://obofoundry.org>.

ou précisés par les catégories de l'ontologie de l'incertitude que nous proposons (voir schéma 5.3, p. 110). Certains de ces termes-concepts disposent d'une définition scientifique ou technique dépendante du domaine de leur ontologie, par exemple :

- *Class: Uncertainty* (source : *Mass spectrometry ontology*, http://purl.obolibrary.org/obo/NCIT_C71478)
 - *Semantic_Type: Conceptual Entity*
 - *Definition: A measure of the inherent variability of repeated observation measurements of a quantity including quantities evaluated by statistical methods and by other means. [Definition Source: NCI]*
- *Label: Coordinate Uncertainty In Meters* (source : *Biodiversity Information Standards (TDWG)*, <http://rs.tdwg.org/dwc/terms/version/coordinateUncertaintyInMeters-2021-07-15.htm>)
 - *Type: Property*
 - *Definition: The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term.*

A cette étape de recherche d'une ontologie, le besoin était de découvrir les principales catégories d'incertitude scientifique qui soient organisables de manière fonctionnelle, et qui soient transversales à l'ensemble des champs scientifiques du changement climatique. Nous avons donc aussi recherché des termes-concepts d'ontologies liés à la fois aux notions d'incertitude et au changement climatique. C'est par exemple la définition de l'incertitude dans l'ontologie Linked Paleo Data (LiPD) ⁸, ontologie développée pour structurer les descriptions de jeux de données données en paléoclimatologie : la LiPD reprend une définition de l'incertitude définie par l'IPCC ; et cette définition inclue notamment que l'incertitude peut être représentée par des mesures quantitatives et par des déclarations qualitatives. Ces deux aspects quantitatif et qualitatif de l'incertitude sont

⁸Version du 4 avril 2017 disponible sur <http://linked.earth/ontology/core/1.2.0/index-en.html>.

à rapprocher de ceux que nous avons retenus pour le niveau intermédiaire des catégories de l'ontologie à laquelle nous avons nous-mêmes abouti.

Les termes-concepts d'incertitude relativement isolés, les uns des autres par leurs définitions et par leur inclusion dans des ontologies de domaines différents, ont laissé penser que les catégories de l'ontologie de l'incertitude que nous proposons peuvent contribuer à améliorer des notions d'incertitude d'autres ontologies en les précisant ou en les étendant.

Selon le rapport [W3C, 2008] qui cherche à cerner des moyens de raisonnement dans les données du Web en prenant en compte l'incertitude sur le web :

« un format standardisé pour représenter l'incertitude permettrait aux agents qui reçoivent une information imparfaite de l'interpréter suivant l'intention de l'agent envoyeur »⁹.

L'architecture du web sémantique, dit aussi web des données, se compose d'un empilement de couches de technologies informatiques théorisé et décrit par Tim Berners-Lee en une *Pile du web sémantique* (*Semantic Web Stack*). La représentation graphique de cette Pile du web sémantique (voir graphique 5.1, p. 99, tel que présenté par Tim Berners-Lee¹⁰) indique lesquelles de ces technologies servent de bases pour mettre en place les suivantes (supérieures dans le graphique), et désigne les ontologies comme l'une de ces technologies.

Dans la suite du rapport [W3C, 2008] déjà cité, [Djebri *et al.*, 2019a, Djebri *et al.*, 2019b, Djebri *et al.*, 2020] proposent d'inclure un formalisme RDF standardisé d'incertitude à la pile du web sémantique, décrit par une ontologie nommée *mUnc Vocabulary Specification*¹¹ (*mUnc* pour *meta-Uncertainty*). Cette ontologie, qui selon nos recherches est la seule en dehors de la nôtre à être consacrée à l'incertitude, a pour fonction de permettre le calcul et la propagation d'incertitudes à partir d'incertitudes déjà évaluées et représentées et publiées en formalismes RDF à l'intérieur du web sémantique : elle ne correspond pas à nos besoins de description de catégories génériques des incertitudes scientifiques présents dans des expressions linguistiques, qui n'expriment pas toujours des intensités d'incertitude calculables ou échelonnables sur une dimension ; et elle

⁹Traduction de « a standardized format for representing uncertainty would allow agents receiving imperfect information to interpret it in the same way as was intended by the sending agents ».

¹⁰Tim Berners-Lee (2000), « Semantic web on XML » in « Semantic Web XML - 2000 », p. 10. Présentation. URL : <https://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>. Consulté le 24/09/2018.

¹¹Disponible sur <http://ns.inria.fr/munc/v2/index.html>.

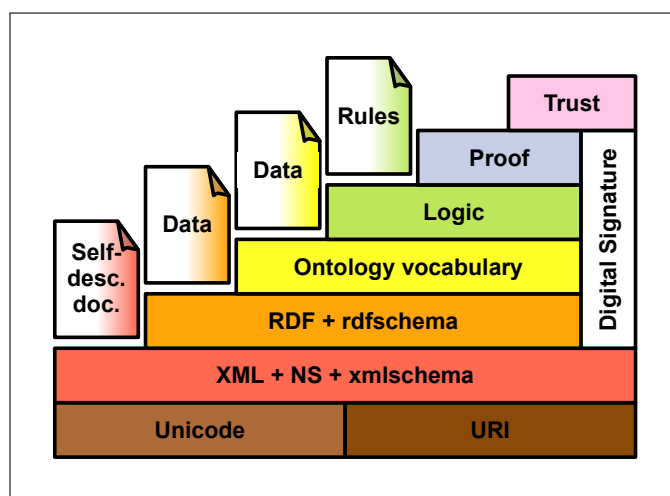


Schéma 5.1: La pile du web sémantique (*Semantic Web Stack*), selon Tim Berners-Lee.

ne coïncide d'ailleurs en rien non plus avec nos premiers essais d'ontologie finalement écartés, [Rey *et al.*, 2018, schéma 2, p. 238], (voir ce même schéma numéroté 5.2, p. 109). Cependant, cette recherche différente, mais pas contradictoire, qui pourrait confluer avec la nôtre dans le futur, illustre à sa manière le besoin de traitements globaux des incertitudes, qu'elles soient exprimées dans des formalismes du web sémantique ou dans une langue naturelle.

5.3 Une définition de l'incertitude

L'étude des expressions linguistiques marqueurs de l'incertitude, dans le domaine du changement climatique, a pour objectif de reconnaître les particularités de l'incertitude liées au domaine et d'en tirer des conséquences méthodologiques et pratiques. La diversité et l'éloignement des domaines scientifiques représentés dans le corpus d'articles sont une indication que le domaine du changement climatique réunit des concepts d'incertitude jusqu'ici difficilement unifiables ou formalisables ensemble. Il a pourtant paru souhaitable de considérer une perspective interdisciplinaire qui dépasse les concepts d'incertitude si nombreux, puisque propres à chaque (sous-)discipline. Autrement dit, l'objectif est de considérer une perspective globale de l'incertitude en envisageant des recherches futures.

Dans notre travail, nous avons cherché à identifier des concepts généralisables de l'incertitude scientifique qui, à la fois, sont communs à l'ensemble des do-

5.3. UNE DÉFINITION DE L'INCERTITUDE

maines scientifiques et se prêtent à une classification. Nous avons remarqué au long de notre recherche qu'une proportion très majoritaire des auteurs (toutes disciplines confondues) aborde les incertitudes du point de vue quantitatif ; et cela a renforcé notre intérêt pour aborder aussi les incertitudes qualitatives.

Après l'examen de définitions générales principalement dans les usages francophones et anglophones, et après l'examen de définitions ou d'approches scientifiques, nous précisons les concepts suivants en commençant par le plus englobant, pour les synthétiser et les situer entre eux de manière pratique ou opérationnelle :

subjectivité (notre définition) : (en anglais *subjectivity*) segments du discours par lesquels le locuteur indique des jugements, impressions, sentiments ou implications personnels, ou bien des références à une situation imprécise ou changeante.

Dans cette subjectivité, notre approche, elle, se tourne vers le discours scientifique, et donc vers une modalité sur les faits, à valeur épistémique :

modalité (notre définition) : (en anglais *subjectivity* ou *subjectiveness*) élément de la langue, qui exprime un degré ou une variation d'une assertion. Par exemple, à propos d'une vérité ou à propos de l'état d'une chose ou d'un être.

Dans les domaines scientifiques ou techniques que nous avons explorés à propos de leur usage de l'incertitude, les méthodes pour appréhender et exprimer le concept d'*incertitude* changent en fonction du domaine. Il en ressort aussi que l'incertitude est considérée comme un phénomène inhérent à la recherche. C'est un objet informationnel inséparable des résultats à transmettre. Transmettre les données de l'incertitude de la recherche, parfois complexes ou composites, permet de ne pas fausser les résultats de la recherche et de les prendre méthodiquement en compte dans des travaux ultérieurs.

Le caractère évolutif de la nature et de la catégorie de la certitude est parfois déterminant, en fonction de l'évolution de l'objet de la recherche, dans l'espace (p. ex. : nature d'un terrain) ou dans le temps (p. ex. : augmentation d'un corpus) ou dans d'autres dimensions (interactions complexes dans une simulation, instrument, méthode, savoirs et présupposés, savoir-faire, etc.). Des remarques critiques sur l'usage scientifique des logiciels de systèmes d'information Géographique (GIS), [Kerdjoudj, 2015], suggèrent que la recherche de la catégorie de

certitude dépend non seulement de la question et de l'actualisation des données, mais aussi du mode de représentation des résultats intermédiaires et finaux. Par ailleurs, [Lyons, 1977, p. 797] signale la difficulté de distinguer nettement entre les modalités épistémiques objectives (qu'il essaye de distinguer de celles subjectives) et les modalités aléthiques, ce qui est repris par [Holmes, 1982, p. 12].

expression de l'incertitude scientifique (notre définition) : évidence explicite que l'auteur modifie la valeur épistémique d'au moins une proposition de la phrase.

Nous abordons donc l'expression d'incertitude non seulement par les valeurs mesurées et estimées, mais aussi par sa place dans le raisonnement scientifique.

D'une part, les incertitudes peuvent être exprimées de manière quantitative, avec des expressions de mesures chiffrées ou d'approximation de grandeurs : notre hypothèse est que ces expressions d'incertitude mesurée ont des formats et degrés communs à l'ensemble des disciplines scientifiques qui les emploient. D'autre part, les incertitudes peuvent être exprimées de manière qualitative, par des expressions de raisonnements ou de parties ou indices de raisonnements.

5.4 Les raisonnements formalisant l'incertitude

Les trois principaux types de raisonnements formalisés sont : déduction, induction et abduction. Parmi ces trois types de raisonnements formalisés, la déduction n'introduit pas d'incertitude, alors que l'induction et l'abduction intègrent une incertitude liée au résultat du raisonnement. Les quantités incertaines, elles, s'expriment soit par des mesures de valeurs ponctuelles soit par des mesures d'intervalles fermés ou ouverts.

Le raisonnement par abduction est généralement moins évoqué que les autres, voire ignoré, parmi les présentations de raisonnements formalisés. Il nous est donc apparu utile d'aborder les acceptions de ce concept moins familier.

5.4.1 Le raisonnement abductif

Le terme *abduction* est dérivé¹² du latin *abductio* (action d'enlever, de séparer). Le concept d'abduction rend compte d'un mode de raisonnement dont le but

¹²Dictionnaire Larousse en ligne. Consulté le 19/08/2018. URL : <http://www.larousse.fr>.

est de retrouver les causes plausibles d'un phénomène. [Catellin, 2012] lie ce raisonnement à l'imagination et à la créativité, et indique sa formalisation par le philosophe Charles S. Peirce à la fin du XIX^e siècle. Charles S. Peirce fait lui-même remonter le concept à Aristote ; lequel Aristote, [Zegarelli, 2007, p. 20], a considéré que la logique structure la connaissance scientifique, et en a traité les structures d'argumentations valides dans ses ouvrages composant l'Organon¹³.

Approches du concept d'abduction

Reconnu comme le pionnier de la notion d'abduction, Charles Sanders Peirce, [Peirce, 1958, vol. 1 § 65, vol. 2 § 776, vol. 5 § 144, vol. 7 § 251, vol. 8 § 209] en fait remonter les racines à Aristote. Il signale que, dû à ce qu'il suppose être une incompréhension du texte¹³ grec corrompu, l'usage est de traduire par 'abduction' le terme ἀπαγωγὴ¹⁴ [apagoogè] d'Aristote qui devrait en fait être traduit par 'rétroduction'. En se référant à Aristote, il indique que la *déduction*, l'*induction* et la *rétroduction* sont les trois raisonnements scientifiques fondamentaux ; et qu'ils sont accompagnés d'un quatrième raisonnement, l'*analogie*, qui combine l'induction et la rétroduction. Il observe, [Peirce, 1958, vol. 2 § 98], que l'on peut presque dire que ces trois raisonnements « sont les clefs de la logique »¹⁵.

[Peirce, 1958, vol. 2 § 102] commente lui-même que sa propre théorie de l'abduction a fortement évolué¹⁶. [Dumez, 2012, p. 3-4] remarque que « chacun semble donner de l'abduction sa propre définition, et toutes ces acceptions diffèrent assez largement entre elles », et que Peirce « en donne des dizaines de définitions au cours de sa vie, dont certaines apparaissent contradictoires ».

Dans sa quête de perfectionnement terminologique, Peirce nomme par ex-

¹³Peirce se réfère au chap. 25 du 2^e livre Premiers Analytiques (en anglais *Prior Analytics*) de la *Logique* d'Aristote (série de six traités constituant la *Logique*, dite aussi Organon, du milieu du IV^e siècle avant notre ère). Premiers Analytiques traite des structures de raisonnements faites pour être valides, et spécialement les syllogismes. Ce chapitre, dans la traduction française de [Saint-Hilaire, 1839, p. 334-337] est intitulé *De l'Abduction*.

¹⁴Cherchant à en cerner mieux un faisceau de sens traduisibles en plus de celui d'*abduction* qui est bien établi dans la langue moderne, nous notons d'autres traductions du terme grec ἀπαγωγὴ, selon Bailly A. (1901), *Abrégé du dictionnaire GREC FRANÇAIS*, p. 85. Hachette, Paris. Dépôt légal n° 8167.3.1984 : *action d'emmener, action en justice pour un procès pour un flagrant délit, action de faire dévier du droit chemin, paiement d'une contribution*.

¹⁵« *The division of all inference into Abduction, Deduction, and Induction may almost be said to be the Key of Logic* ».

¹⁶« Upon this subject, my doctrine has been immensely improved since my essay "A Theory of Probable Inference" was published in 1883. In what I there said about "Hypothetic Inference" I was an explorer upon untrodden ground ».

emple ce concept *présomption* en même temps qu'il propose d'employer pour celui-ci le terme d'*abduction* [Peirce, 1958, vol. 2 § 774]¹⁷, ou bien il paraphrase par « l'activité au cours de laquelle une hypothèse explicative est adoptée » [Peirce, 1958, vol. 5 § 189]¹⁸ ou par « Le premier pas d'une hypothèse et son alimentation, qu'elle soit une simple interrogation ou qu'elle possède un degré quelconque de confiance » [Peirce, 1958, vol. 6 § 525]¹⁹. Autre exemple, parmi les différentes appellations de ce concept utilisées par d'autres auteurs traitant le sujet, [Harman, 1965, p. 88-89] choisit d'utiliser l'expression *inférence de la meilleure explication*^{20 21}, « qui correspond approximativement à ce que d'autres ont appelé 'abduction' », parce que « cela évite la plupart des suggestions de terminologies alternatives, qui induisent en erreur »²².

L'observation moderne du raisonnement abductif n'est donc plus une nouveauté et a fait l'objet de considérations ou de remarques dans diverses disciplines scientifiques, par exemple en économie où [Viviani, 1994, p. 118] soumet déjà l'idée que ce raisonnement s'applique « au choix d'un investissement, de nouvelles technologies, de nouveaux biens ». [Peirce, 1958, vol. 2 § 96] note que le raisonnement abductif est original par le fait qu'il est le seul à faire naître une nouvelle idée. Il va même jusqu'à affirmer que c'est l'abduction qui est à l'origine de toutes les idées scientifiques [Peirce, 1958, vol. 5 § 145]²³. D'autres auteurs lui emboîtent le pas, par exemple [Andreewsky et Bourcier, 2000, p. 839], en insistant aussi sur le fait que l'abduction s'applique à des observations inattendues, des anomalies dans des résultats d'expériences, des inconsistances avec la théorie ; et par là ils font aussi le lien avec la créativité : celle de la découverte de causes plausibles, et, au-delà, celle de l'extension des théories et celle du développement de nouvelles théories.

¹⁷« Reasoning is of three elementary kinds; but mixed reasonings are more common. These three kinds are induction, deduction, and presumption (for which the present writer proposes the name abduction) »

¹⁸Traduction de « the operation of adopting an explanatory hypothesis »

¹⁹Traduction de « The first starting of a hypothesis and the entertaining of it, whether as a simple interrogation or with any degree of confidence »

²⁰Traduction de : « The inference to the best explanation ».

²¹Actuellement, selon [Beckwith, 2018, p. 3, 17], les définitions de l'abduction relèvent de cette 'inférence de la meilleure explication' de Gilbert Harman.

²²Traduction de : « it avoids most of the misleading suggestions of the alternative terminologies ».

²³« All the ideas of science come to it by the way of Abduction. Abduction consists in studying facts and devising a theory to explain them ».

L'abduction en logique

En logique, l'abduction est un « Raisonement par lequel on restreint dès le départ le nombre des hypothèses susceptibles d'expliquer un phénomène donné »²⁴. Ce procédé par élimination, en examinant l'ensemble des faits et des connaissances (propriétés, règles, théories, etc.) sur le contexte des faits, doit permettre notamment d'inférer les causes de ces faits : selon [Catellin, 2004, p. 179], il s'agit d'adopter « des hypothèses plausibles susceptibles d'être vérifiées ultérieurement ». [Bratko, 2012, p. 606, 612-617] montre que des raisonnements abductifs de forme logique sont automatisables.

[Polya, 1985, part. 3 *Short dictionary of heuristic*, sous-part. *Signs of progress*, § 6 *Heuristic syllogism* et § 7 *Nature of plausible reasoning*] choisit d'employer l'expression 'raisonnement plausible' (*plausible reasoning*) bien que celle-ci ne lui convienne pas parfaitement²⁵. Il donne un exemple de raisonnement qui peut être utilisé pour restreindre des hypothèses (syllogisme démonstratif), et un exemple de raisonnement abductif proprement dit (syllogisme heuristique), les deux apparaissant en fait complémentaires :

- raisonnement par syllogisme démonstratif (*demonstrative syllogism*) :

Si A est vrai, alors B est aussi vrai, comme nous savons. (Prémisse)

Or, il s'avère que B est faux. (Prémisse)

Donc, A est faux. (Conclusion)

Noté autrement :

$$\frac{\begin{array}{l} \text{Si } A \text{ alors } B \\ B \text{ faux} \end{array}}{A \text{ faux}}$$

- raisonnement par syllogisme heuristique (*heuristic syllogism*) :

Si A est vrai, alors B est aussi vrai, comme nous savons. (Prémisse)

Or, il s'avère que B est vrai. (Prémisse)

Donc, A devient plus crédible. (Conclusion)

Noté autrement :

²⁴Dictionnaire Larousse en ligne. Consulté le 19/08/2018. URL : <http://www.larousse.fr>.

²⁵Polya précise : « *for lack of a better term*, plausible reasoning ».

$$\frac{\begin{array}{l} \text{Si } A \text{ alors } B \\ B \text{ vrai} \end{array}}{A \text{ plus crédible}}$$

- Exemple de syllogisme heuristique :

$$\frac{\begin{array}{l} \text{If we are approaching land, we often see birds.} \\ \text{Now we see birds.} \end{array}}{\text{Therefore, it becomes more credible that we are approaching land.}}$$

Concernant la relation entre abduction et plausibilité, [Desclés et Jackiewicz, 2006, p. 35-36, 38-39, 43-44] réaffirment (ce qui est moins visible dans d'autres travaux) que l'abduction mène à la plausibilité d'une hypothèse et non pas à la probabilité de celle-ci ; et qu'elle « laisse toujours place à une négociation, éventuellement à l'expression d'un doute ». Ils signalent l'importance de la dépendance que chacun des liens causaux hypothétiques entretient avec un ou des points de vue et avec le contexte. Ils notent qu'une recherche d'indices complémentaires peut être entreprise pour étayer cette hypothèse, ou des hypothèses qui pourront être concurrentes. Ils précisent qu'un raisonnement abductif part d'un faisceau d'indices convergents « dûment explicités ou explicites » pour remonter à l'hypothèse d'une cause dont la plausibilité se renforce d'autant plus que les indices abondent.

5.4.2 Comparaison des raisonnements formalisés

Selon [Degoulet et Fieschi, 1997, p. 49-50] :

- « Le raisonnement déductif est basé sur les principes d'implication logique »²⁶.
- « Le raisonnement inductif fait des généralisations basées sur des exemples spécifiques, pour formuler des règles générales »²⁷.
- « Le raisonnement abductif, souvent appelé la méthode scientifique, est une partie importante de la recherche scientifique. Nous tentons d'établir des liens entre des observations telles que la cause et l'effet. Des hypothèses peuvent aider à formuler une règle pour établir un lien entre la procédure et

²⁶Traduction de « *Deductive reasoning is based on the principles of logical implication* ».

²⁷Traduction de « *Inductive reasoning makes generalizations based on specific examples to formulate general rules* ».

5.4. LES RAISONNEMENTS FORMALISANT L'INCERTITUDE

les faits résultants (p. ex., une hypothèse diagnostique), ou bien elles peuvent concerner la formulation d'une nouvelle règle (p. ex., une découverte scientifique) »²⁸.

Les étapes de découverte scientifique utilisent des processus de raisonnement dont certains sont formalisés en logique. C'est le cas du *raisonnement déductif* et du *raisonnement inductif* qui sont les plus couramment mis en avant ; et c'est aussi le cas du *raisonnement abductif* (voir § 5.4.1, p. 104) :

- Le *raisonnement déductif* (prémisses \Rightarrow conclusion) s'appuie sur des propositions liées considérées vraies pour en déduire une autre proposition conséquente. Exemples :
 - « $A = B$, $B = C$, donc $A = C$ ».
 - « La Terre est une planète, les planètes sont des corps célestes. Donc la Terre est un corps céleste. ».
- Le *raisonnement inductif* (prémisses \Rightarrow conclusion) s'appuie sur des cas particuliers répétitifs pour en déduire une généralisation sans vérifier tous les cas. Exemples :
 - « Tous les cocons ouverts contenaient un papillon, donc le reste des cocons devrait aussi en contenir ».
 - « Les langues dont la grammaire a été étudiée ont au moins une façon d'exprimer le futur. Il est donc raisonnable de considérer que l'expression du futur est une caractéristique de toutes les langues ».
- Le *raisonnement abductif* (prémisses \Leftarrow conclusion) part du constat d'un ou plusieurs faits non élucidés, et s'appuie sur des propriétés de leur univers pour proposer des hypothèses pouvant expliquer ce ou ces faits. Semblablement à la logique d'une enquête policière, l'hypothèse qui est considérée en premier est celle qui est la plus vraisemblable (identifiée, décrite, filtrée, classée, choisie) selon l'expérience des scientifiques. Exemples :

²⁸Traduction de « *Abductive reasoning, often referred to as the scientific method, is an important part of scientific research. We attempt to establish links between observations such as cause and effect. Hypotheses may help formulate a rule to establish a link between the preceding and subsequent facts (e.g., a diagnosis hypothesis), or they may concern the formulation of a new rule (e.g., a scientific discovery).* ».

- « Le cobaye est allé délivrer son compagnon enfermé, au lieu d'aller manger une friandise proposée. Son comportement pourrait être associé à une empathie via des neurones miroirs ».
- « Le compteur électrique « intelligent » a enregistré un pic de 1000 Watts entre 12h00min00sec et 12h05min00sec. Une cause envisageable de ce pic est l'utilisation d'une plaque de cuisson de cuisinière électrique programmable pour réchauffer un déjeuner ».

Dans la recherche scientifique, le moyen de l'abduction permet donc de retourner la perspective de raisonnement :

- La déduction (via la logique) et l'induction (via la probabilité) sont orientées des causes vers la conclusion à l'aide d'un automatisme d'inférence.
- L'abduction est orientée (via la crédibilité liée à l'expérience du monde) de la conclusion vers une ou des causes plausibles, à l'aide de l'imagination d'une hypothèse qui pourra éventuellement être vérifiée.

5.4.3 Du raisonnement à l'incertitude

La déduction qui est porteuse de certitude d'une part, et l'induction et l'abduction qui sont porteuses d'incertitude d'autre part, n'épuisent pas forcément la diversité des raisonnements qui peuvent être distingués dans la réflexion scientifique ; même si, selon les diverses sources consultées, les autres formes de raisonnement peuvent être ramenées à ces trois-là. Par exemple, la créativité scientifique est liée au raisonnement par analogie [Gentner, 2002, p. 35] ; laquelle analogie fait partie du raisonnement inductif selon [Polya, 1954, chap. 2], et combine l'induction et l'abduction selon [Peirce, 1958, vol. 1 § 65 - n.b.p. n° 2] (voir § 5.4.1, p. 102). L'induction, l'abduction et l'analogie font partie des processus importants pour découvrir des hypothèses, [Cellucci, 1998, p. 11-13], bien que l'analogie ne soit pas une justification de ces hypothèses.

Le raisonnement par analogie peut se présenter sous deux formes, [Cellucci, 1998, p. 12-13] : soit la ressemblance en laissant volontairement de côté ce qui dans la connaissance tacite n'entre pas dans cette ressemblance, soit le partage de la présence ou absence de mêmes attributs. Le raisonnement par analogie peut donc s'exprimer par exemple sous la forme *A est à B comme C est à D* et sous la forme *A diffère de B comme C diffère de D*. Par exemple,

[Mikolov *et al.*, 2013, fig. 2], 'Berlin est à l'Allemagne ce que Paris est à la France'.

Nous constatons que des formalismes logiques permettent déjà le calcul automatique de certaines relations d'analogies, pour les vérifier ou les découvrir en se basant sur la similarité des contextes d'expressions linguistiques, et, par exemple chez [Ollinger, 2014, p. 2 et chap. 1], pour générer automatiquement de nouvelles connaissances. Pour leur écriture, ces formalismes utilisent des opérateurs logico-mathématiques tels que dans $A : B :: C : D$ ou $A - B = C - D$, ce qui permet de poser des questions telles que « 'Germany' : 'Berlin' :: 'France' : ? », ou « Germany - Berlin = France - ? » ou l'équivalent « Berlin - Germany + France = ? ». Ce qui donnera des réponses telles que « Germany - Berlin = France - Paris » ou « Berlin - Germany + France = Paris ». Ces méthodes peuvent aussi faire apparaître comme résultat des réponses qui divergent des attentes, par exemple « doctor - man + woman = nurse »²⁹, ce qui peut révéler ou éclairer des phénomènes factuels inattendus, inexplorés, peu approfondis ou en manque d'indices du point de vue scientifique. Il pourrait donc être intéressant d'identifier, classifier et codifier automatiquement des (sous-)expressions d'analogies exprimées dans les raisonnements, et de les faire fonctionner logiquement pour découvrir, proposer scientifiquement, et utiliser automatiquement des relations entre concepts. Cependant, le raisonnement par analogie n'a pas été décelé dans notre corpus de phrases (voir corpus en annexe G, p. 247) sous la forme d'expressions linguistiques autonomes des expressions linguistiques de l'induction, de l'abduction ou d'hypothèse : nous n'avons donc pas rencontré la nécessité d'intégrer l'analogie à l'ontologie de l'incertitude dans notre expérimentation. Identifier, classifier et codifier automatiquement les analogies des raisonnements scientifiques pourrait éventuellement faire l'objet de futures expérimentations.

En ce qui concerne les *hypothèses*, qui constituent une catégorie d'incertitude dans l'ontologie : c'est dans la pratique de l'expérimentation, en observant les expressions d'incertitude scientifiques, que les *hypothèses* sont apparues en tant que troisième forme de raisonnement : les expressions linguistiques d'hypothèses étaient suffisamment distinguables et pas toujours associées aux expressions d'induction ou d'abduction dans une même phrase.

Nous n'excluons pas a priori que d'autres formes de raisonnements existent, que nous n'aurions pas décelé chez des auteurs ou dans le corpus, et qui soient associés à des phénomènes d'incertitude.

²⁹ *doctor* était attendu au lieu de *nurse*, du point de vue « doctor - man = doctor - woman ».

5.5 Proposition d'une ontologie de l'incertitude

Nous avons étudié les différentes relations entre l'incertitude et les objets textuels et informationnels, et construit une première représentation des relations de l'incertitude. Le schéma 5.2 présente ces relations sous forme d'arbre.

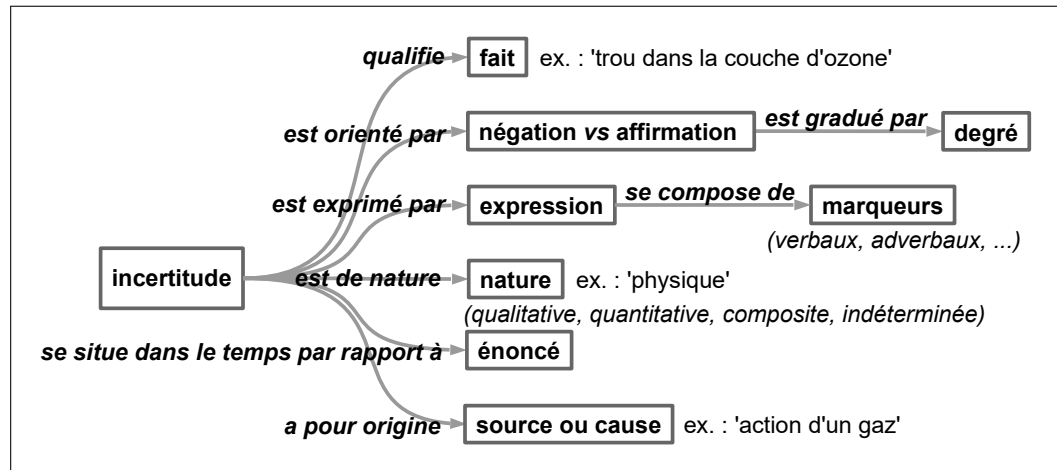


Schéma 5.2: Caractéristiques de l'incertitude scientifique.

Enfin, nous avons construit une ontologie de l'incertitude en prenant en compte les types de raisonnement formalisés pouvant induire de l'incertitude et les réflexions exposées dans les sections précédentes. Le schéma 5.3 présente cette ontologie, avec les deux principaux types d'incertitude : qualitative et mesurée.

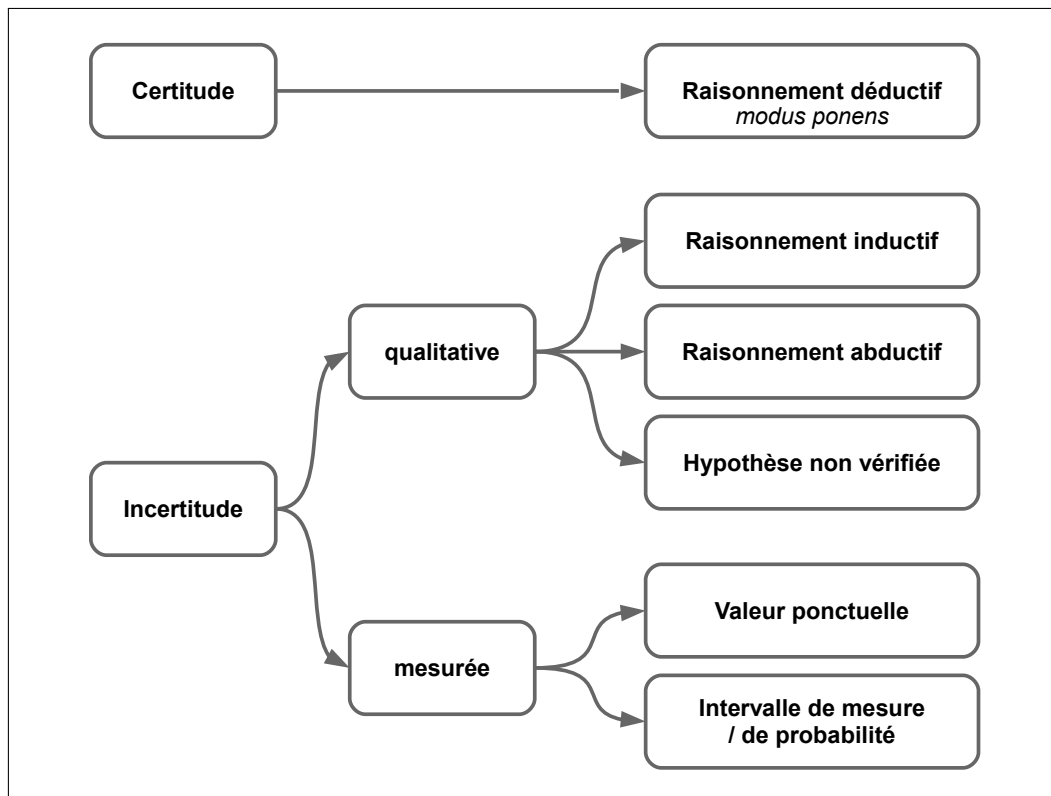


Schéma 5.3: Ontologie de l'incertitude.

Chapitre 6

Constitution de corpus pour le traitement de l'incertitude

Dans ce chapitre nous décrivons le processus de constitution du corpus de 4916 articles scientifiques dans le domaine du changement climatique. Ce corpus d'articles va ensuite être utilisé pour constituer d'une part le corpus d'étude et le corpus d'évaluation.

6.1 Principales étapes du travail

Les étapes les plus significatives structurant notre travail et méthodologie sont les suivantes (voir schéma 6.1) :

1. Constitution du *corpus d'articles scientifiques* liés au changement climatique, sous la forme de deux corpus complémentaires :
2. Analyse des expressions de l'incertitude d'une partie des articles choisis manuellement pour leur diversité (de domaines scientifiques et de nombres de pages).
3. Création de l'ontologie de l'incertitude scientifique dont les catégories d'incertitude mesurée et qualitative peuvent s'appliquer à tous les domaines scientifiques.
4. Constitution du *corpus de 1000 phrases* contenant des expressions-candidates de l'incertitude, issu des trois corpus suivants :
 - corpus de 500 phrases extraites automatiquement du corpus d'articles scientifiques à l'aide d'expressions régulières.

6.1. PRINCIPALES ÉTAPES DU TRAVAIL

- corpus de 500 phrases extraites manuellement des 25 premiers articles du corpus d'articles scientifiques publiés à partir de janvier 2000. Parmi ces phrases, 200 seront utilisées pour le corpus d'étude, et 300 pour le corpus d'évaluation.
- 5. Annotation manuelle du corpus de 1000 phrases par trois experts.
- 6. Proposition du *gold standard* basé sur le corpus de phrases annotées.
- 7. Composition de la *grammaire formelle* de règles de description linguistique d'expressions textuelles complexes.
- 8. Rédaction d'un ensemble de *règles* d'annotation. Intégration de ces règles au système d'annotation de l'incertitude.
- 9. Évaluation du *système* sur le corpus d'évaluation de 300 phrases. Calcul des mesures de Précision, Rappel et F-mesure.

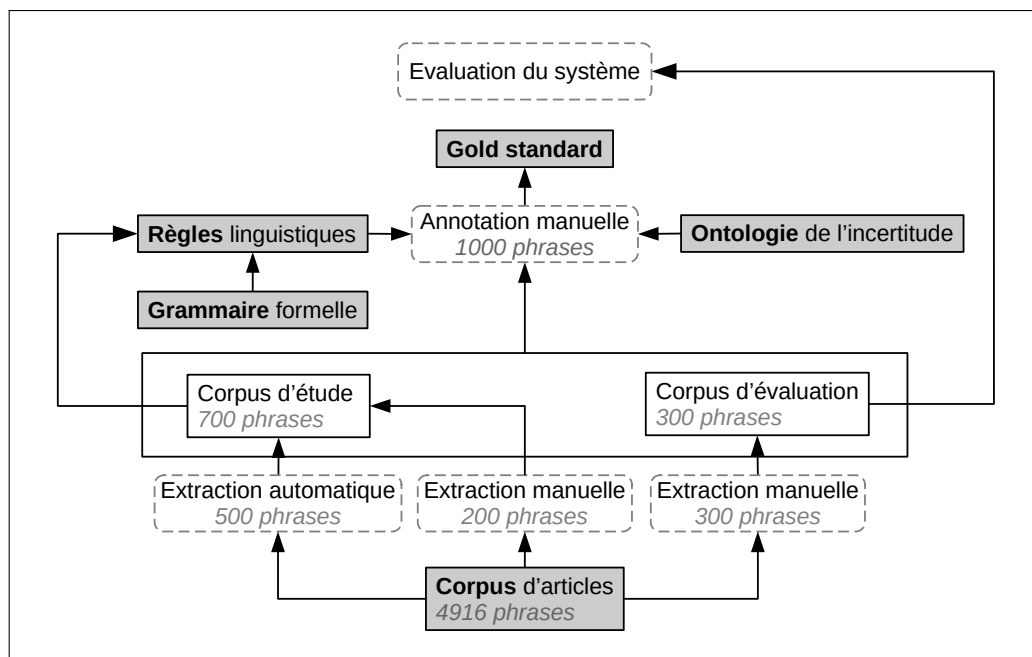


Schéma 6.1: Etapes principales de la méthodologie.

6.2 Contexte de l'incertitude : les articles scientifiques

Notre objectif est d'étudier l'expression de l'incertitude dans les productions scientifiques textuelles et notamment dans les articles scientifiques. Il est donc utile de commencer par présenter les caractéristiques générales de cette matière première que sont les articles scientifiques.

La science, du latin *scire* « savoir »¹, qui est un ensemble de connaissances cohérentes, ouvertes au doute et à la critique, et obtenues de la recherche méthodique, s'est co-construite avec la communication de plus en plus large de ses connaissances, sous la forme d'articles dans des journaux scientifiques publiés depuis le XVII^e siècle. C'est actuellement principalement par le biais de ces publications devenues la norme - les articles scientifiques - que la science se met à disposition de la communauté scientifique.

Un article scientifique n'est pas l'énoncé d'une suite de faits ou de raisonnements déductifs. Il constitue le vecteur principal de communication des résultats des réflexions scientifiques, et s'inscrit dans un contexte d'autres résultats scientifiques eux aussi reconnaissables dans des articles. Un article scientifique contient également des interprétations des chercheurs, des hypothèses, des pistes pour de futures recherches, etc. Plus largement, le processus scientifique implique la prise en compte d'une certaine incertitude : la plupart des résultats sont accompagnés d'une incertitude pouvant être liée aux intervalles de confiance des outils utilisés, à l'incomplétude des données, aux hypothèses à vérifier, ainsi qu'aux débats relatifs aux cadres théoriques et aux paradigmes en vigueur.

Parmi les documents issus de la recherche scientifique² les articles jouent un rôle privilégié dans le processus de transmission des connaissances scientifiques. Les premières revues scientifiques, vecteurs de diffusion des articles, apparaissent en 1665, [Feyel, 2018]. [Solla Price (de), 1965, p. 562] observe que leur accumulation depuis le milieu du XVII^e siècle fait des articles scientifiques des atomes de connaissance. Il distingue deux moitiés de l'activité de recherche : d'une part la science, qui inclut des sujets tels que l'électronique, l'ingénierie informatique et

¹Bloch Oscar et Wartburg (von) Walther (1989), « Dictionnaire étymologique de la langue française » 8^e édition. Presses Universitaires de France, Paris.

²Articles, thèses, rapports de recherche, manuels, livres d'introductions ou divulgation, diaporamas et comptes-rendus de conférences, posters de congrès, entrevues à la presse, carnets de laboratoire, brevets, textes et vidéos de cours, encyclopédies, monographies, notes de synthèse, ...

6.2. CONTEXTE DE L'INCERTITUDE : LES ARTICLES SCIENTIFIQUES

la chimie industrielle, et d'autre part la « technologie »³, pour mieux affirmer que c'est la science qui a pour produit final les articles de recherche (« *papers as an end product* »).

La scientométrie est la science qui a pour vocation d'évaluer la science pour aider à la gérer⁴. En scientométrie, l'article scientifique est un matériau de base pour analyser ce que produit la science. Même s'il différencie le développement des connaissances de celui des publications scientifiques, [Polanco, 1995, partie 1.1] insiste sur ce qu'il appelle le point de vue du réductionnisme bibliométrique en scientométrie :

« l'article scientifique devient un outil de définition de la science et l'on fait de la publication écrite un indicateur privilégié de l'activité scientifique, considérant que le produit final de la recherche scientifique est la publication d'un texte écrit (articles scientifiques, contributions aux colloques, rapports ou toute autre espèce de "littérature grise") ».

Aujourd'hui, publier des articles est aussi une obligation institutionnelle pour le début, le maintient et l'évolution des carrières scientifiques. La science s'industrialise et les chercheurs (ainsi que leurs institutions) sont incités à participer à une course de maximisation d'un indice personnel de rendement et réputation. Pour cela, ils doivent publier le plus possible, en se faisant citer le plus possible, et dans des revues les mieux cotées possibles du point de vue de leur *facteur d'impact*. Il est notable que cette pression se généralise au niveau mondial en générant des effets pervers, tels que le manque d'originalité des articles, des pratiques manquant de rigueur ou de fondements scientifiques, la prévalence monopolisante de la langue anglaise et l'existence de revues prédatrices. Les mesures actuelles d'évaluation de la recherche (facteurs d'impact ou autres mesures) sont controversées⁵, sur leur formulation, leurs utilisations et leurs effets. Malgré cela, il n'en reste pas moins que communiquer en publiant sous le format d'un article, au delà de permettre l'évaluation et l'utilisation des résultats de la recherche par les pairs, sert à la fois à consolider l'évolution de la Science et des sciences

³Les guillemets sont de cet auteur.

⁴Inspiré, entre autres, de [Courtial, 1990, Introduction, p. 7] : « La bibliométrie est à la fois le comptage de tout ce qui peut entrer dans une bibliothèque scientifique et une approche quantitative des techniques de gestion d'une bibliothèque. La scientométrie est la généralisation de ces techniques [...] dans le but d'une gestion de l'activité de recherche scientifique. ».

⁵Par exemple par Hirsh, inventeur de l'indice *h*, [Hirsch, 2005], qui en est venu à critiquer lui-même cet indice, [Hirsch, 2020, p. 4].

et à fonder la condition scientifique des auteurs de la publication, ce que rappelle [Hyland, 1996b, p. 435] en parlant de l'accomplissement « des buts à la fois institutionnels et individuels »⁶.

Les documents scientifiques se citent les uns les autres et sont ainsi organisés en réseau ; et les références bibliographiques qui sont les liens de ce réseau se comptent en milliards⁷. La distribution des articles, qui se faisait via des documents en papier, se convertit actuellement en circulation de flux digitaux. Ces pratiques sociales et changements technologiques permettent l'émergence de nouvelles dimensions et propriétés du réseau, et des sous-réseaux, des informations scientifiques.

Les supports d'articles scientifiques continuent leur mue vers des formats de publication numériques, qui sont de mieux en mieux adaptés à une utilisation automatique. A propos du rôle « central » des périodiques scientifiques électroniques, [Accart et Rivier, 2012, p. 61] expliquent que leur succès d'adoption par les institutions académiques et de recherche fait que celles-ci « ne peuvent plus y renoncer. Peu à peu, les collections papier cèdent la place aux collections électroniques ».

La transformation des formats des articles s'accompagne aussi d'une augmentation de leur disponibilité par leur mise en ligne. C'est particulièrement le cas des publications de la recherche en accès ouvert (ou *open access* même en français). A présent en expansion rapide par nécessité scientifique et économique, le développement récent de ce mode de publication ouvert a bénéficié non seulement des évolutions de la technologie, mais aussi des choix des chercheurs de certains domaines scientifiques, par exemple les biologistes en France, [Boukacem-Zeghmouri *et al.*, 2018, p. 7-8]. Le mode de publication ouvert influence aussi des dispositions législatives et des appuis⁸ et décisions institu-

⁶« La publication de résultats scientifiques cherche à accomplir des buts à la fois institutionnels et individuels. Un article de recherche non seulement étend la compréhension de phénomènes et de théories que le paradigme en cours juge dignes d'étude, mais aussi aide à soutenir ou à établir la réputation personnelle de l'auteur », ou, « *The publication of scientific results seeks to accomplish both institutional and individual goals. A research paper not only extends understanding of phenomena and theories that the current paradigm deems worthy of study, but also helps support or establish the personal reputation of the writer* ».

⁷Par exemple, la plateforme en ligne *Web of science* de la société Clarivate propose un accès à presque 1,9 milliards de références bibliographiques : « *Our comprehensive platform allows you to track ideas across disciplines and time from almost 1.9 billion cited references from over 171 million records* ». Consulté le 23/09/2020 sur <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>.

⁸[Boukacem-Zeghmouri et Rodríguez Bravo, 2019, p. 6] décrivent une mutation vers des *injonctions* françaises, européennes et internationales à publier en accès ouvert.

6.2. CONTEXTE DE L'INCERTITUDE : LES ARTICLES SCIENTIFIQUES

tionnelles allant dans le sens d'une généralisation de l'accès libre aux publications de recherche, par exemple avec le programme Horizon 2020 dans l'Union Européenne. D'un point de vue des politiques scientifiques, l'ouverture des archives est favorisée par :

- les actions des chercheurs considérant que pour exister la science doit être publique. Par exemple, c'est en ce sens que la « Déclaration de Berlin sur le Libre Accès à la Connaissance en Sciences exactes, Sciences de la vie, Sciences humaines et sociales » (« *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* »), publiée en 2003⁹ :

« l'Internet nous offre la possibilité de constituer une représentation globale et interactive de la connaissance humaine [...], et la garantie d'un accès mondial. [...] Nous définissons le libre accès comme une source universelle de la connaissance humaine et du patrimoine culturel ayant recueilli l'approbation de la communauté scientifique. [...] Les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et de métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédia. [...] sous un format électronique approprié ».

- de décisions de législateurs. Par exemple, le code de la propriété intellectuelle a été modifié¹⁰ en France en 2016 pour mieux permettre, dans le cadre de recherches non-commerciales, l'exploration des textes scientifiques et de leurs métadonnées.
- de mises en œuvre institutionnelles favorisant une science ouverte. Par exemple en 2018 en France le « Plan national pour la science ouverte » en appui de la diffusion « sans entrave des publications et des données de la recherche »¹¹.

⁹Disponible sur : https://openaccess.mpg.de/68042/BerlinDeclaration_wsis_fr.pdf (version originale en anglais sur : https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf).

¹⁰LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique, chapitre II : Economie du savoir, article 38, disponible sur https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000033202882.

¹¹Ministère de l'enseignement supérieur, de la recherche et de l'innovation (2018), « Plan

L'Office Français de l'Intégrité Scientifique (OFIS) rapporte¹² que la pratique d'une science ouverte, en plus d'être un constituant inséparable de l'intégrité scientifique, est déjà considérée par certains comme un changement épistémologique : notamment parce qu'elle permet d'intégrer des savoirs de lieux plus divers, de trouver ses sources dans des langues et des collaborations plus diverses, de mettre en œuvre des partages et des participations ; et parce que la transparence permet la reproductibilité des expériences. Il souligne que cela permet de s'engager vers de nouveaux défis scientifiques, mais avertit sur de nouveaux risques liés à la maîtrise des idées et des données : le libre accès doit se prémunir contre des réutilisations abusives ou partisans, et la gestion du cycle des données doit prendre en compte les droits tels que la rémunération lors de la production et le respect de la vie privée.

Cette offre en ligne d'articles scientifiques de plus en plus massive, permet d'améliorer les usages de ressources scientifiques. Par exemple, cela favorise une bibliographie atteignable, dont la nécessité est exposée par [Accart et Rivier, 2012, p. 139] :

« L'information bibliographique doit assurer qu'aucune recherche entreprise ne puisse ignorer les apports précédents, ce qui requiert non seulement que la littérature primaire soit recensée de manière exhaustive, mais aussi qu'elle puisse être retrouvée efficacement ».

[Polanco, 1995, partie 2.8] note qu'il est possible d'échelonner les articles scientifiques sur plusieurs classes en fonction du nombre de références bibliographiques qu'ils contiennent, dont :

- « *articles de synthèse (review)* qui condensent et donnent accès à la littérature récente dans un domaine ou spécialité » ;
- « *articles de recherche* qui rendent compte des résultats d'une expérience de laboratoire, ou d'une enquête, ou des observations de terrain » ;

national pour la science ouverte ». Disponible sur https://www.ouvrirelascience.fr/wp-content/uploads/2018/08/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf. Pour suivre la progression des publications en accès ouvert, un Baromètre français de la Science Ouverte est disponible sur <https://www.enseignementsup-recherche.gouv.fr/cid148931/barometre-francais-de-la-science-ouverte.html>.

¹²Office Français de l'Intégrité Scientifique - OFIS (décembre 2018), « Feuille de route 2020 de l'intégrité scientifique / L'intégrité Scientifique en France : Situation fin 2018 et feuille de route 2020 », p. 16-17. Disponible sur <https://www.hceres.fr>.

6.2. CONTEXTE DE L'INCERTITUDE : LES ARTICLES SCIENTIFIQUES

- « *articles ex cathedra* » qui présentent un nombre nul ou négligeable de références explicites.

Les disciplines scientifiques adoptent pour leurs articles des standards de publication qui sont prescrits dans les revues scientifiques où les chercheurs doivent publier. C'est notamment le cas de la structure *IMReD*¹³ (Introduction, Méthodes, Résultats et Discussion). Cette structure, base d'une norme facultative¹⁴ publiée aux EUA par le [Sous-comité 26 du Comité Z39 de l'ANSI, 1979] en 1972 et (révisée) en 1979, est parfois qualifiée d'universelle¹⁵ car elle s'impose dans beaucoup de domaines scientifiques, par exemple dans ceux de la médecine d'où est parti son développement [Sollaci et Pereira, 2004], [Jacques, 2017b, p. 99], [Jacques, 2017a, p. 102]. L'International Committee of Medical Journal Editors (ICMJE) affirme dans ses recommandations « *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals* »¹⁶ :

« La structure "IMReD" n'est pas un format arbitraire de publication, mais reflète le processus de découverte scientifique ».¹⁷

La structure *IMReD* est aussi présente dans un ordre qui peut varier selon les articles, [Burrough-Boenisch, 1999, p. 304-305], [Atanassova *et al.*, 2018, § 3.1]. Cette caractéristique d'un ordre adaptable des parties est mise en avant par [Wu, 2011, p. 1347] comme avantageuse pour la créativité et l'innovation, en remarquant que la structure *IMReD* est faite en premier lieu pour des articles de recherche originaux. Considérer les divers choix d'ordonnancement de la structure renvoie à une réflexion sur quels seront les usages d'un article qui seront privilégiés [Wu, 2011, p. 1348-1349]. Suivant [Pontille, 2007, p. 3-5, 7, 9-12 et 14 (version article)], la structure *IMReD* est aussi une convention de mise en scène et mise en lien scientifiques avec des objets et substances, personnes et institutions, articles et autres textes, tableaux, figures, références, etc., cadrés dans un article.

¹³Souvent notée avec la version anglaise de ce même acronyme : *IMRAD*.

¹⁴[Blum, 1975, p. 114] : « la mise en conformité avec les normes [du Comité Z39] est volontaire » et reflète un consensus.

¹⁵Jourdain Maud (2006), « Plan de présentation d'un travail de recherche ». Document de travail. Disponible sur : www.cogemspc.fr/theses/guide_redac.pdf.

¹⁶Alias « *ICMJE Recommendations* », alias « *The Uniform Requirements for Manuscripts* » ou « *URMs* ». Version de décembre 2017 consultée. Disponible sur : www.icmje.org/icmje-recommendations.pdf.

¹⁷Traduction de « The "IMRAD" structure is not an arbitrary publication format but a reflection of the process of scientific discovery ».

La structure IMReD est aussi parfois déclinée suivant des variations de son acronyme, par exemple *Introduction, Materials*¹⁸ *and Methods, Results, and Discussion*, et suivant des sous-structurations qui maintiennent son sens premier, comme le présentent [Shane (Mc) et al., 2005, p. 1181-1182]. La place centrale de cette structure permet que lui soient adjointes d'autres structures optionnelles selon les évolutions qui sont jugées améliorer ses utilités, telles que « résumé, mots-clés, remerciements, et références », et pouvant à leur tour s'enrichir de sous-structurations [Wu, 2011, p. 1347-1348]. Cette structure est proche d'autres structures liées à différents types de produits scientifiques, tel que, par exemple, le rapport de laboratoire *objectif, méthode, résultats, implications*, [Skitmore et Runeson, 1996, p. 5].

D'autres structures d'articles existent, qu'elles soient constatées, préconisées ou prescrites. Par exemple, [Devillard et Marco, 1993, p. 38] indiquent les suivantes, en plus d'IMReD pour des articles analytiques, dont les acronymes des quatre dernières « sont de [leur] fait » :

- OPERA : *Observation, Problème, Expérimentation, Résultats et Action*.
→ particulièrement pour des articles analytiques en sciences appliquées.
- ILPIA¹⁹ : *Introduction, Littérature, Problème, Implication, Avenir*
→ pour des articles de synthèse et des *surveys*.
- VO : *Vie et Œuvre*.
→ pour des biographies et des nécrologies.
- TAS : *Thèse, Antithèse, Synthèse*.
→ pour les controverses et les polémiques.
- RAP : *Résumé des textes, Axes communs, Perspectives ouvertes*.
→ pour des introductions, des conclusions de numéros spéciaux et des notes de lecture.

Cependant, même si dans leur structuration tous les articles scientifiques ne suivent pas toujours ou pas au pied de la lettre ces standards, la structure IMReD est particulièrement typique dans la rédaction d'articles scientifiques. S'appuyer sur cette structure est donc un choix pratique dans la recherche sur et dans les

¹⁸*Materials* peut aussi apparaître sans *Methods* pour traduire l'acronyme, par exemple dans [Lindsay, 2011, p. 10].

¹⁹Ou juste *ILIA* selon [Pochet, 2015, p. 121].

6.2. CONTEXTE DE L'INCERTITUDE : LES ARTICLES SCIENTIFIQUES

textes scientifiques, comme le met en avant [Sterken, 2011, p. 18-19] en soulignant l'aptitude particulière de la structure pour le transfert de connaissances scientifiques sur des faits, quelque soit l'ordre adopté des sections de son acronyme ; et comme le prennent en compte, par exemple, [Atanassova *et al.*, 2016] dans leur expérimentation.

Au delà de ces standards de publication coté auteur, il est remarquable que les standards de publication coté éditeur évoluent aussi depuis des années. C'est le cas en particulier dans des revues spécialisées qui présentent ainsi les articles et leurs parties sous divers formats dérivés à partir d'une même trame de document contenue dans un seul fichier XML [Ginsparg, 2009]. Phénomène aussi remarquable, les solutions logicielles *open source* de publications scientifiques qui utilisent ces dialectes XML sont de plus en plus utilisées par des éditeurs de revues scientifiques en accès libre. C'est notamment le cas en France avec le système universitaire d'édition Métopes²⁰ qui a commencé à se répandre dans d'autres pays et utilise le XML-TEI (XML *Text Encoding Initiative*), ou avec le système de gestion de contenus textuels scientifiques Lodel²¹ qui utilise le XHTML. Ces solutions contribuent à l'appropriation et à la définition des standards de publication par les auteurs scientifiques et par les équipes de rédaction scientifiques de ces revues.

L'accroissement du nombre et de la proportion de publications qui participent à la science ouverte, l'adoption de standards partagés dans des proportions significatives, l'emploi croissant de logiciels libres communs par des plateformes publication en ligne, la consolidation d'archives universitaires d'articles de recherche, l'adoption de plateformes publiques comme HAL²² qui encouragent qualitativement et quantitativement les échanges de données de référencement, la communication de la communauté de chercheurs spécialistes de l'édition scientifique vers les autres chercheurs universitaires, ouvrent ensemble un éventail d'opportunités pour mettre en place et généraliser des services aux scientifiques autour des publications scientifiques via le traitement automatique des langues et des données. C'est par exemple le cas des services liés au web sémantique qui reposent en particulier sur les ontologies informatiques, décrits notamment par [Shotton, 2009] et [Peroni, 2014] : *augmentation* des articles, visualisation d'informations sémantiques extraites des textes et leurs métadonnées, proposition de formats sup-

²⁰Site officiel : http://www.unicaen.fr/recherche/mrsh/document_numerique/projets/metopes.

²¹Site de référence : <https://www.lodel.org>.

²²<https://hal.archives-ouvertes.fr>

plémentaires, résumés automatiques, accès à des jeux de données, gestion de références, etc. ; c'est à dire l'installation d'un écosystème sémantique pertinent pour la recherche sur mesure avec chaque publication consultée.

Au départ de notre approche des expressions de l'incertitude dans les articles scientifiques nous avons commencé à constater, [Atanassova *et al.*, 2018, § 3.2], d'une part que ces expressions peuvent être présentes à divers degrés dans toutes les grandes parties structurant le texte ; et d'autre part que la répartition de ces expressions dans ces parties structurantes dépend du domaine scientifique des articles. Puisque le corpus que nous avons ensuite constitué avec des articles concernant le changement climatique est composé d'articles de domaines scientifiques remarquablement divers, nous y avons étudié les expressions dans l'ensemble de chaque texte, plutôt que celles d'une section en particulier.

6.3 Choix du changement climatique comme domaine d'étude

Lors de nos recherches préliminaires sur les études de traitement automatique des langues liées au thème de l'incertitude, nous avons particulièrement remarqué la présence de deux domaines déjà largement abordés : celui de la santé humaine et celui du renseignement militaire ou civil.

Au début de cette thèse, nous avons mené une expérimentation [Atanassova *et al.*, 2018], sur un corpus de près de 10 000 articles scientifiques constitué à partir de *PubMed Open Access Subset*²³ : 9 463 articles en biomédecine, et 488 articles en physique. Dans cette première expérimentation, le corpus a servi à évaluer la distribution des expressions de l'incertitude dans les sections principales (introduction, méthodes, résultats, discussion) structurant les articles. C'est en particulier cette expérimentation qui a été l'occasion de remarquer des prises en compte et des références liées de l'incertitude dans des articles autour du changement climatique. Cela nous a permis de confirmer le fait que le changement climatique constitue un domaine doté d'un intérêt particulier pour le traitement de l'incertitude. En poursuivant la réflexion, nous avons observé que la composante d'incertitude scientifique du domaine du changement climatique fait partie d'une problématique prégnante d'actualité, et est transversalement présente dans des articles de domaines scientifiques très divers et souvent éloignés, ce qui

²³*Pubmed OA subset corpus*, composé de 1,7 millions d'articles disponibles en XML. URL : <https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>

oblige à considérer l'incertitude scientifique d'une manière globale ou généralisable.

À première vue, le domaine du changement climatique est comparable au domaine biomédical dont l'étude par le TAL est déjà très développée ; c'est remarquable en particulier si l'on considère que le domaine du changement climatique traite des (dys)fonctionnements de la biosphère terrestre considérée comme un organisme vivant, suivant l'hypothèse scientifique²⁴ *Gaïa* du médecin, chimiste et ingénieur James Lovelock ; lequel point de vue à influencé, entre autres et surtout depuis les années 80, les sciences de la Terre [Dutreuil, 2018, p. 2]. Cependant, nous avons remarqué que la diversité et l'éloignement des domaines scientifiques qui traitent des aspects du changement climatique sont particulièrement importants, ce qui semble différencier le corpus de ces domaines de celui du corpus biomédical, qui lui est composé de sous-domaines biomédicaux plus contigus. Un intérêt prévisible de l'utilisation d'un corpus du changement climatique est donc de devoir faire autrement, plus globalement, pour traiter les catégories d'incertitude scientifique.

Un autre motif pour cette démarche de mise en valeur des données scientifiques sur la thématique du changement climatique est peut-être le plus important : il s'agit du pouvoir des scientifiques, et donc de leur responsabilité, concernant le fondement des informations, compréhensions, volontés, décisions et actions climatiques. La juriste Dinah Shelton [Shelton, 1998, p. 41-43], explique que la science est un moyen d'action positif, non seulement pour connaître les dommages environnementaux qui sont la conséquence d'actions humaines, mais aussi pour les limiter. Elle signale que c'est notamment sur une prise en compte de l'incertitude scientifique dans l'évaluation de « la forme, la magnitude et les caractéristiques » du risque que peut se baser la politique de l'environnement pour choisir parmi les options de réduction du risque :

Ce choix dépendra qualitativement et quantitativement des données scientifiques servant de base à l'évaluation et aux options », « L'évaluation du risque [...] doit aboutir à un jugement concernant l'importance du problème, prenant en compte les incertitudes inévitables », « l'information sur l'incertitude et la publication de différentes analyses pourraient rendre les institutions plus prudentes ».

²⁴Hypothèse scientifique qu'il faut distinguer des croyances et visions artistiques, par exemple la croyance religieuse en la *Pachamama* dans les Andes, et la fiction littéraire *Solaris* de Stanislas Lem (Pologne, 1961) adaptée cinématographiquement par Andreï Arséniévitch Tarkovski (Russie, 1972) puis par Steven Soderbergh (EUA, 2002).

Le changement climatique, c'est-à-dire le fait et les conséquences de l'actuel réchauffement climatique global et rapide à l'échelle de la vie humaine, est reconnu ou appréhendé au delà de la connaissance scientifique, par exemple par sa prise en compte au niveau d'institutions économiques internationales²⁵ ²⁶, au niveau d'agences gouvernementales en charge de la sécurité²⁷, et au niveau de responsables et organisations politiques. Cependant, hors du champs scientifique les faits, preuves et tendances du changement climatique risquent l'incompréhension et la controverse, et [Callon *et al.*, 2001, § Précaution et « scénarios du pire »] indiquent entre autres causes les dommages créés par les journalistes qui « [jouent] avec les affects du public » avec des « formulations [...] outrancières » qui ne correspondent pas aux scénarios climatiques envisagés et décrits par les scientifiques. [Sluijs (van der) *et al.*, 2008, p. 1-2] parlent par exemple de *science contestée* dans des processus de décisions politiques concernant des risques environnementaux complexes, et de connaissance *contestée* ou *robuste* selon que son acceptabilité est partagée non seulement entre scientifiques, mais aussi entre scientifiques et politiques (et autres parties prenantes et société civile) ; et ils insistent sur le fait que l'analyse des incertitudes est un problème essentiel. Dans ce même cadre, [Sluijs (van der) *et al.*, 2006b, p. 5-6,10-11] parlent d'inclusion de pairs non-scientifiques, d'une science dite post-normale enrichie par le dialogue, liée à l'éthique, à la politique et aux sciences sociales, et liée à la volonté de réconcilier les différents points de vue ; en bref, une Science impliquée dans la société et ses enjeux, ce qui engage le sens de la Science elle-même, et où il faut « Intégrer l'incertitude dans les débats sur les futurs possibles ».

Des savoirs scientifiques à l'action politique les relais ne sont pas directs, [Schlaufer, 2016] : la mise en place de politiques qui s'appuient sur la connaissance et la compréhension des faits et des preuves scientifiques dépend, à la fois, d'une part du discours scientifique qui présente les preuves, et d'autre part du discours politique où ce discours scientifique sera inséré, cadré et orienté (« *the different narrative uses of evidence of opposing coalitions* »). Ainsi, le discours

²⁵Par exemple : World Bank (2019), « Action Plan on Climate Change Adaptation and Resilience - Managing Risks for a More Resilient Future ». Disponible sur <https://www.worldbank.org/en/topic/climatechange>.

²⁶Par exemple : le Fonds monétaire international qui diffuse l'idée que l'incertitude est un « concept nébuleux » et que sa croissance actuelle est liée entre autres à des phénomènes tels que la croissance du produit intérieur brut et des entreprises, et « les élections, les guerres et le changement climatique », [Ahir *et al.*, 2020].

²⁷Par exemple : ODNI (2016), « Implications for US National Security of Anticipated Climate Change ». Memorandum. Office of the Director of National Intelligence, Washington, EUA. Disponible sur : <https://www.dni.gov/index.php>

scientifique peut être pris en compte de manière appropriée ou inappropriée, ou récupéré de manière partisane dans des points de vue ou des stratégies préexistants pour tenter de justifier ou dénigrer²⁸ des choix de solutions. La connaissance du discours scientifique sur le climat peut aussi être délibérément restreinte, [Jones, 2014, p. 644-645].

La difficulté de l'information sur le changement climatique a des conséquences au delà de la science puisque les conclusions scientifiques indiquent que l'enjeu immédiat des actions humaines qui influencent le climat est celui de la réduction ou disparition des populations d'espèces vivantes sur Terre, de l'effondrement de la diversité du vivant qui permet l'existence humaine, et qu'en même temps²⁹ il est difficile « d'appréhender les questions environnementales d'un point de vue politique ». Cette difficulté qu'ont les responsables politiques avec le changement climatique est particulièrement liée à l'incertitude³⁰ dans ce domaine, d'autant plus que, [Smith *et al.*, 2018], les choix politiques eux-mêmes ont une incidence remarquée sur l'expansion complexe des incertitudes des scénarios climatiques (*cascade of uncertainty in scientific modeling*).

Actuellement, une prise en compte correcte de l'incertitude scientifique sur les mécanismes et risques climatiques a des conséquences sur les stratégies d'attention, de prévention, d'atténuation et d'adaptation au *dérèglement climatique* de la Terre, allant de la « surveillance et protection de la faune et de la flore d'environnements et écosystèmes uniques »³¹, en passant par la « défense verte »³² militaire, jusqu'à « Prévenir le développement anarchique des solutions

²⁸[Schlaufer, 2016, p. 13] : « *evidence of recent cooling of the global temperature is invoked to deny the climate change problem* ».

²⁹Bignon Jérôme (2019), « Examen de la note scientifique Biodiversité : extinction ou effondrement », p. 2-5 in *Assemblée nationale (2019), « Compte rendu de séance n° 34 de l'Office parlementaire d'évaluation des choix scientifiques et technologiques de l'Assemblée nationale »*. Disponible sur <http://www.assemblee-nationale.fr/15/cr-oecst/18-19/c1819034.asp>.

³⁰Par ex., [Lim *et al.*, 2004, p. 201, § A.8.1.6] signalent à propos des méthodes pour décider la priorité, sélectionner et adapter les politiques et les mesures à prendre face au changement climatique : « *Climate change is a process that is characterised by a number of uncertainties relative in particular to the magnitude, timing and nature of the changes. Decision makers are more familiar with processes/problems that are not subject to this degree of uncertainty* ».

³¹Traduction de : Cherlet M. *et al.* (2018), *World Atlas of Desertification*, 3^e éd., p. 232. Publication Office of the European Union, Luxembourg. Disponible sur <https://wad.jrc.ec.europa.eu/download>

³²M. le Ministre des affaires étrangères et du développement international du Gouvernement français (2015), « Réponse du Ministère des affaires étrangères et du développement international » in *Question d'actualité au gouvernement n° 0595G*. Journal Officiel du Sénat du 16/10/2015, p. 9658.

de géo-ingénierie »³³. Au delà, la correcte prise en compte de la connaissance scientifique du changement climatique a vocation à contribuer à l'amélioration d'autres types de connaissances et activités humaines, liées par exemple à l'étude et ingénierie de la vie dans les biosphères. Par exemple, la connaissance scientifique du changement climatique est déjà liée à la compréhension du climat des exoplanètes habitables³⁴ ; et, selon [Beaulieu *et al.*, 2019, p. 104, 112], ce sujet de l'habitabilité d'exomondes est fédérateur parmi les scientifiques, notamment en France, et participe à la création scientifique de « l'écologie planétaire ».

L'orientation de cette démarche de recherche vers les domaines scientifiques qui traitent du changement climatique, a pour but d'en maximiser l'utilité en cherchant à prendre en compte la remarque épistémologique de [Barreau, 2013, p. 123] :

« la science est une aventure qui, au même titre que toute aventure humaine, n'est en définitive justifiée que par la noblesse évidente de ses buts ».

Nous n'avons pas trouvé de corpus d'articles scientifiques constitué uniquement sur le concept du changement climatique (*climate change*), ni sur des concepts parfois employés comme équivalent tels que réchauffement climatique (*climate warming*) ou dérèglement climatique (*climate disruption*). Pourtant, selon [Masson-Delmotte, 2020, p. 258] qui se base sur des données de Web of Science, les publications dans des journaux à comité de lecture correspondant au mot clef « *climate change* » sont passées de centaines par an en 1990 à des dizaines de milliers par an actuellement.

Nous avons donc décidé de constituer notre propre corpus d'articles scientifiques liés au changement climatique, quels que soient les domaines scientifiques dans lesquels s'expriment leurs auteurs, afin d'y étudier l'incertitude scientifique. Ce choix thématique répond à des priorités scientifiques actuelles : par exemple, le changement climatique fait notablement³⁵ partie des thématiques prioritaires où des scientifiques de tous horizons désirent engager leur responsabilité.

³³Perrin C. *et al.* (2015), Rapport sénatorial sur les conséquences géostratégiques du dérèglement climatique, p. 80-82 et 185. Sénat français, Paris. Disponible sur <http://www.senat.fr/rap/r15-014/r15-0141.pdf>

³⁴« Pour modéliser le climat sur les exoplanètes, nous utilisons les mêmes modèles que ceux du GIEC (Groupe d'expert intergouvernemental sur l'évolution du climat) pour notre Terre [...] » : Emeline Bolmont citée in « L'astrophysicienne qui joue avec les planètes » in Campus n° 137 juin 2019 p. 44-47, Université de Genève, Genève, Suisse.

³⁵En témoignent les signaux forts réitérés sur l'importance de problèmes contemporains dont fait partie le changement climatique, tels que celui « des 15.000 » publié par [Ripple *et al.*, 2017].

6.4 Constitution d'une liste d'expressions clefs du domaine

Une liste d'expressions clefs du domaine du changement climatique en langue anglaise a été créée manuellement. Cette liste a servi à nous familiariser avec le domaine composite du changement climatique et ses grandes sous-problématiques, en particulier les problématiques abordées au niveau des institutions scientifiques et politiques internationales.

Pour constituer la liste, nous avons d'abord observé et listé manuellement des termes communément employés dans divers types de supports informatifs spécialisés sur le thème du changement climatique : documents et sites scientifiques, institutionnels, et d'organisations non-gouvernementales. Ensuite, ont été retirées de la liste finale les expressions qui paraissaient n'appartenir exclusivement qu'à un seul des domaines de recherche fortement associés au changement climatique ; principalement les domaines de la climatologie, de la météorologie, de l'agriculture et de la migration humaine.

La liste finale se trouve en annexe B, p. 219. La démarche de constitution manuelle de la liste suit les étapes suivantes :

1. Recherche sur internet de sources scientifiques et encyclopédiques, institutionnelles et d'information, et d'organisations non-gouvernementales spécialisées.
2. Recueil des expressions clefs (vocabulaire listé, glossaires, expressions mises en avant) qui se répètent dans plusieurs sources, et compilation en une liste générale commune.
3. Filtrage manuel des expressions candidates dans la liste générale :
 - (a) suppression des expressions qui appartiennent au langage commun : chaque expression gardée dans la liste (parfois après consultation de sa définition) doit laisser supposer que le texte de l'article scientifique où elle se trouve traite du changement climatique ou est directement lié à ses problématiques.
 - (b) Elimination des expressions homonymes du vocabulaire de la langue courante, par exemple : 'air pollution', 'frozen ground', 'fingerprint', "sequestration" et 'irreversibility'.

- (c) Elimination des expressions qui ne font pas partie du vocabulaire partagé entre une grande diversité des domaines des documents consultés, telles que les expressions plutôt réservées à :
- la politique, par exemple : 'eco-refugee', 'community-based adaptation', 'Partnership for Market Readiness (PMR)', 'Monitoring, Evaluation and Learning (MEL)';
 - la technique, par exemple : 'bituminous coal', 'cogeneration', 'hazards-based approach' ;
 - les sciences en général, par exemple : 'carbon cycle', 'absorption of radiation', 'British thermal unit (Btu)';
 - les sciences de la terre, par exemple : 'glacier', 'thermocline', 'geoid', 'geosphere', 'tropospheric ozone precursor' ;
 - l'écologie, par exemple : 'diatoms', 'biomass', 'deforestation', 'ecosystem', 'terrestrial and marine biosphere'
 - la météorologie, par exemple : 'atmosphere', 'heat wave', 'Indian Ocean Dipole (IOD)'
 - et même la climatologie, par exemple : 'climate system', 'extreme climate event', 'energy budget of the Earth', 'climate-carbon cycle feedback'.
- (d) Elimination des expressions de moins de 4 lettres correspondant à des acronymes homonymes d'un mot.
- (e) Elimination semi-manuelle des doublons.
- (f) Vérification des définitions des concepts.
- (g) Ajout de pluriels ou de singuliers aux expressions qui s'y prêtent.

La motivation initiale pour la création de cette liste d'expressions était la recherche d'articles liés au changement climatique, pour en constituer le corpus. Cependant, nous avons constaté que la recherche de la seule expression clef la plus générique « *climate change* » sur la plateforme ISTEX permet d'obtenir un nombre suffisant d'articles pour le corpus. La méthodologie de la création de cette liste est générale et pourrait servir à la constitution de corpus d'articles d'autres domaines ; l'ajout de pluriels et de variantes orthographiques ('centre' et 'center', etc.) pouvant aussi se faire automatiquement.

6.5 Constitution du corpus d'articles du domaine

Parmi les caractéristiques des articles scientifiques qui sont avantageuses pour leur traitement automatique, il y a par exemple les suivantes :

- Ils ont normalement une bonne orthographe.
- Ils possèdent des métadonnées, telles que titre, auteurs, année, journal, résumé, mots clefs, corps de l'article, bibliographie et renvois bibliographiques.
- Ils sont de plus en plus disponibles ouvertement et gratuitement en ligne.
- Ils sont parfois fournis par leurs sources dans des formats spécialement adaptés aux traitements automatiques, tels que les formats XML avec JATS (Journal Article Tag Suite) ou TEI (Text Encoding Initiative).
- Ils suivent souvent des standards de publication, tels que la structure IMReD.

6.5.1 Source et critères pour les choix des articles

La structure IMReD a été retenue au départ comme une caractéristique de sélection des articles du corpus. Ce choix méthodologique exclue des articles non rédigés avec une variante de la structure IMReD (voir les structures au § 6.2), notamment parmi ceux de sciences humaines et sociales. Le but de ce choix a été de faciliter de possibles recherches ultérieures prenant en compte la présence d'expressions de l'incertitude en fonction de la structuration des articles, suivant ce que nous avons fait avec d'autres corpus dans [Atanassova *et al.*, 2018].

L'archive numérique ISTEX³⁶ a été retenue comme source des articles. Nous avons sélectionné les articles ayant une structure apparentée à IMReD et contenant l'expression clef « *climate change* ». Plus précisément, pour télécharger le corpus nous avons utilisé la formule de recherche suivante :

« "*climate change*" AND "*global warming*" AND Introduction AND keywords AND Methods AND Results AND Discussion AND language:eng AND genre:"research-article" », avec les options de téléchargement des types de fichiers « Texte intégral : TEI » et « Métadonnées : XML ».

³⁶Archive numérique d'articles scientifiques, d'accès libre pour les chercheurs des universités qui y participent. URL : <https://www.istex.fr/>

6.5.2 Composition du corpus d'articles

Nous avons obtenu le corpus qui se compose de 4916 articles scientifiques liés au changement climatique et issus de domaines scientifiques variés. Les champs de ces articles ne se restreignent absolument pas à la climatologie et vont par exemple de l'ornithologie aux énergies renouvelables : biologie, géophysique, agronomie, économie, politique, sociologie, droit, architecture, etc.

Nous avons téléchargé ces articles au format XML-TEI (pour des traitements automatiques) et PDF (pour des traitements manuels). Le corpus est constitué de deux parties :

- le premier dossier contient 4 759 articles dont la date de publication va jusqu'au 7 février 2019. C'est sur ces articles que l'extraction automatique de phrases a été effectuée.
- Le second dossier contient 157 articles dont la date de publication va du 1er janvier 2000 jusqu'au 28 janvier 2021. C'est sur ces articles que l'extraction manuelle de phrases a été effectuée, pour ne pas extraire manuellement les mêmes phrases que celles extraites automatiquement.

Tous les articles de ce corpus qui ont été consultés manuellement, lors de la recherche d'expressions de catégorie d'incertitude, ont permis de vérifier que leur contenu est effectivement en rapport avec le changement climatique, quel que soit leur domaine scientifique.

6.5.3 Exploration préliminaire

Pour commencer, un premier essai de recherche manuelle d'expressions de l'incertitude a été effectué dans deux articles scientifiques pris dans le corpus sur le changement climatique sans connaître ni leur titre ni leur contenu à l'avance : les expressions de l'incertitude qui y ont été repérées ont été systématiquement soulignées, puis examinées pour repérer des formes linguistiques de reconnaissance d'expressions de l'incertitude. Par exemple, « considerations are likely to be » et « trends may be expected to influence » peuvent correspondre à une règle ayant une structure telle que :

« sujet+(may)+verbe be+[likely|expected]+to+verbe »

Cette connaissance d'expressions modèles a permis de composer des expressions régulières, de longueur limitée et incluant des alternatives, utilisées ensuite

pour extraire automatiquement du corpus d'articles les premières 500 phrases contenant des expressions-candidates à des catégories d'incertitude. Ce premier repérage manuel a servi aussi à se familiariser avec les aspects pratiques du processus, ce qui a aidé à constituer un mode opératoire d'annotation pour les experts annotateurs.

6.6 Constitution du corpus d'étude et du corpus d'évaluation

Le corpus d'étude, qui a servi à l'élaboration du système, est composé de 700 phrases divisées en deux sous-corpus : 500 phrases extraites automatiquement (voir corpus en annexe G.1, p. 247) ; et 200 phrases extraites manuellement (voir corpus en annexe G.2, p. 267).

Le corpus d'étude est destiné à être annoté manuellement, afin que les expressions-candidates de chaque phrase puissent servir de modèles pour la rédaction des règles linguistiques du système. En particulier, ce corpus est utilisé pour :

- la rédaction des règles linguistiques selon la grammaire formelle (voir chap. 7, p. 145).
- la vérification et l'optimisation du fonctionnement des règles lors de l'élaboration du système (voir § 7.6, p. 153).
- la constitution d'un *gold standard* dont il fait partie avec ses annotations (voir § 6.7.2, p. 138).

Le corpus d'évaluation, quant à lui, contient 300 phrases extraites manuellement (voir corpus en annexe G.3, p. 275). Il est utilisé pour l'évaluation du système, en comparant les annotations manuelles des catégories d'incertitude à celles faites automatiquement par le système.

6.6.1 Extraction automatique de 500 phrases

Les premières 500 phrases (voir corpus en annexe G.1, p. 247) ont été extraites par la recherche d'expressions-candidates d'incertitude scientifique au sein du corpus de 4 759 articles publiés avant le 7 février 2019. Ceci a été fait en employant des expressions régulières relativement courtes.

Du point de vue de la codification des caractères, ces 500 phrases sont extraites telles quelles, en gardant leurs défauts de transcription automatique, provenant de l'article au format TEI proposé par ISTEEX, par exemple :

- « 'efficiency paradox' » \Rightarrow « â€~eï¬fciciency paradoxâ€™™ »
- « detect 3-6-month » \Rightarrow « detect 3â€"6-month »
- « near 50°S, 120°W » \Rightarrow « near 50Â°S, 120Â°W »

6.6.2 Extraction manuelle de 500 phrases

Les phrases ont été extraites manuellement à partir de 25 articles publiés entre le 1er janvier 2000 et le 28 janvier 2021. La démarche consiste à analyser manuellement le texte des articles pour y trouver les phrases qui expriment des incertitudes. Parmi ces 500 phrases, 200 (voir corpus en annexe G.2, p. 267) viennent enrichir le corpus d'étude et 300 autres (voir corpus annexe G.3, p. 275) serviront à l'évaluation du système.

L'analyse manuelle a été faite par un expert qui a trouvé un total de 525 phrases contenant des expressions-candidates de l'incertitude au sein de 25 articles. Une vérification et élimination des phrases doublons à ensuite laissé un total de 500 phrases disponibles.

Du point de vue de la codification des caractères, ces 500 phrases sont ajoutées manuellement des articles au format original PDF vers le corpus de phrases en évitant tout défaut de transcription (à la différence des phrases extraites automatiquement, voir § 6.6.1).

6.7 Proposition d'un corpus annoté - *gold standard*

L'un des objectifs de notre travail est l'annotation manuelle et automatique des différents corpus afin de constituer un Gold standard : corpus annoté selon les catégories de notre ontologie et dont la qualité des annotations est contrôlée. De tels corpus annotés sont en général coûteux à produire, et peuvent servir, entre autres, à la mise en place de méthodes par apprentissage automatique.

Ce Gold standard que nous proposons est composé de 1000 phrases annotées (voir corpus en annexe G, p. 247), avec des sous-corpus présentés sur le schéma 6.2. Ci-dessous nous présentons les différentes étapes du traitement et de la constitution du Gold Standard. Il peut être divisé en plusieurs parties :

6.7. PROPOSITION D'UN CORPUS ANNOTÉ - GOLD STANDARD

- du point de vue de l'extraction des phrases (automatique vs manuelle) à partir du corpus d'articles : parmi les 700 phrases, 500 sont extraites automatiquement, et 200 sont extraites manuellement. De leur côté, les 300 phrases destinées à l'évaluation du système sont extraites manuellement.
- du point de vue de l'annotation (automatique vs manuelle) : toutes les 1000 phrases sont annotées manuellement. Les 500 phrases sont annotées par trois experts. Les ensembles de 200 et de 300 phrases sont annotés ensuite par l'un des experts.
- du point de vue de l'utilisation des sous-corpus (élaboration du système vs évaluation) : 700 phrases sont utilisées durant l'élaboration du système comme modèles pour rédiger les règles. Les 300 autres phrases sont utilisées pour évaluer le système.

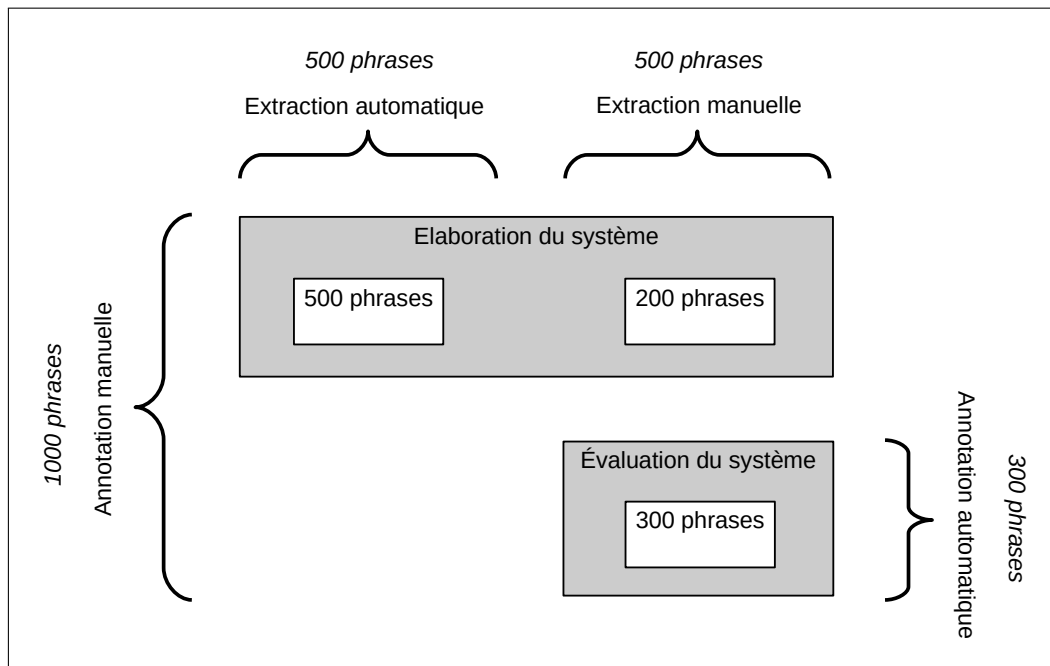


Schéma 6.2: Composition du Gold Standard.

6.7.1 Annotation manuelle

Ces annotations manuelles ont été faites indépendamment par trois experts de la langue anglaise. Ces experts ont participé à l'annotation comme suit :

- deux experts ont annoté la première partie de 500 phrases du corpus d'étude.
- le troisième expert a annoté la totalité des 1000 phrases des corpus d'étude et corpus d'évaluation. Il a révisé les annotations des deux autres experts et, dans les cas de conflit d'annotation entre les experts, il a pris les décisions finales d'annotation.

Pré-requis. Les compétences d'experts liées à l'annotation manuelle des phrases candidates, sont en particulier les suivantes : maîtriser la langue anglaise pour savoir évaluer et clarifier le sens de ses expressions complexes, être familiarisé avec les concepts de l'ontologie pour classer et nommer les expressions de l'incertitude, avoir une compréhension commune de la méthode d'identification et classification des expressions potentielles de l'incertitude.

Ces experts de la langue anglaise n'étaient pas au départ habitués à l'analyse des formes textuelles de l'incertitude et des catégories de l'ontologie tels que les raisonnements, et cela aura pu avoir une incidence significative sur la proportion de notations non consensuelles. Plusieurs moyens ont été employés afin de limiter ce problème en renforçant leur compréhension individuelle et commune : des explications orales d'introduction de l'expérimentation, un débat après une première analyse d'articles (voir ci-dessous), et un mode d'emploi (voir mode d'emploi en annexe C, p. 225) accompagné d'explications orales concernant les catégories de l'ontologie (notamment sur les raisonnements déductifs et inductifs et abductifs) et les annotations.

Avant l'annotation, les experts se sont familiarisés avec les expressions d'incertitudes et leurs annotations, notamment par un premier essai d'analyse : les deux experts ont analysé et annoté chacun les expressions-candidates de 6 articles, et le troisième expert a analysé les mêmes 12 articles afin de comparer l'ensemble des annotations des trois. Ces annotations de 12 articles ont été suivies d'un débat entre les trois experts pour favoriser par la suite une harmonisation de leurs analyses et annotations sur les 500 phrases du corpus.

Les types d'incertitude des expressions-candidates pour chaque phrase du corpus ont été annotés suivant l'ontologie. Dix abréviations ont été utilisées pour annoter ces différents catégories³⁷ (voir schéma 6.3, p. 135) :

- huit annotations de catégories d'incertitudes :

³⁷Pour mémoire, pour une éventuelle consultation de fichiers de l'expérience, un changement des abréviations de catégories concernant des incertitudes a été effectué : *i-m* pour *im*, *i-q* pour *iq*, *i-m-int* pour *mdi*, *i-m-val* pour *mvp*, *i-q-rin* pour *rin*, *i-q-rab* pour *rab*, *i-q-hnv* pour *hnv*.

i : « incertitude ».

Ex.: However, due to the uncertainties inherent in model parameters and driving variables [...], these should not be taken for solid year predictions.

i-m : « incertitude mesurée ».

Ex.: [...] in regions where [...], the potential for increased carbon sequestration in agricultural soils is much greater, perhaps by an order of magnitude.

i-q : « incertitude qualitative ».

Ex.: The implications of conservation tillage on methane and nitrous oxide emissions requires additional research.

i-m-int : « incertitude : mesure d'intervalle » (ou approximation pouvant renvoyer ou non à une échelle d'intervalles).

Ex.: [...] elevations are between 500 and 1200 m.

i-m-val : « incertitude mesurée : mesure de valeur ponctuelle ».

Ex.: This endangered population of 250 individuals has a 50% chance of extinction over the next 100 years.

i-q-rin : « incertitude qualitative : raisonnement inductif ».

Ex.: Under such conditions, [...] will likely shift westward to maintain an adequate water balance.

i-q-rab : « incertitude qualitative : raisonnement abductif ».

Ex.: But the measured concentrations were less than the standards, presumably because of [...].

i-q-hnv : « incertitude qualitative : hypothèse non vérifiée ».

Ex.: Although this has to be confirmed in future work, for now we assume that [...].

- deux annotations autres :

ai : « aucune incertitude » : par exemple pour la simple présentation d'un fait, pour un raisonnement déductif, pour une définition, etc.

Ex.: A mitigation project may be defined as a [...].

nd : « non décidé » (classification temporaire).

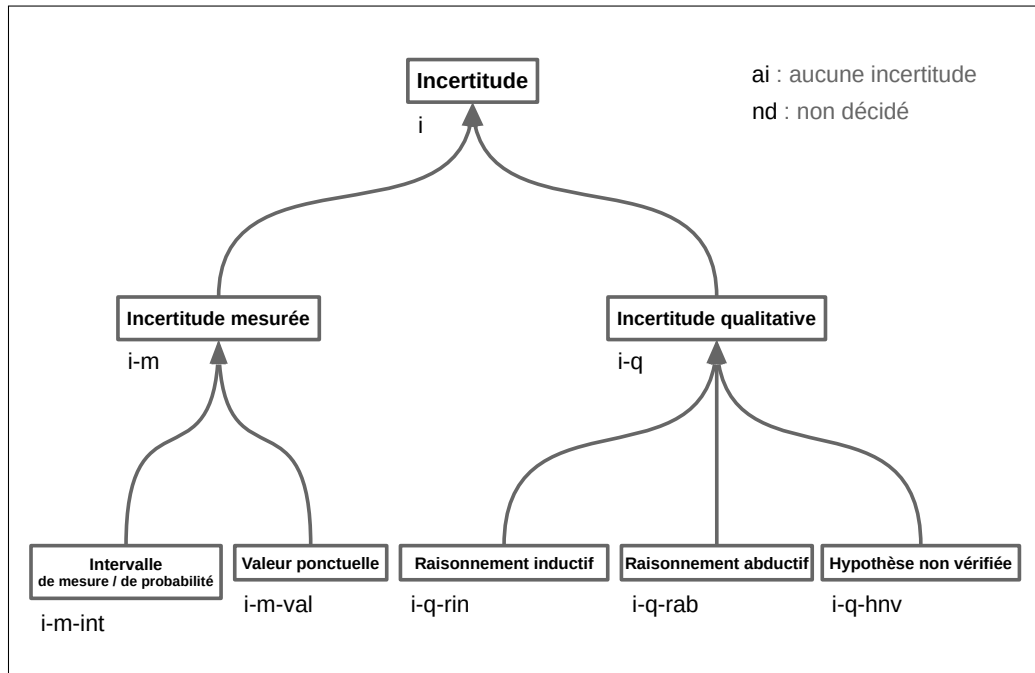


Schéma 6.3: Ontologie de l'incertitude scientifique, avec les abréviations utilisées pour l'annotation.

Suivant l'expression de *condensation de l'incertitude* employée par [Shackley et Wynne, 1996, p. 285-287] pour qualifier d'une manière ascendante la relation habituellement nommée *héritage* entre des catégories, chaque catégorie d'incertitude de l'ontologie se retrouve condensée dans la catégorie d'incertitude immédiatement plus générale, jusqu'à la catégorie d'incertitude la plus générale. L'organisation des catégories est prise en compte dans le fonctionnement et l'évaluation du système, car l'annotation d'une catégorie entraîne aussi l'annotation des catégories où elle se condense. C'est à dire que :

- les catégories *i-m-int* et *i-m-val* sont présentes dans la catégorie *i-m* ;
- les catégories *i-q-rin*, *i-q-rab* et *i-q-hnv* sont présentes dans la catégorie *i-q* ;
- les catégories *i-m* et *i-q* sont présentes dans la catégorie *i*.

Par conséquent, lorsqu'une expression d'incertitude correspond par exemple à la catégorie *i-q-rab*, alors elle correspond aussi à la catégorie *i-q* qui est d'une granularité moins fine pour cette expression, ainsi qu'à la catégorie *i* qui est la plus générale. Les expressions d'incertitude identifiées ont donc été catégorisées

chacune selon la granularité la plus fine possible ; et l'annotation de cette catégorie a été jointe à sa phrase³⁸. Par exemple pour des expressions de raisonnements :

- Dédution : partant de causes certaines (prémisses³⁹ considérées vraies), leur enchaînement certain fait constater leurs effets.

L'énoncé est certain. → catégorie **ai**

Exemple :

1. Si la toundra se réchauffe, le pergélisol dégèle.
2. Si le pergélisol⁴⁰ dégèle, du méthane est émis dans l'atmosphère.
3. La toundra se réchauffe, donc du méthane est émis dans l'atmosphère.

- Induction : partant d'observations certaines, leur généralisation mène à des effets probables.

L'énoncé est probable. → catégorie **i-q-rin**

Exemple :

- La force des ouragans a augmenté chaque année les 10 dernières années, donc il est prévisible que la force des ouragans augmentera cette prochaine année.

- Abduction : partant d'un fait certain, ses causes les plus plausibles sont inférées.

L'énoncé est plausible. → catégorie **i-q-rab**

Exemple :

- L'épaisseur des cernes d'un arbre est régulière, donc il est plausible, là où cet arbre est planté et durant les cycles annuels de sa croissance, que la pluviométrie ait été régulière et que la température ambiante soit restée positive.

³⁸Lors de l'évaluation, avant de calculer les indices, le système ajoute aux annotations manuelles les annotations des catégories où ces annotations manuelles se condensent.

³⁹Le terme *Prémisse* « désigne chacune des deux propositions d'un syllogisme d'où résulte la conclusion ; par extension, ce nom désigne aussi tout argument d'où découle une conséquence ». Académie française. Consultée le 10/04/2019. URL : <http://academie-francaise.fr/premisse-premices>

⁴⁰Equivalent anglais plus connu : *permafrost*.

Les experts ont eu pour tâche première de reconnaître et indiquer pour chaque phrase la catégorie de ses expressions d'incertitude ; une même phrase pouvant correspondre à plusieurs catégories selon ses contenus. Le contexte de chaque phrase à analyser, phrase à droite et la phrase à gauche, était disponible.

Chronologie Les annotations manuelles suivent les étapes suivantes :

1. Chacune des 500 phrases a été soumise aux trois experts :
 - Annotation manuelle par chaque expert individuellement avec les catégories de l'incertitude définie.
 - Les phrases dont les 3 annotations sont identiques gardent cette annotation.
 - Les phrases dont 2 annotations sont identiques gardent cette annotation.
 - Les phrases dont les annotations sont différentes sont réexaminées en fonction des annotations et l'un des experts fait une 4e annotation.
2. Plus de 2 semaines plus tard, pour les phrases dont les 4 annotations étaient différentes :
 - Si 2 des 4 annotations sont identiques, la phrase garde cette annotation.
 - Si les annotations sont encore toutes différentes, la phrase est réexaminée en fonction de l'expérience sur les autres phrases, et une décision d'annotation est prise par l'un des experts (voir schéma 6.4).
3. Lors de l'élaboration des règles d'annotation, après chaque fin des cycles d'annotation automatique, réitérés pour perfectionner le système :
 - Les phrases dont l'annotation manuelle et l'annotation automatique (par les règles en préparation) sont différentes sont réexaminées :
 - soit c'est la règle qui est changée,
 - soit c'est l'annotation qui est changée.

Cette étape s'est achevée lorsque le corpus des phrases a été jugé conforme aux attentes en termes de qualité d'annotation manuelle, notamment par la concordance avec les annotations automatiques. Cela a correspondu à une révision de plusieurs règles et de plusieurs annotations manuelles.

6.7. PROPOSITION D'UN CORPUS ANNOTÉ - GOLD STANDARD

4. Les 500 phrases restantes sont soumises à l'annotation d'un unique expert.
La procédure de l'étape 3 est aussi suivie pour ces phrases.

C	D	E	F	G	H	I	J
	Ex1	EX2	EX3	Cat. Accord	n° de choix identique	Choix le plus normal	Choix quand désaccord
is too short, popu	hny	ai	hny	hny	2	hny	
e, many fragment	hny	ai	hny	hny	2	hny	
al we assume see	rab hny	i	rin hny	hny	2	hny	
which mimic the	rin	ai	hny		0		nd
idered for V. fimb	rin	ai	hny		0		hny
conducting lAs an	ai	ai	hny	ai	2	ai	

Schéma 6.4: Comparaison des annotations des experts : aperçu du document CSV.

6.7.2 Corpus annoté final - *gold standard*

Le corpus de 1000 phrases (voir corpus en annexe G, p. 247) a été annoté manuellement à partir des différentes catégories d'incertitude de ses expressions, ce qui traduit une expertise humaine : l'annotation correcte des phrases due à l'expertise humaine permet d'utiliser ce corpus comme un standard pour d'autres expérimentations. Ce corpus est décrit comme suit :

- **contenu du corpus** : 1000 phrases en langue anglaise contenant des expressions-candidates des catégories d'incertitude scientifique.
- **origine des phrases du corpus** : articles liés au thème du changement climatique, issus de domaines scientifiques les plus divers.
- **catégorie d'incertitude annotés** : *i-q-hny*, *i-q-rab*, *i-q-rin*, *i-m-val*, *i-m-int*, *i-q*, *i-m*, *i*, et *ai*.
- **format du corpus** : document CSV en UTF-8. Options d'ouverture avec un logiciel tableur : séparation par tabulation et virgule, et délimitation de chaîne de caractères par guillemets (").
- **composition et parties du document CSV** : 1000 lignes x 4 colonnes (voir schéma 6.5, p. 139). Chaque ligne contient :
 - **1^{re} colonne** : une phrase du corpus.
 - **2^e colonne** : annotation abrégée des catégories d'incertitude des expressions contenues dans la phrase.

- **3^e colonne** : identifiant alphanumérique ISTE^X de l'article d'où la phrase est extraite, pour la traçabilité (voir § 6.7.3, p. 139).
- **4^e colonne** : la mention de l'extraction soit automatique soit manuelle de la phrase.

637	There is little quantitative information about such se	'i-m'	1F99E444BCE29F2CA230	Extraction automatique
638	While this assumption may be justified in some eco	'i-q-hnv'	1F99E444BCE29F2CA230	Extraction automatique
639	Such maps can demonstrate whether certain parts	'i-q-hnv', 'i-m-int'	1F99E444BCE29F2CA230	Extraction automatique
640	For example, certain insect outbreaks are more like	'i-m-int'	1F99E444BCE29F2CA230	Extraction automatique
641	Parts of a landscape that are highly fragmented by	'i-q-hnv', 'i-m-int'	1F99E444BCE29F2CA230	Extraction automatique
642	Thirty per cent area of the region has an MAT less	'i-m-int'	002F72D60D1D08325942	Extraction automatique
643	Mean annual precipitation (MAP) is 100–300 mm in	'i-m-int'	002F72D60D1D08325942	Extraction automatique
644	About 60% of the area has MAP between 100 mm	'i-m-int'	002F72D60D1D08325942	Extraction automatique
645	Observations show that, over the last 100 years, at	'i-m-int'	02D3A03ABCD3188EC9B	Extraction automatique
646	However, our current understanding of the functioni	'i'	02D3A03ABCD3188EC9B	Extraction automatique

Schéma 6.5: Gold standard : aperçu du document CSV. Contenu par ligne : phrase, annotations, identifiant, type d'extraction.

A titre indicatif, le champ scientifique principal de l'article d'où est extraite chacune des 1000 phrases du corpus a été déterminé manuellement⁴¹ : pour chaque phrase, ce champ scientifique est annoté dans une 5^e colonne. Cependant, le dénombrement qui en résulte (voir la répartition du nombre des phrases dans le schéma 6.6, p. 141 ; voir la répartition proportionnelle dans le schéma 6.7, p. 142 ; et voir le tableau de dénombrement 6.1, p. 143) ne représente que partiellement la diversité des champs scientifiques contextes de ces phrases. Ceci est dû au fait que la quasi-totalité des articles d'où ces phrases sont extraites sont en fait transdisciplinaires : c'est uniquement le champ scientifique le plus représentatif de l'article d'origine qui est attribué à chaque phrase.

6.7.3 Traçabilité de l'article d'origine de chaque phrase

Chaque phrase du corpus est extraite d'un article présent sur l'archive numérique Initiative d'excellence de l'Information Scientifique et Technique (ISTEX). Pour pouvoir retrouver l'origine des phrases du corpus, chaque phrase est sauvegardée avec l'identifiant *idIstex*⁴² de l'article d'où elle est extraite. Par exemple, l'article⁴³ où se trouve la phrase « *It is my belief that 50 years from now,*

⁴¹ Les catégories des champs scientifiques sont le moins regroupées possible.

⁴² Identifiant interne à 40 caractères fourni par ISTE^X, unique pour chaque article sur ISTE^X.

⁴³ Truman P. Young (2000), « Restoration ecology and conservation biology » in *Biological Conservation*, Janvier 2000, vol. 92, n° 1, p. 73-83. DOI : 10.1016/S0006-3207(99)00057-9.

6.7. PROPOSITION D'UN CORPUS ANNOTÉ - GOLD STANDARD

the majority of the world's habitats and species will either be destroyed or on their way to recovery from a degraded state. » correspond à l'identifiant idIstex **0FBF56FD252AD5FA6254ADA6499E2AE1EDB12E80** :

- l'article d'origine : il peut être consulté, et téléchargé au format PDF, depuis la page <https://dl.istex.fr>, en indiquant son identifiant idIstex dans le champ de saisie *Identifiants ARK*, ou alternativement en intégrant son identifiant idIstex dans une URL sous la forme « <https://api.istex.fr/document/{idIstex}/fulltext/pdf> », c'est-à-dire pour l'exemple : « <https://api.istex.fr/document/0FBF56FD252AD5FA6254ADA6499E2AE1EDB12E80/fulltext/pdf> ».
- les métadonnées de l'article d'origine : elles peuvent être consultées, et téléchargées au format JSON, en intégrant son identifiant idIstex dans une URL sous la forme « <https://api.istex.fr/document/{idIstex}> », c'est-à-dire pour l'exemple : « <https://api.istex.fr/document/0FBF56FD252AD5FA6254ADA6499E2AE1EDB12E80> ».

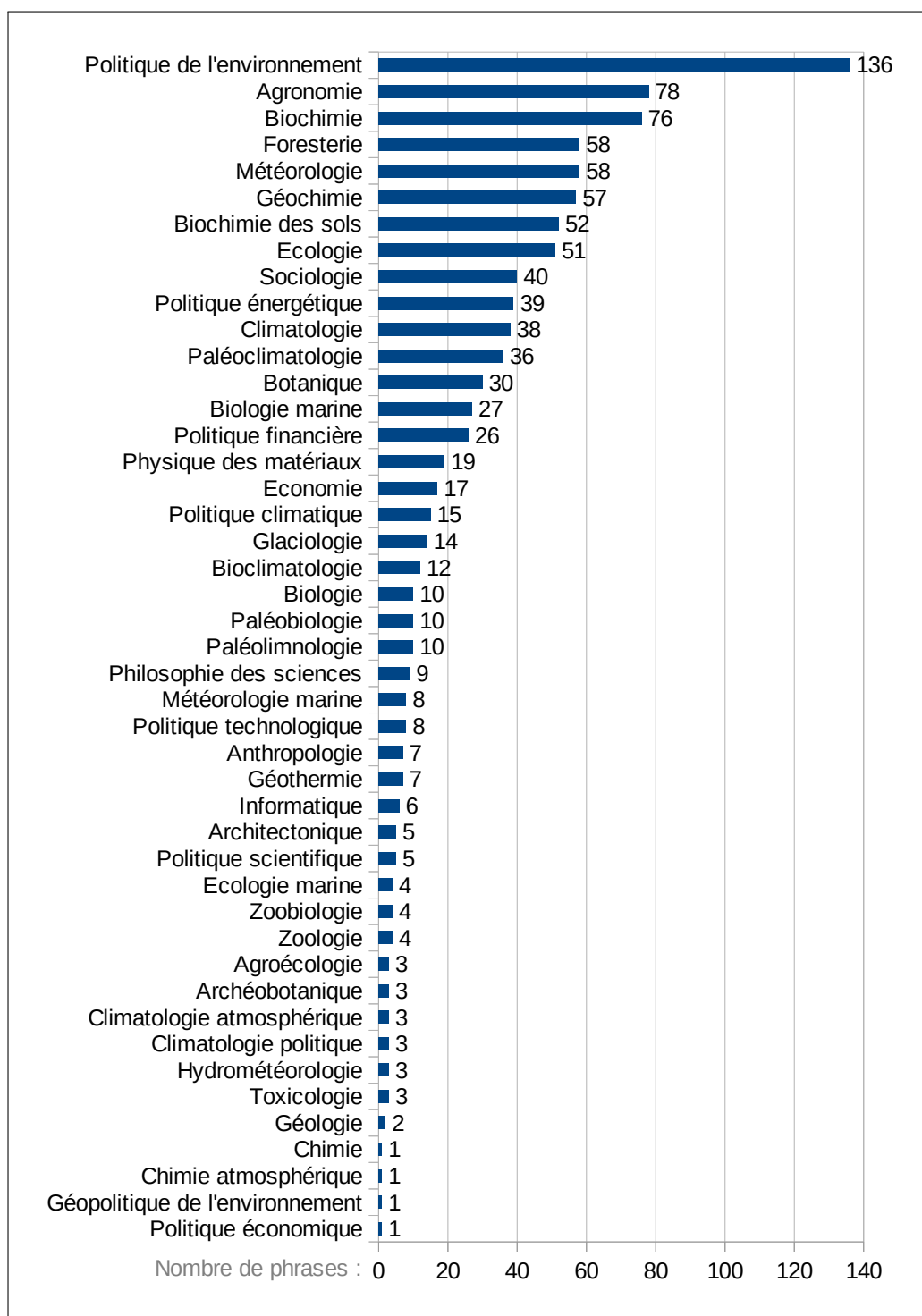


Schéma 6.6: Dénombrement des 1000 phrases du gold standard en fonction du champ scientifique principal de leur article, ordonné par effectifs décroissants des champs. (Les effectifs correspondent donc à un ‰ des phrases de chaque champ).

6.7. PROPOSITION D'UN CORPUS ANNOTÉ - GOLD STANDARD

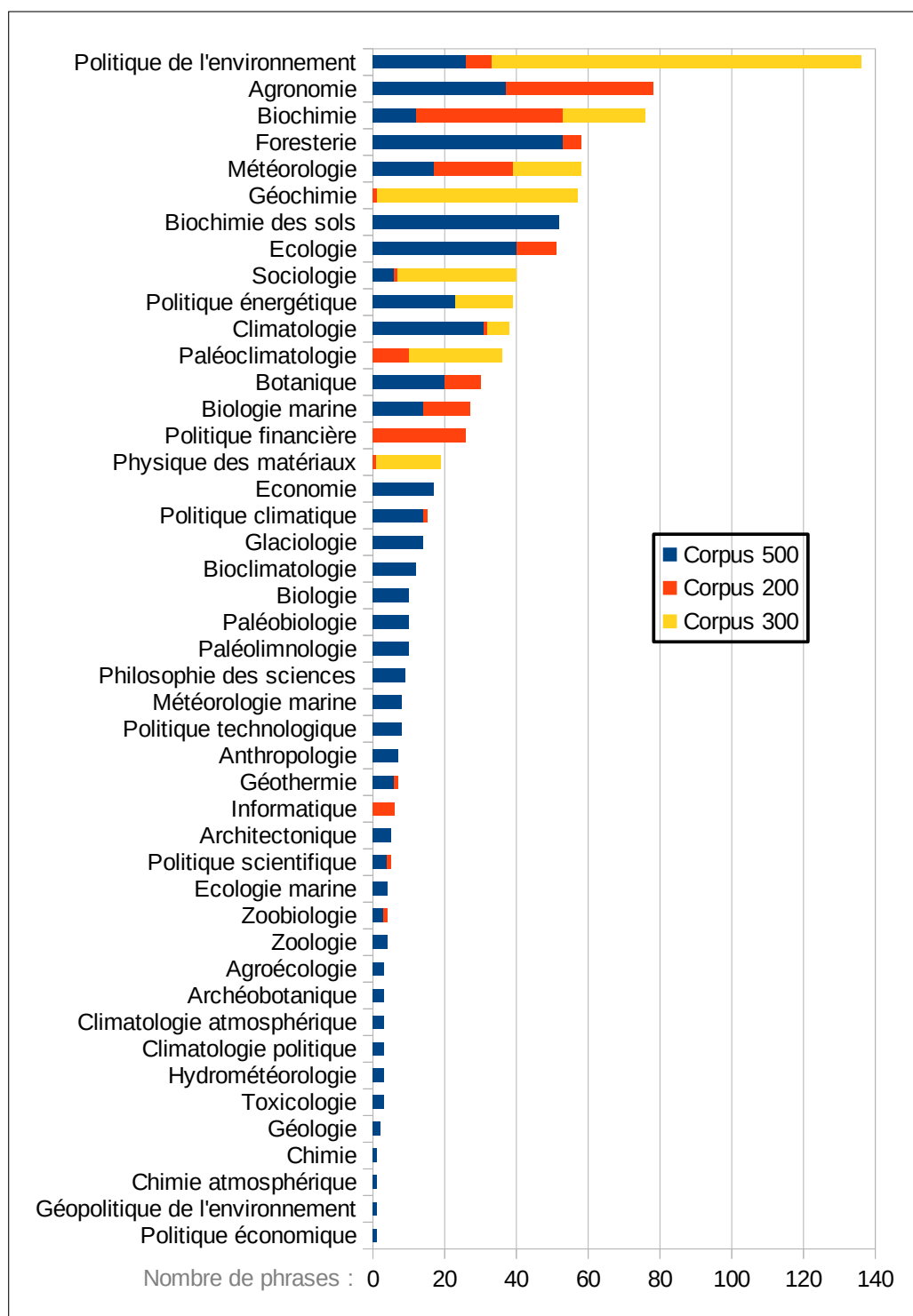


Schéma 6.7: Répartition proportionnelle des 1000 phrases du gold standard en fonction du champ scientifique principal de leur article et de leur corpus d'origine (de 500 ou 200 ou 300 phrases), ordonné par effectifs décroissants des champs.

CHAPITRE 6. CONSTITUTION DE CORPUS POUR LE TRAITEMENT DE L'INCERTITUDE

Champ principal d'article	Corpus 500	Corpus 200	Corpus 300	Total des phrases par champ
Politique de l'environnement	26	7	103	136
Agronomie	37	41	0	78
Biochimie	12	41	23	76
Météorologie	17	22	19	58
Foresterie	53	5	0	58
Géochimie	0	1	56	57
Biochimie des sols	52	0	0	52
Ecologie	40	11	0	51
Sociologie	6	1	33	40
Politique énergétique	23	0	16	39
Climatologie	31	1	6	38
Paléoclimatologie	0	10	26	36
Botanique	20	10	0	30
Biologie marine	14	13	0	27
Politique financière	0	26	0	26
Physique des matériaux	0	1	18	19
Economie	17	0	0	17
Politique climatique	14	1	0	15
Glaciologie	14	0	0	14
Bioclimatologie	12	0	0	12
Paléolimnologie	10	0	0	10
Paléobiologie	10	0	0	10
Biologie	10	0	0	10
Philosophie des sciences	9	0	0	9
Politique technologique	8	0	0	8
Météorologie marine	8	0	0	8
Géothermie	6	1	0	7
Anthropologie	7	0	0	7
Informatique	0	6	0	6
Politique scientifique	4	1	0	5
Architectonique	5	0	0	5
Zoologie	4	0	0	4
Zoobiologie	3	1	0	4
Ecologie marine	4	0	0	4
Toxicologie	3	0	0	3
Hydrométéorologie	3	0	0	3
Climatologie politique	3	0	0	3
Climatologie atmosphérique	3	0	0	3
Archéobotanique	3	0	0	3
Agroécologie	3	0	0	3
Géologie	2	0	0	2
Politique économique	1	0	0	1
Géopolitique de l'environnement	1	0	0	1
Chimie atmosphérique	1	0	0	1
Chimie	1	0	0	1
Total des phrases par corpus	500	200	300	1000

Tableau 6.1: Dénombrement des 1000 phrases du gold standard en fonction du champ scientifique principal de leur article et de leur corpus d'origine (de 500 ou 200 ou 300 phrases), ordonné par effectifs décroissants dans la colonne *Total des phrases par champ*.

Chapitre 7

Règles linguistiques pour l’annotation de l’incertitude

Les expressions porteuses de l’incertitude scientifique en langue naturelle peuvent être constituées de plusieurs mots (unités lexicales, ponctuation, apostrophes et symboles textuels) et peuvent être discontinues. Nous avons choisi de proposer des règles *ad-hoc*, qui s’appuient sur des conditions de présence d’indices positifs et négatifs dans les phrases. Cette approche est fortement inspirée des travaux autour de l’Exploration Contextuelle (EC) par J.-P. Desclés [[Desclés et al., 1997](#), [Desclés, 2006](#)], qui est une méthode opérationnelle permettant la formalisation de règles linguistiques d’annotation.

7.1 Motivations pour la création d’une grammaire formelle

Remarquant là l’opportunité de créer un outil destiné à des linguistes, et dont l’utilisation ne repose pas sur des compétences en informatique, nous avons proposé une grammaire formelle dédiée à la formulation des expressions linguistiques recherchées. Si d’autres alternatives existent, notamment les expressions régulières, elles sont peu intuitives et limitées à la représentation d’expressions contigues. Par ailleurs, elles sont spécifiques aux usages informatiques et caractérisées par une écriture et une lecture complexes éloignées de la linguistique.

De même, l’écriture de règles en Corpus Query Language (CQL), décrite par exemple par [[Klyueva et al., 2018](#)], est éloignée du langage naturel avec une complexité qui empêche l’écriture et lecture humaine de règles linguistiques longues.

Les problèmes d'apprentissage que posent les langages d'interrogation de corpus aux linguistes non informaticiens ne sont pas nouveaux, [Bański *et al.*, 2016, p. 2804].

Parmi d'autres exemples de grammaires formelles, nous avons par exemple observé les détails de systèmes de notation employés pour décrire, dans leurs normes respectives nationales ou internationales :

- les identifiants de comptes bancaires : International Bank Account Number (IBAN) défini par le standard international ISO 13616-1:2007¹, et Business Identifier Code (BIC) géré par l'organisation internationale Society for Worldwide Interbank Financial Telecommunication (SWIFT) et définit par le standard international ISO 9362:2014².
- les identifiants d'assurés sociaux : Numéro d'inscription au Répertoire³ (NIR, dit « numéro de sécurité sociale ») en France, Numéro d'Assurance-Vieillesse et Survivants⁴ (NAVS13) en Suisse, National Insurance Number (NINO) au Royaume-Uni, et Número de la Seguridad Social (NUSS, dit « Número de Afiliación » (NAF)) en Espagne.
- les adresses courrielles : selon les standard de facto des Request for Comments⁵ (RFC) de l'Internet Engineering Task Force (IETF), notamment les RFC 5321 (p. 40, § 4.1.2), RFC 5322 et RFC 3696.
- les numéros de téléphone : selon les décisions⁶ de l'Autorité de régulation des communications électroniques et des postes (ARCEP) en France, et selon les recommandations⁷ de l'Union Internationale des Télécommunications (UIT) à l'international.
- les adresses postales : selon les normes des pays se conformant à une Standardized Address Format Description (SAFD) de Universal Postal

¹ISO 13616-1:2007 disponible sur <https://www.iso.org/standard/41031.html>.

²ISO 9362:2014 disponible sur <https://www.iso.org/standard/60390.html>.

³Répertoire national d'identification des personnes physiques (RNIPP) français.

⁴NAVS13 introduit par l'article 133 du "RS 831.101 Règlement sur l'assurance-vieillesse et survivants (RAVS)" suisse.

⁵RFC disponibles sur <https://www.rfc-editor.org>

⁶ARCEP, Paris, France : décision n°2018-0881 modifiée en date du 24 juillet 2018 établissant le plan national de numérotation et ses règles de gestion. Disponible sur <https://extranet.arcep.fr>.

⁷UIT, Genève, Suisse : recommandations E.123 (2002) et son annexe 1 (2008), et complément à la E.164 (11/2010). Disponibles sur <https://www.itu.int>.

Union⁸ (UPU) à travers sa norme S42 liée à la norme ISO 19160-4:2017, dont par exemple l'implémentation par La Poste en France se fait à travers la norme AFNOR NF Z10-011.

La grammaire que nous proposons a pour objectif de rendre possible la description manuelle de motifs (dits aussi formes ou *patterns*) correspondant à des expressions linguistiques continues ou discontinues en langues naturelles. En particulier, elle doit pouvoir décrire des expressions des catégories de l'incertitude dans des textes scientifiques en langue anglaise, pour aboutir à leur traitement automatique.

Les limitations notables que nous avons suivies lors de l'élaboration de cette grammaire ont été :

- employer des symboles qui facilitent la lecture et l'écriture « naturelles » des expressions.
- éviter les symboles inutiles. Par exemple, l'utilisation d'un caractère '+' pour agréger des sous-expressions a été envisagée au départ ; cela s'est avéré inutile, et n'a donc pas été retenu.
- permettre une traduction automatique vers un langage informatique, notamment pour être implémentée dans le moteur de repérage.

7.2 Définition de la grammaire

La mise au point de la grammaire s'est faite par des cycles de réductions successives du nombre de symboles faisant partie de la grammaire pour permettre :

- des délimitations de sous-expressions,
- des propositions d'alternatives,
- des emboitements,
- l'inclusion de sous-expressions possibles mais non obligatoires.

⁸UPU, Berne, Suisse. <https://www.upu.int>.

7.2.1 Caractères réservés

La liste des caractères réservés que nous proposons est la suivante :

/ Délimiteur entre deux alternatives.

- Ex. : '[could/may/might]' ⇒ « could » ou « may » ou « might »

[et] Délimiteurs de début et de fin d'une série d'alternatives.

- Ex. : '[C/c/W/w]ould' ⇒ « Could » ou « could » ou « Would » ou « would »
- Ex. : 'it [[c/w]ould/may]' ⇒ « it could » ou « it would » ou « it may »

(et) Délimiteurs de début et de fin d'une expression non obligatoire.

- Ex. : 'may be(more)' ⇒ « may be » ou « may be more »
- Ex. : 'may(...)' ⇒ « may » ou « may have » ou « may start a new growth », etc.
- Ex. : 'may(_) just' ou 'may (_)just' ⇒ « may just » ou « may have just » ou « may be just », etc.

... Chaîne de un ou plusieurs caractères quelconques, espaces et ponctuations incluses.

- Ex. : '... are likely to' ⇒ « prices are likely to » ou « over a long period are likely to », etc.

_ (Caractère dit *tiret bas* ou *tiret du 8*.) Chaîne de un ou plusieurs caractères alphabétiques, tiret '-' (trait d'union) inclus.

- Ex. : '_ed' ⇒ « increased » ou « observed » ou « considered » ou « challenged », etc.
- Ex. : 'may be _er than' ⇒ « may be deeper than » ou « may be higher than », etc.
- Ex. : 'could be _[ing/ed]' ⇒ « could be approaching » ou « could be proposed », etc.

Chaîne de un ou plusieurs caractères numériques, virgules ',' et points '.' inclus, pouvant être précédée d'un plus '+' ou d'un moins '-'. Représente un nombre positif ou négatif, en notation anglaise ou française⁹.

- Ex. : 'approximately # °C' ⇒ « approximately 1,200 °C » ou « approximately -25.30 °C », etc.
- Ex. : 'about # km' ⇒ « about 2,000,000 km » ou « about 3.5 km », etc.
- Ex. : 'around #()%' ⇒ « around 5% » ou « around -47.35 % », etc.

^ (Caractère de l'accent circonflexe, seul.) Début ou fin d'une suite de lettres.

- Ex. : 'may be _ed^' ⇒ « may be induced by oxcarbazepine », mais pas *« may be medically better ».
- Ex. : '^all in this' ⇒ « are all in this category », mais pas *« could fall in this area ».

Tous les autres caractères sont des caractères non réservés. Les caractères utilisés librement dans les règles sont :

- les lettres de l'alphabet majuscules et minuscules de la langue anglaise (ce qui n'exclue pas à priori ceux d'autres langues comme la française avec ses signes diacritiques) ;
- les signes de ponctuation intermédiaire ',' (virgule) et ';' (point-virgule) et ':' (deux points) , et le '?' (point d'interrogation).
- les apostrophes (leurs écritures alternatives sont prises en compte) ;
- les symboles courants de l'écriture scientifique '%' et '°' (degré, comme dans °C).

⁹Dû au fait que les points et virgules des chiffres sont utilisés inversement en notation française et en notation anglaise, et que les articles qui paraissent en anglais peuvent incorporer dans le texte des données numériques d'origine non anglophone non réécrites. Par exemple : « 12,4 » en français est « 12.4 » en anglais, et « 53.485.234,30 » en français est « 53,485,234.30 » en anglais.

7.2.2 Observations sur les caractères

Caractères spéciaux des expressions régulières : le `'\'` (barre oblique inversée dite *backslash*), le `'*'` (astérisque), le `'.'` (point) et le `'!'`, ne font pas partie de notre grammaire, bien qu'il s'agit de caractères réservés en expressions régulières. Notre moteur de repérage traduit automatiquement les expressions linguistiques définies selon la grammaire formelle en expressions régulières pour les besoins de l'implémentation. Cependant, cette traduction ne prend pas en compte ces quatre caractères. Si, par exemple, le caractère `'\'` était utilisée dans une expression linguistique (ce qui n'est pas le cas dans nos expressions), il aurait la même signification que celle qui est définie dans le langage des expressions régulières.

Représentations alternatives de caractères : certains signes de l'écriture, tels que les apostrophes et les points de suspension, peuvent se trouver représentés par des caractères différents selon les logiciels. Lors de l'implémentation du moteur de repérage, nous avons pris en compte toutes les alternatives possibles pour chaque caractère¹⁰.

7.3 Définition des règles

Chaque règle correspond à une catégorie de l'incertitude, et renvoie à un ensemble d'expressions en langue naturelle qui l'expriment. Les règles utilisent des expressions écrites à l'aide de la grammaire formelle définie ci-dessus.

Chaque règle est constituée d'une partie positive, et d'une partie négative optionnelle (voir les règles en annexe F, p. 239) :

- **partie positive** : elle décrit des expressions linguistiques (d'une catégorie seulement) dont la présence d'au moins une d'elles est nécessaire dans une phrase afin de déclencher l'annotation de la phrase avec la catégorie correspondante. Si la règle ne contient pas de partie négative, la présence de l'expression linguistique dans la phrase déclenche automatiquement son annotation.

¹⁰Par exemple, lors d'une étape d'ajustement des règles linguistiques sur le premier corpus de 500 phrases, l'ajout d'un apostrophe alternatif (remplacement de « ' » par « [''] ») a changé sensiblement les résultats d'évaluation pour les catégories *i* et *i-q* et *i-q-hnv* : il a permis une augmentation du Rappel, avec une diminution de la Précision liée au nombre de faux positifs (FP). Pour *i* : le Rappel est passé de 0,991 à 1, la Précision de 0,880 à 0,701, et les FP de 33 à 104. Pour *i-q* : le Rappel est passé de 0,991 à 1, la Précision de 0,856 à 0,675, et les FP de 38 à 110. Pour *i-q-hnv* : le Rappel est passé de 0,976 à 1, la Précision de 0,661 à 0,455, et les FP de 64 à 153.

Exemple de partie positive :

"(...) [U/u]n doute [subsiste/reste présent] [concernant/à propos de] ..."

- **partie négative, optionnelle** : elle décrit des expressions linguistiques dont la présence est cherchée uniquement si la partie positive de la règle a proposé d'annoter la phrase. Si au moins l'une de ces expressions est présente dans la phrase, la règle ne produit pas d'annotation. Si elles sont absentes, l'annotation se fait avec la catégorie correspondante comme proposé par la partie positive.

Exemple de partie négative :

"[Si/Quand] un doute [subsiste/reste présent] [concernant/à propos de] ..."

Nous avons implémenté un moteur d'annotation sous python, permettant d'exécuter les règles de manière automatique. Lors de l'annotation, la même phrase peut être annotée avec une ou plusieurs catégories différentes.

7.4 Conception des règles

Chacune des catégories de l'ontologie est représentée par plusieurs règles de repérage. Chaque règle est conçue à partir des observations des expressions dans les phrases du corpus d'étude, en cherchant à prendre en compte un ensemble d'expressions le plus large possible. Par exemple, la règle présentée plus haut correspondrait entre autres à la phrase "Un doute reste présent à propos de l'efficacité de la bourse de carbone." (phrase imaginaire).

Dans un premier temps, nous avons construit des ensembles de règles en prenant pour modèle les expressions contenues dans le corpus d'étude. Nous avons analysé la totalité des 700 phrases de ce corpus. Dans un deuxième temps, l'efficacité de chaque règle a été testée, et des améliorations ont été apportées jusqu'à ce que la qualité des résultats soit satisfaisante. Les règles sont aussi consolidées en ajoutant des variantes à leurs sous parties, notamment des sous-expressions alternatives synonymes. Les phrases annotées manuellement avec *aucune incertitude* dans le corpus, ont été utiles en particulier pour créer les parties négatives des règles.

Pour faciliter à la fois l'écriture manuelle des règles et leur lecture par le système, elles sont écrites dans un fichier CSV, chacune sur une ligne comportant : un numéro unique identifiant de la règle, l'annotation de la catégorie d'incertitude,

7.5. EXEMPLES DE RÈGLES

la partie positive de la règle, et la partie négative de la règle si elle existe (voir schéma 7.1).

75	255 i-m-val	[... may be # .../^[a(n)/the] #()[%/p] [[/A/a]/[R/r]ound /[/A/a]bout /[/A/a]pproximately /
76	256 i-q-rab	[[might/could /seem([s/ing]) to] be (p) [stipulate(s) (...)that ... may be .../... [factor/re
77	257 i-m-int	[[([/A/a]/[R/r]ound /[/A/a]bout /[/A/a]p) estimate of [a/the] ...
78	258 i-q-rin	[[may be/might be/could be/seem(s) [^was^ (...), as may be expected/rather than(_)(

Schéma 7.1: Règles linguistiques : aperçu du document CSV. Colonnes visibles : numéro identifiant de la règle, catégorie d'incertitude, partie positive.

7.5 Exemples de règles

Exemple d'une règle courte déclenchant l'annotation *hypothèse non vérifiée*

- **Partie positive :**

"[A/a]lthough ... [suggest(s)/[is/are] suggesting] that ... [may/might/could] ..."

- **Partie négative :**

"[A/a]lthough ... [is/are/do(es)][not/n['/']t] suggest(ing) that ... [may/might/could] ..."

- **Une phrase où cette règle doit déclencher l'annotation *i-q-hnv* :**

« *Although AMO signal is correlated with the tropical Atlantic SSTG, previous studies suggest that it may be physically independent from the SST changes in the Table 2 Number of years with statistically significant at 5% level positive (PP) and negative (NP) phases for the SST variability indices considered here DJF* »

Exemple d'une règle déclenchant l'annotation *raisonnement abductif*

- **Partie positive :**

"[is [likely/possible]/seems/appears/] that [the/a(n)] ...
[is/are/was/[may/might/could] be] [due to/([the/a])result/a consequence]
[of/from]] ..., [while/but/however/though/nevertheless] [the/a(n)] ...
[is/are/was/[may/might/could] be] [due to/([the/a])result/a consequence]
[of/from]] ..."

- **Partie négative** : \emptyset
- **Une phrase où cette règle doit déclencher l'annotation *i-q-rab*** :

« *It is likely that the night-time difference is the result of poor relative calibration between two different types of Ta sensors, while the greater discrepancy during the day may be due to solar heating of the metal screen in which the Ta /RH probe is housed.* »

7.6 Exemples de modifications de règles

En fonction des expressions d'incertitude qu'elles décrivent, les règles sont rédigées et améliorées de façon itérative selon plusieurs procédés, dont les suivants :

- **création** : soit de la partie positive (lors de la première création d'une règle), soit de la partie négative pour préciser des exceptions, soit des deux en même temps.
- **extension ou correction** : soit de la partie positive, soit de la partie négative.
- **regroupement** : plusieurs règles en deviennent une seule.
- **différenciation** : une seule règle est remplacée par plusieurs règles.
- **requalification** : la catégorie d'incertitude associée à une règle est modifiée.

Dans cette section, nous présentons des exemples. Pour faciliter la lecture, certaines règles sont indiquées en réduisant leur longueur ou complexité, ce qui est signalé à chaque occurrence. La version complète des règles est consultable en annexe F (p. 239).

Ajout d'une partie négative, à une règle qui ne comportait qu'une partie positive

Données :

7.6. EXEMPLES DE MODIFICATIONS DE RÈGLES

- phrase *P* : « *The clean development mechanism (CDM)2, for example, limits the accounting of carbon pools for certain types of forestation projects to those within a project boundary and stipulates three different ways that project-specific baseline scenarios may be set* ».
- règle identifiante (réduite pour le propos) 256-*i-q-rab* uniquement positive : "[T/t]he(_)(_)(possible) mechanism(s)(...) [may/might/could] be ..."

Problème :

- la phrase *P* est annotée manuellement *ai*, mais détectée automatiquement par 256-*i-q-rab*, ce qui est incorrect.

Remarque :

- c'est le sens de la sous-expression « [...] *stipulates* [...] *that* [...] » qui permet de reconnaître que la phrase *P* est *ai* puisqu'elle est une assertion qui indique des procédures alternatives ; contrairement, par exemple, à une hypothétique autre phrase « *The clean development mechanism may be* [...] ».

Solution adoptée :

- pour constituer une exception à la règle 256-*i-q-rab* positive, une règle *i-q-rab* négative est créée :
 - 256-*i-q-rab* négative : "stipulate(s) (...)that ... may be ...".

Extension de la partie négative d'une règle

Données :

- phrase *P* : « *One reason that efforts at science communication may be ineffective, especially on controversial issues, is because they are often based on a 'deficit model' of public understanding of science.* ».
- règle identifiante (réduite pour le propos) 256-*i-q-rab* positive : "[[T/t]he/[A/a](n)/[O/o]ne/[A/a]nother](_)(_)([possible/plausible]) [factor/reason/explanation/motive/motivation/cause/mechanism/causation/determinant/origin/provenance](s)(...) [may/might/could/seem([s/ing]) to] be ...".
- 256-*i-q-rab* négative : "stipulate(s) (...)that ... may be ...".

Problème :

- la phrase *P* est annotée manuellement *ai*, mais détectée automatiquement par *256-i-q-rab*, ce qui est incorrect.

Remarque :

- c'est le sens de la sous-expression « ... *that ... is because ...* » qui permet de reconnaître que la phrase *P* est *ai* puisqu'elle est une assertion qui indique la certitude d'une cause fréquente ; contrairement, par exemple, à une hypothétique autre phrase « *One reason of efforts at science communication may be because they are often based on a 'deficit model' of public understanding of science.* ».

Solution adoptée :

- pour constituer une nouvelle exception à la règle *256-i-q-rab* positive, la règle *i-q-rab* négative est modifiée par l'ajout d'une nouvelle exception :
 - *256-i-q-rab* négative : "[stipulate(s) (...)that ... may be .../
... [factor/reason/explanation/motive/motivation/cause/mechanism/
causation/determinant/origin/provenance](s) (...)that ... may be ...
[is/are] because ...]".

Reformulation de deux règles en une

Données :

- phrase *P* : « *Taken together, therefore, it is possible that in shoots the reduced DM and N content as well as N concentration with elevated [CO₂] alone could be from faster and favorable translocation of photosynthates and N from shoots to grains.* ».
- règle identifiante *205-i-q-hnv* uniquement positive : "[(remaining)possibility is/it [is/[may/might/can/could] be] (_)possible/there [exists/is](_) the possibility](_) that ... [may/might/could]".
- règle identifiante *256-i-q-rab* uniquement positive : "[might/could] be (originated)from ...".

Problème :

7.6. EXEMPLES DE MODIFICATIONS DE RÈGLES

- la phrase *P* est annotée manuellement uniquement *i-q-rab*, mais détectée automatiquement à la fois par *205-i-q-hnv* et *256-i-q-rab*.

Remarque :

- l'expression « *it is possible that [...] could* » détectée par la règle *205-i-q-hnv* est en fait une partie de l'expression *i-q-rab* de la phrase *P*, dont l'expression *could be from* détectée par la règle *256-i-q-rab* est le pôle.

Solution adoptée :

- la règle positive correspondant à l'expression *i-q-rab* est ajouté comme règle négative de la règle *i-q-hnv*. La nouvelle règle *205-i-q-hnv* est :
 - 205-i-q-hnv* partie positive (identique) "[remaining]possibility is/it [is/[may/might/can/could] be] (_)possible/there [exists/is](_) the possibility](_) that ... [may/might/could]"
 - 205-i-q-hnv* partie négative : "[might/could] be (originated)from ..."

Dissociation d'une seule expression en deux expressions de catégories différentes

Données :

- phrase *P* : « *The minimum future time period of validity for the baseline will be dictated by GHG mitigation program rules and by the intervals over which data become available, which may be 5 years or more for deforestation reduction projects that rely on remote sensing imagery.* ».
- règle identifiante *254-i-q-hnv* partie positive : "... may be ...".

Remarque :

- l'expression « *may be 5 years* » dans la phrase *P* correspond à la fois à la catégorie *i-q-hnv* déjà identifié et à la catégorie *i-m-val* non identifié.

Problème :

- comment prendre en compte les deux catégories de l'expression ?

Solution adoptée :

- d'une part, l'expression continue à être détectée par la règle *254-i-q-hnv* qui préexistait et n'est pas modifiée ; et, d'autre part, une nouvelle règle *i-m-val* est ajoutée :
 - règle *255-i-m-val* uniquement positive : "... may be # ..."

Priorité d'une expression d'une catégorie sur une expression d'un autre catégorie

Données :

- phrase *P* : « *Hence an uncertainty of around 5% may be expected in this study.* ».
- règle identifiante 257-*i-m-int* uniquement positive : "[[A/a][round/bout] /±()]#".
- règle identifiante 254-*i-q-hnv* partie positive : "... may be ...".
- règle 254-*i-q-hnv* partie négative : règle négative longue de 1395 caractères (voir règles en annexe F, p. 239).

Problème :

- la phrase *P* est annotée manuellement uniquement *i-m-int*, mais détectée automatiquement à la fois par 257-*i-m-int*, ce qui est correct, et par 254-*i-q-hnv*, ce qui est incorrect.

Remarque :

- dans la phrase *P*, l'expression *i-m-int* « *around 5%* » dirige le sens de l'expression « *may be expected* », et cette dernière ne peut alors plus être la marque de la catégorie *i-q-hnv*.

Solution adoptée :

- la règle 257-*i-m-int* est ajoutée à la partie négative de la règle 254-*i-q-hnv*, pour y constituer une nouvelle exception.

Remarque à posteriori :

- le raisonnement qui mène à cette solution illustre particulièrement le fait que des catégories d'incertitudes annotées automatiquement du point de vue des règles (ici deux catégories) peuvent être annotées manuellement différemment du point de vue des experts (ici une seule catégorie). Dans ces cas de contradiction, il serait utile de pouvoir reconsidérer de manière plus systématisée qui du système ou des experts apporte la solution valable (comment/pourquoi). Sur ce point, parce que des notes informelles

ont été ajoutées spontanément lors de l'annotation manuelle par les experts, nous avons remarqué l'aide méthodologique que pourraient apporter des notes explicatives des choix d'experts : pour l'amélioration future de la méthodologie, et pour d'éventuelles découvertes de raisonnements sous-jacents, des règles explicatives associées aux choix des experts nous paraissent donc envisageables et souhaitables.

Augmentation de la précision de la catégorie d'une expression

Données :

- phrase *P* : « *In tropical lianas it may be that an active process associated with the abundant xylem parenchyma is especially well-developed, and thus protects against hydraulic failure.* ».
- règle identifiante (réduite pour le propos) *256-i-q-rab* partie positive : "[may/might/could/seem([s/ing]) to^] be (_)(_)(_)(_)[associated(_) with^/linked(_) to^/related(_) to^] ...".
- règle *256-i-q-rab* partie négative : règle négative longue (voir règles en annexe F, p. 239).

Problème :

- la phrase *P* est annotée manuellement uniquement *i-q*, ce qui est correct mais insuffisant, et détectée automatiquement par *256-i-q-rab*, ce qui est correct.

Remarque :

- l'annotation *i-q-rab* inclue aussi la catégorie plus général *i-q*.

Solution adoptée :

- la phrase *P* est ré-annotée manuellement *i-q-rab* au lieu de *i-q*.

Arbitrage de la catégorie d'une phrase

La phrase « *The GPR survey results suggest there may be intense ice body melting in some areas of the glacier.* » a d'abord été annotée manuellement avec la catégorie *i-q-rab*, en considérant qu'elle pourrait indiquer que « *intense ice body*

melting in some areas of the glacier » est une cause plausible des « *GPR survey results* ». Cependant, l'annotation automatique lui a attribué la catégorie *i-q-hnv*, puisque l'expression « *suggest there may be* » peut indiquer l'hypothèse d'un événement situable dans le passé, le présent ou le futur. Après analyse, c'est finalement la catégorie *i-q-hnv* qui a été attribuée à la phrase.

7.7 Les expressions contenant « *may be* »

Les expressions contenant « *may be* » se sont révélées particulièrement diverses et pouvant correspondre à plusieurs catégories de l'incertitude ou même à aucune incertitude. Nous avons aussi constaté leur fréquence relativement élevée lors de l'examen manuel d'articles scientifiques. Elles constituent donc un phénomène sur lequel il est utile de s'attarder en particulier.

La complexité de traitement manuel des phrases contenant l'expression « *may be* » a révélé un besoin de critères unifiés et explicites, alors que l'explication de leur choix d'annotation, phrase par phrase, n'avait pas été demandé aux experts.

Les 500 phrases du corpus d'étude ont donc été réexaminées par le rédacteur des règles, pour un résultat contenant pour chaque phrase :

- la phrase entière,
- la ou les catégorie(s) d'incertitude annotée(s) pour les expressions dans la phrase,
- le pattern considéré pour l'annotation,
- un commentaire nommant la cause générique de rattachement à une ou plusieurs catégorie(s).

La plupart des phrases observées contenant « *may be* » expriment une hypothèse non vérifiée ('*i-q-hnv*'). Pour certaines, un désaccord entre les annotations des trois experts a été constaté ; et elles ont fait l'objet d'un arbitrage final par l'expert auteur des règles afin d'utiliser des critères unifiés.

Exemples de phrases non '*i-q-hnv*' contenant « *may be* »

- la phrase « *Recognizing and quantifying this equivalence in solar radiation regimes may be helpful in selecting the type of study that can be carried out in the B2C with most real-world relevance.* » sera classée '*i-q-hnv*', car

sa forme « may be ... that can be ... » ne contient pas d'expression qui l'empêche clairement d'être une hypothèse ;

- tandis que la phrase « *At the ABRACOS Reserva Jaru site, for instance, monthly rainfall may be less than 10 mm per month in the dry season, but at other times of the year, rainfall at this site (and at the other ABRACOS sites) can be several hundreds of millimeters.* » sera classée 'ai', car sa forme « may be ... but ... can be ... » contient une expression qui l'empêche clairement d'être une hypothèse, car 'may' pourrait être remplacé par 'can' et indique une alternative au lieu d'une possibilité, et elle n'exprime pas d'incertitude.

Parmi les phrases contenant 'may be' et classées différemment de 'i-q-hnv', se trouvent par exemple :

- notée 'ai' : « *because government capacity may be limited* ».
- notée 'i-m-int' et 'i-q-rin' : « *Hence an uncertainty of around 5% may be expected in* ».
- notées 'i-q-rab' : « *may be related to water availability* », « *The reasons for the high sensitivity to rainfall may be that* », « *Another factor influencing our findings may be the presence of* », « *The different observations may be due to the* », « *these circulation anomalies [...] may be responsible for* ».
- notée 'i-q-rin' : « *a change in diet composition [...] may be expected* », « *In the future [...] the central positions of intense ablation may be disconnected* ».

7.8 Algorithme d'exécution des règles

Le moteur d'annotation que nous avons produit (voir algorithme en annexe E, p. 233) traduit automatiquement la partie positive et négative de chaque règle linguistique en expressions régulières. Ensuite, pour chaque phrase avec chaque règle, il vérifie la présence dans la phrase d'une expression correspondant à la partie positive de la règle. Si une expression correspondant à la partie positive de la règle est présente, alors le moteur vérifie dans la phrase l'absence d'une expression correspondant à la partie négative. Si aucune expression correspondant

à la partie négative n'est présente, le moteur déclenche l'annotation comme décrit plus haut.

Durant la traduction des expressions linguistiques en expressions régulières, les caractères réservés sont remplacés dans un ordre prédéterminé. Par exemple, tous les '[' sont traduits puis tous les ']' puis tous les '/'.

Exemple de traduction d'une expression linguistique en expression régulière.

L'exemple ici est celui d'une règle particulièrement courte, pour que sa traduction en expression régulière reste lisible :

- **Partie positive :**

"[A/a]n uncertainty of ([around/about/approximately/≈])#"

- **Partie négative :** Ø

- **Traduction des sous-parties de la partie positive** (dans cet ordre) :

- Délimiteurs de parties non obligatoires (et)
→ "([aroundlaboutlapproximatl≈])?"
- Délimiteurs de parties alternatives [A/a]
→ "(Ala)"
- Délimiteurs de parties alternatives [around/about/approximately/≈]
→ "(aroundlaboutlapproximatl≈)"
- Nombre #
→ "[+\\-]?[0-9][0-9,.]*"

- **Expression régulière résultante :**

"(Ala)n uncertainty of ((aroundlaboutlapproximatl≈))?[+\\-]?[0-9][0-9,.]*"

- **Une phrase annotée par cette règle avec la catégorie *i-m-int* :**

« *Hence an uncertainty of around 5% may be expected in this study.* »

Chapitre 8

Résultats et évaluation des annotations

Evaluer le moteur d’annotation de l’incertitude signifie évaluer sa capacité à identifier correctement les catégories d’incertitude exprimées dans les phrases. Ce chapitre tente de répondre aux questions suivantes :

- Dans quelle mesure l’annotation automatique (grâce aux règles) des catégories de l’incertitude est-elle conforme aux annotations manuelles de ces mêmes phrases ?
- Le système est-il adapté au traitement de nouveaux corpus de phrases ?
- Le cadre de rédaction des règles linguistiques (grammaire formelle) permet-il de les améliorer progressivement en prenant en compte de nouveaux corpus d’étude ?

8.1 Protocole d’évaluation du système

Le corpus d’évaluation que nous avons utilisé contient 300 phrases annotées manuellement (voir annexe [G.3](#), p. [275](#)). Nous avons calculé les indices de *Précision*, *Rappel* et *F-mesure*.

Les résultats de l’annotation automatique du corpus d’évaluation sont comparés avec les annotations correctes (manuelles), pour chacune des catégories de l’incertitude (voir schéma [6.3](#), p. [135](#)). Ces comparaisons produisent les valeurs suivantes :

8.1. PROTOCOLE D'ÉVALUATION DU SYSTÈME

- *Vrais positifs* (VP) : le nombre de phrases annotées correctement.
⇒ Pour ces phrases, la catégorisation automatique fonctionne.
- *Faux positifs* (FP) : le nombre de phrases annotées automatiquement, mais pas manuellement. Ce sont des détections erronées.
⇒ Pour ces phrases, la catégorisation automatique ne fonctionne pas.
- *Faux négatifs* (FN) : le nombre de phrases non annotées, mais qui sont pourtant identifiées manuellement. Ce sont des non-détections erronées.
⇒ Pour ces phrases, la catégorisation automatique ne fonctionne pas.
- *Vrais négatifs* (VN) : le nombre de phrases où aucune expression de l'incertitude n'est identifiée, ni automatiquement ni manuellement. Ce sont des non-détections correctes.
⇒ Pour ces phrases, la catégorisation automatique fonctionne.

Les valeurs mentionnées ci-dessus peuvent se présenter sous la forme d'un tableau de contingence¹ appelé *matrice de confusion* (*confusion matrix*), tel qu'indiqué dans le tableau 8.1.

	Incertainie détectée	Incertainie non détectée
Incertainie présente	<i>VP</i>	<i>FN</i>
Incertainie absente	<i>FP</i>	<i>VN</i>

Tableau 8.1: Matrice de confusion (modèle explicatif).

La Précision (*Precision*) indique la proportion de phrases identifiées correctement parmi toutes les phrases identifiées.

$$Précision = \frac{\text{Nombre de phrases à incertitude correctement identifiées}}{\text{Nombre de toutes les phrases identifiées}}$$

Exprimée autrement :

$$Précision = \frac{VP}{VP + FP}$$

¹Pour chaque catégorie d'incertitude de l'ontologie

Le Rappel (*Recall*, ou *Sensitivity*, ou *true positive rate* ou *probability of detection*) indique la proportion de phrases identifiées correctement parmi l'intégralité des phrases qui auraient dû être identifiées.

$$Rappel = \frac{\text{Nombre de phrases à incertitude correctement identifiées}}{\text{Nombre de toute les phrases à incertitude}}$$

Exprimé autrement :

$$Rappel = \frac{VP}{VP + FN}$$

La F-mesure (*F-score* ou *F₁ score*) est dite aussi *moyenne harmonique de la Précision et du Rappel*.

$$F\text{-mesure} = \frac{2 \times Précision \times Rappel}{Précision + Rappel}$$

8.2 Expérimentations lors de la phase de construction des règles

Dans ce chapitre, nous présentons les résultats de l'expérimentation sur le corpus d'étude. Ces résultats montrent dans quelle mesure l'annotation automatique des phrases par les règles linguistiques (voir les règles en annexe F, p. 239) est correcte par rapport aux annotations manuelles des experts. Les évaluations se font de manière indépendante pour chaque catégorie de l'ontologie : *incertitude* (i), *incertitude mesurée* (i-m), *incertitude qualitative* (i-q), *intervalle de mesure ou de probabilité* (i-m-int), *valeur ponctuelle* (i-m-val), *raisonnement inductif* (i-q-rin), *raisonnement abductif* (i-q-rab), *hypothèse non vérifiée* (i-q-hnv).

Rappelons que le corpus d'étude comporte un total de 700 phrases, dont les 500 ont été extraites automatiquement (partie 1) et les 200 autres manuellement (partie 2).

8.2.1 Expérimentation sur le corpus d'étude (partie 1)

Ici nous traitons les 500 premières phrases du corpus d'étude (voir le corpus en annexe G.1, p. 247).

8.2. EXPÉRIMENTATIONS LORS DE LA PHASE DE CONSTRUCTION DES RÈGLES

Première version des règles linguistiques

Une première version des règles linguistiques a été produite au nombre de 245, rédigées spécifiquement en s'appuyant sur les expressions-candidates de ce premier corpus.

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	1,00	0,85	0,92	327	0	114	59
i-m	1,00	0,78	0,88	7	0	491	2
i-q	0,99	0,86	0,92	325	1	121	53
i-m-int	1,00	1,00	1,00	6	0	494	0
i-m-val	1,00	1,00	1,00	1	0	499	0
i-q-rin	0,93	1,00	0,97	30	2	468	0
i-q-rab	1,00	0,75	0,86	87	0	384	29
i-q-hnv	1,00	0,90	0,95	217	0	259	24
Moyenne	0,99	0,86	0,92				

Tableau 8.2: Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec la première version des règles.

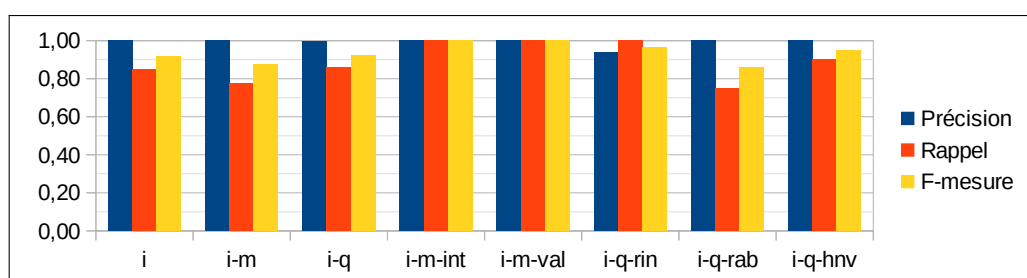


Schéma 8.1: Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec la première version des règles : Précision, Rappel, F-mesure.

Les résultats de cette première expérimentation sont présentés dans le tableau 8.2, et le schéma 8.1. Les indices moyens pondérés ont été calculés selon le mode de calcul présenté dans l'annexe D, p. 229.

Ces résultats montrent une Précision de 1 pour toutes les catégories d'incertitude, sauf pour 3 faux positifs (1 *i-q* et 2 *i-q-rin*). Pour les catégories d'incertitude mesurée les phrases sont moins nombreuses, mais le Rappel est à 1, ce qui montre que les expressions d'incertitude de ces catégories sont correctement reconnues, même si cela reste à démontrer sur plus d'occurrences. Les

valeurs élevées de la Précision (0.99 en moyenne), pour un Rappel de 0.86, démontrent que la méthodologie de rédaction des règles fonctionne : les règles, rédigées suite à l'analyse de ce corpus, permettent d'annoter ce même corpus avec une bonne qualité globale des annotations.

Les faux négatifs sont dus à des expressions de catégories *i-q-rab* et *i-q-hnv*, qui sont celles les plus présentes dans le corpus à ce niveau de granularité maximal : le système a donc tendance à ne pas reconnaître certaines expressions de ces catégories. Pour la catégories la plus générale d'incertitude (*i*) qui inclue toutes les autres catégories, le nombre de vrais négatifs coïncide avec les 114 phrases du corpus qui n'ont aucune incertitude : toutes les phrases qui n'ont aucune incertitude ont donc été correctement identifiées. En général, le nombre élevé de vrais négatifs et relativement bas de faux négatifs montre que le système discrimine relativement bien les expressions de catégories les unes des autres.

La catégorie la plus fréquente dans le corpus au niveau de plus fine granularité est celle des hypothèses non vérifiées, suivi des raisonnements abductifs, suivi des raisonnements inductifs. Le faible nombre de phrases *i-m*, *i-m-int* et *i-m-val*, montre que ces catégories sont trop peu représentées et donc que leurs résultats sont non représentatifs.

Les résultats de l'expérimentation avec ce premier corpus et ces premières règles sont donc satisfaisants sur la performance du système, mais ils demandent une diversification des expressions-candidates dans les catégories, et un plus grand nombre de ces expressions-candidates dans le corpus utilisé.

Les règles linguistiques dans leur version finale

L'ensemble final de règles que nous avons obtenues après les corrections contient 79 règles (voir annexe F, p. 239). Elles ont été rédigées en unifiant une partie des précédentes 245 règles, et en s'appuyant sur les expressions d'incertitude des phrases du corpus d'étude de 200 phrases. Les résultats de cette expérimentation sont présentés sur le tableau 8.3 et le schéma 8.2.

La valeur du Rappel et celle de la Précision restent les mêmes pour cette version finale des règles, 0,86 et 0,99, mais le peu de phrases sur lesquelles reposent les données dans les catégories *i-m-int* et *i-m-val* ne permettent pas de conclure sur l'efficacité des règles de ces catégories. La Précision élevée montre que cette dernière version des règles permet d'annoter correctement une grande majorité des phrases, et le Rappel élevé montre que la majorité des phrases de chaque catégorie ont été reconnues. Par conséquent, nous pouvons conclure que la réduction

8.2. EXPÉRIMENTATIONS LORS DE LA PHASE DE CONSTRUCTION DES RÈGLES

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	0.99	0.85	0.92	327	1	113	59
i-m	0.47	0.78	0.58	7	8	483	2
i-q	0.99	0.86	0.92	325	1	121	53
i-m-int	0.43	1.00	0.60	6	8	486	0
i-m-val	1.00	1.00	1.00	1	0	499	0
i-q-rin	0.94	1.00	0.97	30	2	468	0
i-q-rab	1.00	0.75	0.86	87	0	384	29
i-q-hnv	1.00	0.90	0.95	217	0	259	24
Moyenne	0.99	0.86	0.92				

Tableau 8.3: Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec les règles finales.

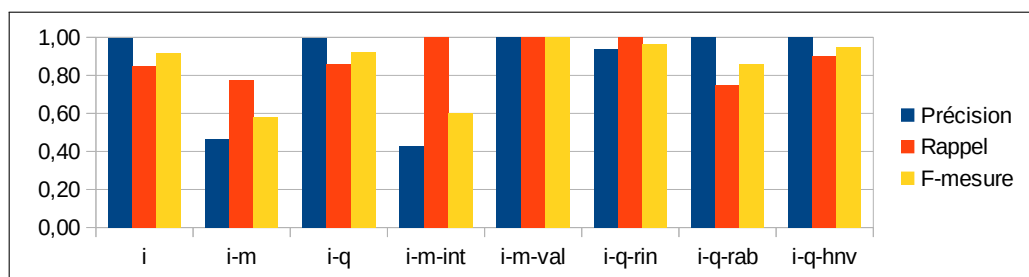


Schéma 8.2: Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec les règles finales : Précision, Rappel, F-mesure.

du nombre des règles, de 245 à 79, et leur adaptation pour prendre en compte des expressions supplémentaires, s'est faite sans perte significative de leur performance.

8.2.2 Expérimentation sur le corpus d'étude (partie 2)

Cette expérimentation avec la partie 2 du corpus d'étude, de 200 phrases (voir corpus en annexe G.2, p. 267), à été faite avec la dernière version des règles (voir annexe F, p. 239), dont la rédaction s'appuie en partie sur l'analyse de ce même corpus. Les résultats sont présentés dans le tableau 8.4 et le schéma 8.3.

Puisque ces nouvelles 200 phrases ont été extraites manuellement, elles représentent plus exhaustivement les expressions d'incertitude de leurs articles d'origine. Les valeurs moins élevées de la Précision et du Rappel reflètent la moindre facilité de prise en compte de ces nouvelles phrases, dont la diversité des

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	0.99	0.45	0.62	87	1	5	107
i-m	0.91	0.52	0.66	42	4	115	39
i-q	0.91	0.37	0.52	42	4	81	73
i-m-int	0.90	0.59	0.71	37	4	133	26
i-m-val	1.00	1.00	1.00	4	0	196	0
i-q-rin	0.57	0.22	0.32	4	3	179	14
i-q-rab	0.88	0.24	0.38	7	1	170	22
i-q-hnv	0.82	0.38	0.52	27	6	123	44
Moyenne	0.91	0.43	0.58				

Tableau 8.4: Résultats de l'expérimentation sur le corpus d'étude (200 phrases, partie 2).

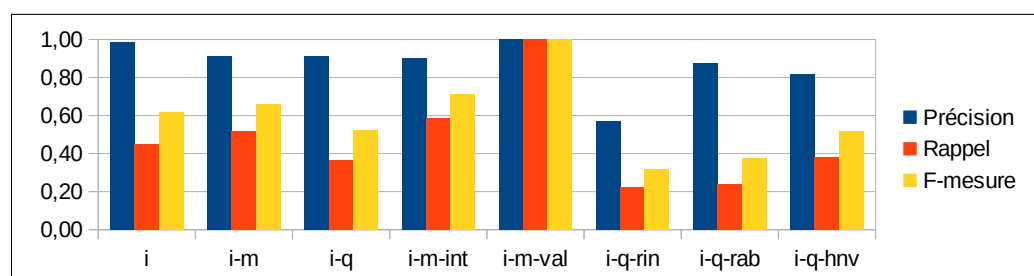


Schéma 8.3: Résultats de l'expérimentation sur le corpus d'étude (200 phrases, partie 2). : Précision, Rappel, F-mesure.

expressions est plus grande. Par exemple, les dernières règles vont jusqu'à 2876 caractères pour la partie négative. La moindre F-mesure reflète donc que les 79 règles ont été perfectionnées jusqu'à un niveau qui dépend de la complexité² à modifier les règles déjà en place, de l'expérience de l'expert, et du temps imparti.

Malgré la complexité de rédaction qui devra être mieux gérée pour l'amélioration future des règles, la Précision reste à un niveau relativement élevé : le système est encore performant pour annoter correctement les phrases.

Les résultats obtenus sur tout le corpus d'étude de 700 phrases sont présentés sur le tableau 8.5 et le schéma 8.4.

Les valeurs de Précision élevées montrent que la dernière version des règles décrit correctement une proportion importante des catégories d'incertitude sans

²La complexité des changements des règles déjà en place dépend elle-même de l'ergonomie du système de rédaction des règles. Cette ergonomie a potentiellement une forte marge d'amélioration.

8.3. EVALUATION

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	0.99	0.71	0.83	414	2	118	166
i-m	0.80	0.54	0.65	49	12	598	41
i-q	0.99	0.74	0.85	367	5	202	126
i-m-int	0.78	0.62	0.69	43	12	619	26
i-m-val	1.00	1.00	1.00	5	0	695	0
i-q-rin	0.87	0.71	0.78	34	5	647	14
i-q-rab	0.99	0.65	0.78	94	1	554	51
i-q-hnv	0.98	0.78	0.87	244	6	382	68
Moyenne	0.97	0.72	0.82				

Tableau 8.5: Résultats de l'expérimentation sur le corpus d'étude (700 phrases).

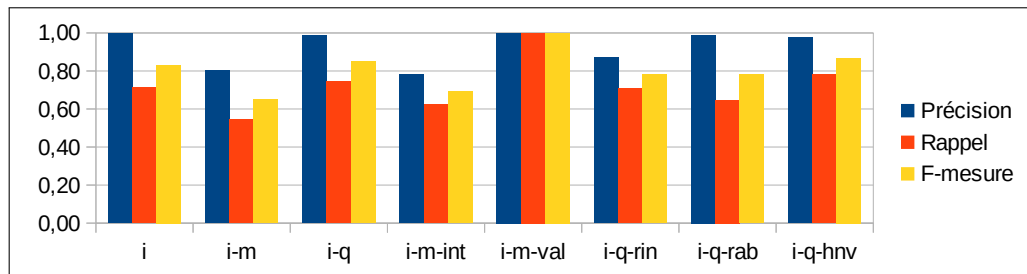


Schéma 8.4: Résultats de l'expérimentation sur le corpus d'étude (700 phrases) : Précision, Rappel, F-mesure.

les confondre. Les valeurs de Rappel nettement au dessus de la moyenne, mais majoritairement au-dessous de 0.80, montrent que la description des expressions d'incertitude par les règles devra encore être améliorée pour prendre en compte de nouvelles variantes d'expressions.

8.3 Evaluation

L'évaluation a été faite sur le corpus d'évaluation de 300 phrases³ (voir annexe G.3, p. 275) qui n'ont pas été utilisées lors de la création des règles. Les résultats de l'évaluation sont présentés sur le tableau 8.6 et le schéma 8.5.

³Le corpus totalisait 328 phrases lors de son extraction, et il a été réduit aux 300 phrases prévues en enlevant parmi ses dernières phrases (en commençant par la fin du corpus) les 28 annotées uniquement avec la catégorie *i-q-hnv* (catégorie la plus représentée dans les phrases de ce corpus).

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	0.90	0.24	0.37	62	7	31	200
i-m	0.88	0.28	0.43	30	4	190	76
i-q	0.76	0.17	0.28	29	9	122	140
i-m-int	0.91	0.38	0.54	30	3	218	49
i-m-val	N/A	0.00	N/A	0	0	285	15
i-q-rin	0.50	0.04	0.08	2	2	250	46
i-q-rab	0.67	0.09	0.15	2	1	276	21
i-q-hnv	0.44	0.15	0.22	14	18	186	82
Moyenne	0.78	0.21	0.33				

Tableau 8.6: Résultats de l'évaluation.

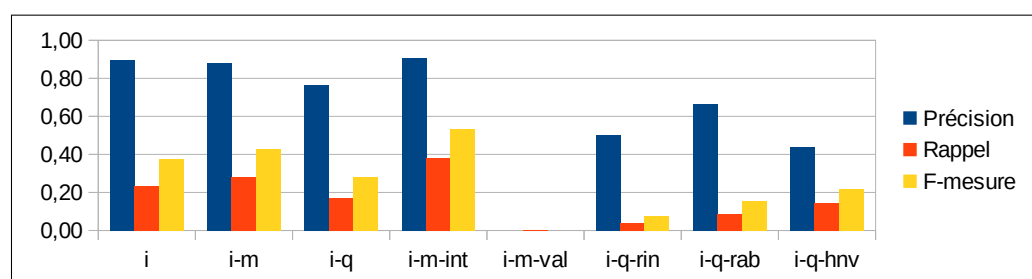


Schéma 8.5: Résultats de l'évaluation : Précision, Rappel, F-mesure.

La Précision moyenne est de 0,78. Cependant, les valeurs de la Précision sont assez basses pour certaines catégories (par ex. 0,44 et 0,50), ce qui montre que les règles que nous avons proposées ne décrivent pas totalement certaines des catégories d'incertitude.

Le Rappel moyen est de 0,21, ce qui montre que les règles laissent non annotées une grande proportion des phrases. Ceci reflète avant tout la complexité de l'expression de l'incertitude dans les articles.

Les règles que nous avons proposées peuvent être considérées comme un premier pas vers la construction d'un système performant d'annotation de l'incertitude. Les valeurs de la Précision restent relativement élevées et ceci indique la faisabilité de l'approche. Cependant, de nouvelles règles restent à être ajoutées afin pour pouvoir augmenter les valeurs du Rappel.

Nous notons que des imperfections des annotations automatiques peuvent être dues aux imperfections des règles, mais aussi à plusieurs autres facteurs, parmi lesquels : des erreurs d'annotation manuelle ; des problèmes liés à l'orthographe ou aux caractères mal encodés ; des coupures de ligne ou de mauvaises segmen-

8.3. EVALUATION

tations des phrases.

Catégorie	Précision	Rappel	F-mesure	VP	FP	VN	FN
i	0.98	0.57	0.72	476	9	149	366
i-m	0.83	0.40	0.54	79	16	788	117
i-q	0.97	0.60	0.74	396	14	324	266
i-m-int	0.83	0.49	0.62	73	15	837	75
i-m-val	1.00	0.25	0.40	5	0	980	15
i-q-rin	0.84	0.38	0.52	36	7	897	60
i-q-rab	0.98	0.57	0.72	96	2	830	72
i-q-hnv	0.91	0.63	0.75	258	24	568	150
Moyenne	0.94	0.56	0.70				

Tableau 8.7: Résultats sur le gold standard.

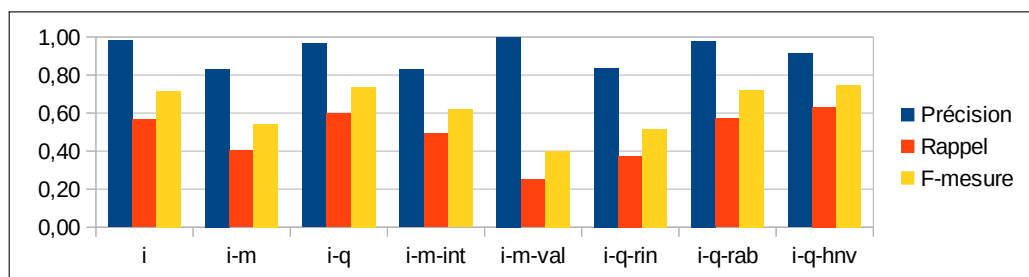


Schéma 8.6: Résultats sur le gold standard : Précision, Rappel, F-mesure.

Le corpus gold standard réunit le corpus d'étude et le corpus d'évaluation (voir le corpus en annexe G, p. 247). Les résultats sur ce corpus sont présentés sur le tableau 8.7 et le schéma 8.6. La Précision moyenne est de 0,94, ce qui indique que globalement les règles décrivent et discriminent bien les motifs d'expressions candidates déjà prévues. Le Rappel moyen est de 0,56.

Enfin, le schéma 8.7 présente le récapitulatif des indices de Précision, Rappel et F-mesure obtenus lors des expérimentations. Nous avons représenté (voir mode de calcul en annexe D, p. 229) la Précision moyenne, le Rappel moyen et la F-mesure moyenne des corpus d'étude (avec ses parties de 500 et 200 phrases), du corpus d'évaluation et du gold standard. Ce schéma montre les valeurs élevées de la Précision, entre 0,78 et 0,99 pour les différents corpus, avec une diminution nette de la Précision entre le corpus d'étude et le corpus d'évaluation. Les valeurs du Rappel sont moins importantes, entre 0,21 et 0,86.

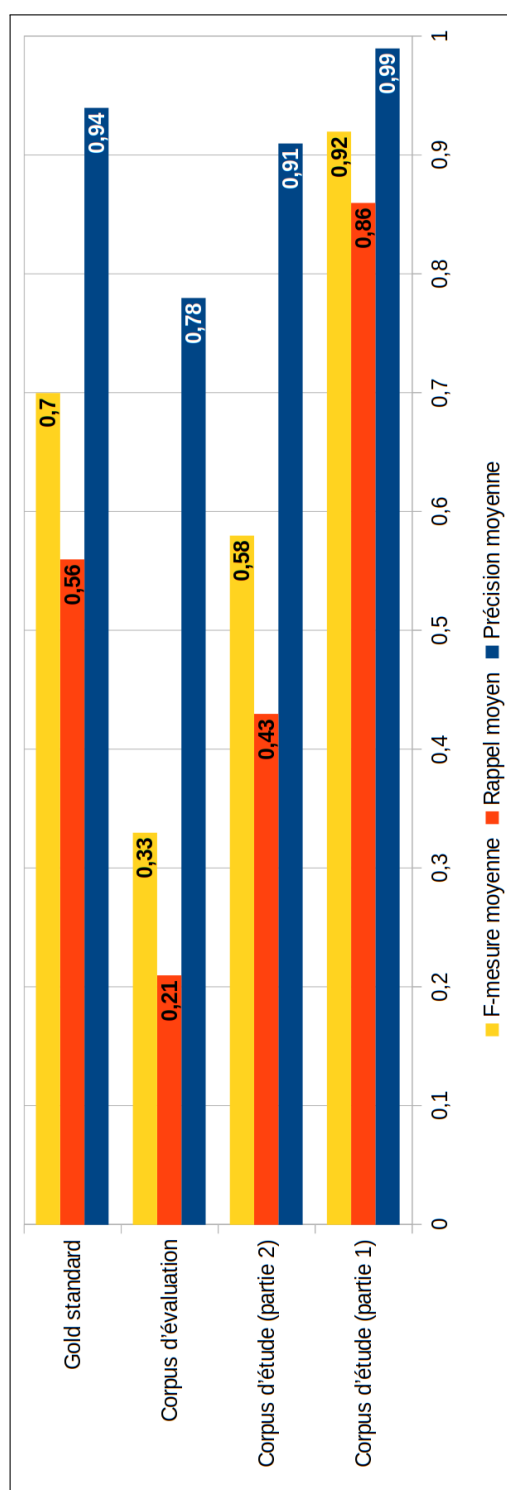


Schéma 8.7: Précision, Rappel et F-mesure pour les différents corpus.

Chapitre 9

Conclusion

Cette thèse doctorale a étudié et modélisé le concept d'incertitude scientifique dans toute sa diversité à travers les différents champs disciplinaires. L'incertitude occupe une place importante dans le processus de la recherche scientifique. Les hypothèses scientifiques, les marges d'erreur dans les mesures expérimentales, les intuitions des chercheurs sont autant de sources d'incertitude. Leur trace est identifiable dans les articles scientifiques qui sont le principal vecteur de communication des résultats scientifiques et de construction des raisonnements. Nous rappelons également le rôle que joue l'incertitude scientifique au sein des différentes institutions comme l'autorité européenne de sécurité des aliments ou les organisations devant gérer des informations de santé à travers la labellisation des sources et des certifications de sites d'information médicale afin de proposer un service de confiance auprès de la population. Il en va de même pour certaines données environnementales où la labellisation joue un rôle important. La gestion intelligente de l'incertitude est alors un défi sociétal.

Le concept d'incertitude véhicule des notions et des terminologies différentes d'un champ disciplinaire à l'autre. L'un des objectifs de cette thèse est de comprendre l'incertitude à travers son expression dans les articles scientifiques de différentes disciplines. Ce travail de nature linguistique sur la sémantique textuelle à partir de l'étude de corpus, propose une ontologie de l'incertitude scientifique qui s'articule autour de deux types qui sont respectivement : l'incertitude qualitative et l'incertitude mesurée. Nous proposons un algorithme d'identification et de classification des incertitudes dans les textes scientifiques, qui fonctionne grâce à des règles linguistiques. L'outil que nous avons construit permet l'annotation automatique des contenus textuels pour pouvoir extraire et classer les expressions d'incertitudes exprimées dans de grandes collections d'articles scientifiques.

L'originalité de cette thèse, nous semble-t-il, porte sur la dimension multidisciplinaire selon laquelle est abordé le concept d'incertitude. L'état de l'art montre que selon les différentes institutions et disciplines qui tentent de formaliser cette notion, l'incertitude est considérée d'un point de vue d'une graduation : du plus incertain au moins incertain, à des fins d'analyses quantitatives. Notre objectif était, au contraire, de travailler sur le réseau de concepts qu'implique l'incertitude en sciences, sans réduire la complexité sous-jacente à une échelle graduée. L'ontologie de l'incertitude est le produit de cet effort de conceptualisation.

Nous avons choisi comme domaine d'application le domaine du changement climatique. Notre expérimentation, qui porte sur un corpus de presque 5000 articles scientifiques, nous a permis de valider et évaluer notre approche de l'annotation automatique. Disposer d'outils permettant l'analyse du discours scientifique est essentiel pour répondre à des problématiques sociétales actuelles, telles que le changement climatique.

Les principaux résultats de la thèse consistent en :

- la proposition d'une ontologie du concept d'incertitude scientifique, d'un point de vue transdisciplinaire ;
- la construction d'un corpus de 1000 phrases annotées manuellement avec les catégories de l'incertitude selon l'ontologie proposée - gold standard. Ce corpus a pour vocation d'être publié en accès libre ;
- la construction de ressources linguistiques - grammaire formelle et règles de repérage et de classification de l'incertitude dans les textes - qui ont été mises à disposition de la communauté scientifique. La qualité de ces règles a été évaluée.

Les développements dans le domaine du Traitement Automatique des Langues ces 20 dernières années exploitent les possibilités des calculs vectoriels pour l'analyse de données textuelles. Cette tendance s'est accentuée avec l'introduction des réseaux de neurones qui ont rendu possible des avancées notables. Cependant, toutes les méthodes d'apprentissage automatique et d'apprentissage profond nécessitent de grandes masses de données prétraitées. Dans le cas de l'annotation de l'incertitude dans les textes, la mise en place de telles méthodes exigerait la disponibilité de grands corpus annotés. Le Gold standard que nous proposons, et dont la qualité des annotations a été contrôlée, pourrait servir à de telles approches.

Dans ce contexte, notre méthodologie est tournée résolument vers les ressorts de la linguistique et de la sémantique textuelle. Le travail sur les marqueurs linguistiques et la production de règles d'annotation permet, au-delà de la construction d'un outil informatique, de décrire et comprendre les moyens linguistiques qui sont mobilisés par un concept. Ainsi, les algorithmes produits et le processus d'annotation restent compréhensibles pour l'être humain, et les ressources linguistiques peuvent servir à des études ultérieures, plus approfondies du concept d'incertitude. Dans ce sens, l'apport principal de la thèse se situe sur le plan linguistique, et l'expérimentation informatique nous permet de valider cette approche et les résultats obtenus.

Références bibliographiques

- [Académie française, 1935] ACADÉMIE FRANÇAISE (1932-1935). *Dictionnaire de l'Académie française*, 8^e édition. Académie française, Paris.
- [Accart et Rivier, 2012] ACCART, J.-P. et RIVIER, A. (2012). *Mémento de l'information numérique*. Electre - Editions du Cercle de la Librairie, Paris.
- [Ahir *et al.*, 2020] AHIR, H., BLOOM, N. et FURCERI, D. (2020). 60 Years of Uncertainty - Our new index provides novel insights into an amorphous concept. *Finance Development*, mars 2020, vol. 57, n^o 1.
- [Aikhenvald, 2004] AIKHENVALD, A. Y. (2004). *Evidentiality*. Oxford University Press Inc., New York, EUA.
- [Andreewsky et Bourcier, 2000] ANDREEWSKY, E. et BOURCIER, D. (2000). Abduction in language interpretation and law making. *Kybernetes*, 2000, vol. 29, n^o 7/8, pages 836–845.
- [Assanzadeh *et al.*, 2011] ASSANZADEH, S., MARTINAIS, V., STEYER, E., TILLEMENT, S. et REVERDY, T. (2011). Approches de l'incertitude et son impact sur la décision. Rapport technique in Cahiers de la Sécurité Industrielle, n^o 2011-06, Institut pour une Culture de Sécurité Industrielle, Toulouse, France (ISSN 2100-3874), Fondation pour une Culture de Sécurité Industrielle - FonCSI, Toulouse, France.
- [Atanassova, 2012] ATANASSOVA, I. (2012). *Exploitation informatique des annotations sémantiques automatiques d'Excom pour la recherche d'informations et la navigation*. Thèse de doctorat.
- [Atanassova *et al.*, 2016] ATANASSOVA, I., BERTIN, M. et LARIVIERE, V. (2016). On the Composition of Scientific Abstracts. *Journal of Documentation*, 2016, vol. 72, n^o 4.

- [Atanassova et Rey, 2021] ATANASSOVA, I. et REY, F.-C. (2021). Categorising scientific uncertainty in papers. *In SciNLP 2021, 8 October 2021, 2nd Workshop on Natural Language Processing for Scientific Text*, page <https://scinlp.org>.
- [Atanassova et al., 2018] ATANASSOVA, I., REY, F.-C. et 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.
- [Atkins et al., 1993] ATKINS, B. T., DUVAL, A., MILNE, R. C., COUSIN, P.-H., LEWIS, H. M., SINCLAIR, L. A., BIRKS, R. O. et LAMY, M. N. (1993). *Le Robert & Collins, dictionnaire français-anglais anglais-français Sénior. 3e édition*. HarperCollins Publishers et Dictionnaires Le Robert, Paris.
- [Auger et Roy, 2008] AUGER, A. et ROY, J. (2008). Expression of Uncertainty in Linguistic Data. *11th International Conference on Information Fusion*, 2008, pages 1860–1867.
- [Bachimont, 2007] BACHIMONT, B. (2007). *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Collection Science informatique et SHS. Lavoisier - Hermes science publications, Paris.
- [Bally, 1965] BALLY, C. (1965). *Linguistique générale et linguistique française - Quatrième édition revue et corrigée*. Editions Francke, Berne.
- [Barreau, 2013] BARREAU, H. (2013). *L'épistémologie*. Presses Universitaires de France, Paris.
- [Bastin et al., 2011] BASTIN, L., WILLIAMS, M., GOSLING, J. P., TRUONG, P., CORNFORD, D., HEUVELINK, G. et ACHARD, F. (2011). Web Based Expert Elicitation of Uncertainties in Environmental Model Inputs. *Geophysical Research Abstracts*, 2011, vol. 13, EGU2011-5384-1, EGU General Assembly 2011. Résumé.
- [Bański et al., 2016] BAŃSKI, P., FRICK, E. et WITT, A. (2016). Corpus Query Lingua Franca (CQLF). *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, mai 2016, pages 2804–2809.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Beaulieu *et al.*, 2019] BEAULIEU, J.-P., BIVER, N., BOCCALETTI, A., BROCHIER, C., FRAY, N., GOURIER, D., LEMELLE, L., MUSTIN, C., PASCAL, R., QUIRICO, E., SCHOEPP-COTHENET, B., SELSIS, F., SZOPA, C., VISO, M. et VUITTON, V. (2019). Rapport du « Groupe de travail thématique Exobiologie, exoplanètes et protection planétaire », p. 104-113 in Rapport préliminaire de la prospective 2019 du CNES « Séminaire de prospective scientifique - Le Havre, 8 au 10 octobre 2019 ». Rapport technique, Centre National d'Etudes Spatiales (CNES), Paris.
- [Beckwith, 2018] BECKWITH, A. (2018). C. S. Peirce and Abduction Inference. *JCCC Honors Journal*, automne 2018, vol. 10, n° 1, article n° 2.
- [Benamara *et al.*, 2012] BENAMARA, F., CHARDON, B., MATHIEU, Y., POPESCU, V. et ASHER, N. (2012). How do Negation and Modality Impact on Opinions?
- [Benamara *et al.*, 2017] BENAMARA, F., GROUIN, C., KAROUI, J., MORICEAU, V. et ROBBA, I. (2017). Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du Défi Fouille de Textes DEFT2017. *Actes de l'atelier « Défi Fouille de Textes » (DEFT 2017), 24^e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), Orléans, France – 26-30 juin 2017*, pages 1–12.
- [Benford *et al.*, 2018] BENFORD, D., HALLDORSSON, T., JEGER, M., KNUTSEN, H., MORE, S., NAEGELI, H., NOTEBORN, H., OCKLEFORD, C., RICCI, A., RYCHEN, G., SCHLATTER, J., SILANO, V., SOLECKI, R., TURCK, D., YOUNES, M., CRAIG, P., HART, A., VON GOETZ, N., KOUTSOUMANIS, K., MORTENSEN, A., OSSENDORP, B., MARTINO, L., MERTEN, C., MOSBACH-SCHULZ, O. et HARDY, A. (2018). Guidance on Uncertainty Analysis in Scientific Assessments. *EFSA Journal*, janvier 2018, vol. 16, n° 1, e05123, 39 p.
- [Benveniste, 1958] BENVENISTE, E. (1958). De la subjectivité dans le langage. *Journal de psychologie* n° 55, 1958. Article repris dans Benveniste E. (1966), *Problèmes de linguistique générale ; 5^e partie : L'homme dans la langue, chapitre XXI*, pages 258–266. Gallimard, Paris.
- [Bernhard et Ligozat, 2011] BERNHARD, D. et LIGOZAT, A.-L. (2011). Analyse automatique de la modalité et du niveau de certitude : application au domaine médical. *TALN 2011, Montpellier, 27 juin – 1er juillet 2011*.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Bhati, 2019] BHATI, R. G. (2019). A Survey on Sentiment Analysis Algorithms and Datasets. *Review of Computer Engineering Research*, juin 2019 (publié en décembre 2019), vol. 6, n° 2, pages 84–91.
- [BIPM, 2006] BIPM (2006). Le système international d’unités (SI) - 8^e édition. Rapport technique, Bureau international des poids et mesures (BIPM), Sèvres, France.
- [Blasch *et al.*, 2013] BLASCH, E., LASKEY, K. B., JOUSSELME, A.-L., DRAGOS, V., COSTA, P. C. G. et DEZERT, J. (2013). URREF Reliability versus Credibility in Information Fusion (STANAG 2511). *Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, July 9-12, 2013*, pages 1600–1607.
- [Blum, 1975] BLUM, F. (1975). Standards Update: ANSI Committee Z39. *The Indexer*, avril 1975, vol. 9, n° 3, pages 113–115.
- [Boukacem-Zeghmouri *et al.*, 2018] BOUKACEM-ZEGHMOURI, C., DILLAERTS, H., LAFOUGE, T., BADOR, P. et SAUER-AVARGUES, A. (2018). French publishing attitudes in the open access era: The case of mathematics, biology, and computer science. *Learned Publishing*.
- [Boukacem-Zeghmouri et Rodríguez Bravo, 2019] BOUKACEM-ZEGHMOURI, C. et RODRÍGUEZ BRAVO, B. (2019). Présentation du dossier 2019. Une information scientifique, entre évaluation et médiatisation. *Les enjeux de l’information et de la communication*, n° 20/2, pages 5–11.
- [Boukhelifa *et al.*, 2012] BOUKHELIFA, N., BEZERIANOS, A., ISENBERG, T. et FEKETE, J.-D. (2012). Evaluating Sketchy Lines for the Visualization of Qualitative Uncertainty. Rapport technique, Project-Teams AVIZ, Institut national de recherche en sciences et technologies du numérique - INRIA, Orsay, France. RR-7910.
- [Bourgin, 2014] BOURGIN, F. (2014). *Comment quantifier l’incertitude prédictive en modélisation hydrologique ? - Travail exploratoire sur un grand échantillon de bassins versants*. Thèse de doctorat.
- [Brankart, 2014] BRANKART, J.-M. (2014). *Traitement des Incertitudes en Océanographie*. Mémoire d’HDR.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Bratko, 2012] BRATKO, I. (2012). *Prolog Programming for Artificial Intelligence - Fourth edition*. Addison Wesley - Pearson Education, Harlow, Essex, Royaume-Uni.
- [Brown et Shuford, 1973] BROWN, T. et SHUFORD, E. H. (1973). Quantifying uncertainty into numerical probabilities for the reporting of intelligence. Rapport technique, RAND Corporation, Santa Monica, Californie, EUA. Cité par [Auger et Roy, 2008, p. 1865-1866].
- [Bureau de normalisation de l'OTAN, 2018] BUREAU DE NORMALISATION DE L'OTAN (2018). Standards related document « SRD AJMedP-3-1 » Guide to Medical Intelligence Handbook, édition A, version 1, janvier 2018. Rapport technique, Bureau de normalisation (NSO) de l'Organisation du traité de l'Atlantique nord (OTAN), Bruxelles, Belgique.
- [Burrough-Boenisch, 1999] BURROUGH-BOENISCH, J. (1999). International Reading Strategies for IMRD Articles. *Written Communication*, juillet 1999, vol. 16, n° 3, page 296–316.
- [Callon et al., 2001] CALLON, M., LASCOUMES, P. et BARTHE, Y. (2001). *Agir dans un monde incertain - Essai sur la démocratie technique*. collection La couleur des idées. Le Seuil, Paris.
- [Capet et Revault d'Allonnes, 2013] CAPET, P. et REVAULT D'ALLONNES, A. (2013). La cotation dans le domaine militaire : doctrines, pratiques et insuffisances. *Chapitre 4 in CAPET Philippe et DELAVALLADE Thomas (2013), « L'évaluation de l'information Confiance et défiance », éditions Lavoisier, Paris.*
- [Cardey, 1987] CARDEY, S. (1987). *Traitement algorithmique de la grammaire normative du français pour une utilisation automatique et didactique*. Thèse de doctorat.
- [Cardey, 2003] CARDEY, S. (2003). Modélisation, systémique, traductibilité. *BULAG n° 28. Presses Universitaires de de Franche-Comté, Besançon.*
- [Cardey, 2013] CARDEY, S. (2013). *Modelling Language*. John Benjamins Publishing Company, Amsterdam, Pays-Bas.
- [Cardey et al., 2010] CARDEY, S., ANANTALAPOCHAI, R., BEDDAR, M., CORNALLY, T., DEVITRE, D., GREENFIELD, P., JIN, G., MIKATI, Z., RENAHY, J.,

RÉFÉRENCES BIBLIOGRAPHIQUES

- KAMPEERA, W., MELIAN, C., SPAGGIARI, L. et VUITTON, D. (2010). Le projet LiSe, « Linguistique, normes, traitement automatique des langues et sécurité » : du data et sense mining aux langues contrôlées. *In Actes du Workshop Interdisciplinaire sur la Sécurité Globale (WISG)*, Université de Technologie de Troyes.
- [Cardey *et al.*, 2006] CARDEY, S., CHAN, R. et GREENFIELD, P. (2006). The Development of a Multilingual Collocation Dictionary. *In Proceedings of COLING ACL 2006: MLRI 2006 workshop*, pages 32–40, Sydney, Australie.
- [Cardey *et al.*, 1997] CARDEY, S., EL HAROUCHY, Z. et GREENFIELD, P. (1997). La forme des mots nous renseigne-t-elle sur leur nature ? *In Actes des 5^{èmes} journées scientifiques, Réseau Lexicologie, Terminologie, Traduction. La mémoire des mots. Collection « actualité scientifique » de l'AUPELF-UREF*, Tunis.
- [Cardey et Greenfield, 2003] CARDEY, S. et GREENFIELD, P. (2003). Disambiguating and Tagging Using Systemic Grammar. *In Actes du 8^{ème} International Symposium on Social Communication*, pages 559–564, Santiago de Cuba, Cuba.
- [Cardey et Greenfield, 2008] CARDEY, S. et GREENFIELD, P. (2008). Micro-systemic Linguistic Analysis and Software Engineering: a synthesis. *Revue Maghrébine des Langues (RML)*, 2008, vol. 6, n° 1, pages 5–24.
- [Catellin, 2004] CATELLIN, S. (2004). L'abduction : une pratique de la découverte scientifique et littéraire. *Hermès, La Revue*, n° 39 - 2004/2, pages 179–185.
- [Catellin, 2012] CATELLIN, S. (2012). Sérendipité et réflexivité. *Alliage*, juillet 2012, n° 70.
- [Cellucci, 1998] CELLUCCI, C. (1998). The scope of logic: deduction, abduction, analogy. *Theoria*, août 1998, vol. 64, n° 2-3, pages 217–242.
- [Cessac et Treherne, 1967] CESSAC, J. et TREHERNE, G. (1967). *Incertitude des mesures et calculs approchés*. Chapitre in *Physique - Classe Terminale D*. Fernand Nathan, Paris. Cité par [Robert-Schwartz et Treiner, 2002].
- [Chafe, 1986] CHAFE, W. (1986). Evidentiality in English Conversation and Academic Writing. *Chafe Wallace et Nichols Johanna (éd.), Evidentiality :*

RÉFÉRENCES BIBLIOGRAPHIQUES

- the linguistic coding of epistemology, 1986, vol. XX in Advances in Discourse Processes (Freedle Roy O. (éd.)), pages 261–272. Ablex Publishing Corporation, Norwood, New Jersey, EUA.*
- [Chen et Song, 2017] CHEN, C. et SONG, M. (2017). *Representing Scientific Knowledge - The Role of Uncertainty*. Springer International Publishing, Cham, Suisse.
- [Chen et al., 2018] CHEN, C., SONG, M. et HEO, G. H. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *Journal of Informetrics*, février 2018, vol. 12, n° 1, pages 158–180.
- [Ciucci et al., 2015] CIUCCI, D., DUBOIS, D. et PRADE, H. (2015). Structures of Opposition in Fuzzy Rough Sets. *Articles written on the occasion of the 50th anniversary of fuzzy set theory, Rapport Interne IRIT RR-2015-11-FR, décembre 2015*, pages 119–144. Institut de Recherche en Informatique de Toulouse - IRIT, Toulouse, France.
- [CoFAT, 2001] CoFAT (2001). *Manuel du sous-officier - Traité toutes armes n° 150 (TTA 150), titre VI - Renseignement, édition 2001*. Commandement de la Formation de l'Armée de Terre - CoFAT, Paris.
- [Cooke et Noortwijk (van), 1999] COOKE, R. M. et NOORTWIJK (VAN), J. M. (1999). Generic Graphics for Uncertainty and Sensitivity Analysis. « *Safety and Reliability* » (éd. Schuëller G.I. et Kafka P.), *Proceedings of ESREL '99 - The Tenth European Conference on Safety and Reliability, Munich-Garching, Germany, 1999*, pages 1187–1192. Balkema, Rotterdam, Pays-Bas.
- [Correia, 2017] CORREIA, M. (2017). Aristotle's Squares of Opposition. *South American Journal of Logic*, 2017, vol. 3, n° 1.
- [Couillaud et al., 2005] COUILLAUD, A., GARNIER, A., DEBONNE, E., FRANCOPOULO, G., GIRE, F., GUILLEMIN LANE, S., LOUPY (DE), C., MAZIÈRES, G., MENON, B. et Sézille de MAZANCOURT, H. (2005). Le traitement automatique des langues dans les industries de l'information - Livre blanc. Rapport technique, Groupe de Réflexion sur les Industries de l'Information et les Industries de la Langue (GRIIL), Paris.
- [Courtial, 1990] COURTIAL, J.-P. (1990). *Introduction à la scientométrie - De la bibliométrie à la veille technologique*. Anthropos - Economica, 49 rue Héricart, 75015 Paris.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Cruz Díaz et Maña López, 2019] CRUZ DÍAZ, N. P. et MAÑA LÓPEZ, M. J. (2019). *Negation and Speculation Detection*. John Benjamins Publishing company, Amsterdam, Pays-Bas.
- [Crystal, 2008] CRYSTAL, D. (2008). *A Dictionary of Linguistics and Phonetics*.
- [Dalloux, 2017] DALLOUX, C. (2017). Fouille de texte et extraction d'informations dans les données cliniques. IRISA - CNRS, Rennes, France. Résumé de travaux de thèse.
- [Dautun et al., 2006] DAUTUN, C., TIXIER, J., CHAPELAIN, J., FONTAINE, F. et DUSSERRE, G. (2006). Le traitement de l'incertitude en gestion de crise : mise en place d'une veille stratégique du territoire. *Colloque de maîtrise des risques et sûreté de fonctionnement « Risques et performances »*, Oct 2006, Lille, France, page 6.
- [Dauzat et al., 1964] DAUZAT, a., DUBOIS, J. et MITTERAND, H. (1964). *Nouveau dictionnaire étymologique et historique*. Librairie Larousse, Paris.
- [Davis et al., 2016] DAVIS, P. K., PERRY, W. L., HOLLYWOOD, J. S. et MANHEIM, D. (2016). Uncertainty-sensitive heterogeneous information fusion - assessing threat with soft, uncertain, and conflicting evidence. Rapport technique, RAND Corporation, Santa Monica, Californie, EUA.
- [Deccache, 1999] DECCACHE, A. (1999). Comment ont évolué les concepts et pratiques de prévention des maladies et de promotion de la santé ? *Santé conjugulée*, octobre 1999, n° 10, pages 30–33.
- [Degoulet et Fieschi, 1997] DEGOULET, P. et FIESCHI, M. (1997). *Introduction to Clinical Informatics*. Springer-Verlag, New York, EUA.
- [Dendale et Tasmowski, 2001] DENDALE, P. et TASMOWSKI, L. (2001). Introduction: Evidentiality and related notions. *Journal of Pragmatics*, mars 2001, vol. 33, n° 3, pages 339–348.
- [Desclés, 1997] DESCLÉS, J.-P. (1997). Systèmes d'exploration contextuelle. *Claude Guimier (éd.)(1997), « Co-texte et calcul du sens » : actes de la table ronde tenue à Caen les 2 et 3 février 1996* », pages 215–232.
- [Desclés, 2006] DESCLÉS, J.-P. (2006). Contextual Exploration Processing for Discourse Automatic Annotations of Texts. *FLAIRS 2006, Melbourne, Floride, 11-13 mai, Invited Speakers*, pages 281–284. AAAI Press.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Desclés, 2009] DESCLÉS, J.-P. (2009). Prise en charge, engagement et désengagement. *Langue française*, février 2009, vol. 162, n° 2 « *La notion de prise en charge en linguistique* », pages 29–53.
- [Desclés et al., 1997] DESCLÉS, J.-P., CARTIER, E., JACKIEWICZ, A. et MINEL, J.-L. (1997). Textual Processing and Contextual Exploration Method. *Context'97, Rio de Janeiro, février 1997*, pages 189–197.
- [Desclés et Guentchéva, 2000] DESCLÉS, J.-P. et GUENTCHÉVA, Z. (2000). Enonciateur, locuteur, médiateur. *Erikson Philippe et Monod-Becquelin Aurore (éd.) (2000), « Les rituels du dialogue », Promenades ethnolinguistiques en terres amérindiennes*, pages 79–112. Société d'ethnologie, Paris X, Nanterre. Cités par [Kronning, 2003, p. 145].
- [Desclés et Jackiewicz, 2006] DESCLÉS, J.-P. et JACKIEWICZ, A. (2006). Abduction et prise en charge énonciative de la causalité. *Linx n° 54*.
- [Devillard et Marco, 1993] DEVILLARD, J. et MARCO, L. (1993). *Ecrire et publier dans une revue scientifique*. Les Editions d'Organisation, Paris, France. Collection Method'Sup.
- [Diab et al., 2009] DIAB, M. T., LEVIN, L., MITAMURA, T., RAMBOW, O., PRABHAKARAN, V. et GUO, W. (2009). Committed Belief Annotation and Tagging. *ACL-IJCNLP 2009 - LAW 2009: Proceedings of the Third Linguistic Annotation Workshop, Suntec, Singapore - August 06-07, 2009*, pages 68–73.
- [Dias et al., 2018] DIAS, L. C., MORTON, A. et QUIGLEY, J. (2018). Elicitation: State of the Art and Science. *Dias Luis C., Morton Alec et Quigley John (eds.), 18 novembre 2017, Elicitation - The Science and Art of Structuring Judgement, International Series in Operations Research & Management Science*, vol. 261, chap. 1, pages 1–14. Springer International Publishing, Cham, Suisse.
- [Djebri et al., 2019a] DJEBRI, A. E. A., TETTAMANZI, A. G. B. et GANDON, F. (2019a). Linking and Negotiating Uncertainty Theories Over Linked Data. *WWW '19: Companion Proceedings of The 2019 World Wide Web Conference*, p. 859–865. *LDOW/LDDL Workshop of the World Wide Web Conference (WWW 2019), Mai 2019, San Francisco, EUA*.

- [Djebri *et al.*, 2019b] DJEBRI, A. E. A., TETTAMANZI, A. G. B. et GANDON, F. (2019b). Publishing Uncertainty on the Semantic Web: Blurring the LOD bubbles. *Graph-Based Representation and Reasoning, Juillet 2019*, p. 42-56. Endres D., Alam M., Şotropa D. (ed.). 24th International Conference on Conceptual Structures (ICCS 2019), Marburg, Allemagne. *Lecture Notes in Computer Science*, vol 11530. Springer International Publishing, Cham, Suisse.
- [Djebri *et al.*, 2020] DJEBRI, A. E. A., TETTAMANZI, A. G. B. et GANDON, F. (2020). Task-Oriented Uncertainty Evaluation for Linked Data Based on Graph Interlinks. *Knowledge Engineering and Knowledge Management, Septembre 2020*, p. 204-215. Keet C.M., Dumontier M. (ed). 22nd International Conference on Knowledge Engineering and Knowledge Management (EKAW 2020), Bozen-Bolzano, Italie. *Lecture Notes in Computer Science*, vol 12387. Springer International Publishing, Cham, Suisse.
- [Djioua *et al.*, 2006] DJIOUA, B., FLORES, J. J. G., BLAIS, A., DESCLÉS, J.-P., GUIBERT, G., JACKIEWICZ, A., LE PRIOL, F., NAIT-BAHA, L. et SAUZAY, B. (2006). EXCOM: An Automatic Annotation Engine for Semantic Information. In *FLAIRS Conference*, pages 285–290.
- [Dossantos-Uzarralde *et al.*, 2017] DOSSANTOS-UZARRALDE, P., CHAU, P., VAYATIS, N. et VARET, S. (2017). Quantification d’incertitudes par hiérarchisation de modèles - Application aux sections efficaces neutroniques. *Revue CHOCS - Revue scientifique et technique de la direction des applications militaires*, 2017, n° 48 « Maîtrise des incertitudes », pages 84–93. CEA - DAM, Institut supérieur des études nucléaires de défense (ISENDé), Bruyères-le-Châtel, France.
- [Dougherty, 2016] DOUGHERTY, E. R. (2016). *The evolution of scientific knowledge : from certainty to uncertainty*. SPIE Press, Bellingham, Washington, EUA.
- [Dragos, 2013] DRAGOS, V. (2013). An ontological analysis of uncertainty in soft data. *Proceedings of the 16th International Conference on Information Fusion, FUSION 2013, Istanbul, Turkey, July 9-12, 2013*, pages 1566–1573.
- [Dubois *et al.*, 2007] DUBOIS, J., GIACOMO, M., GUESPIN, L., MARCELLESI, C., MARCELLESI, J.-B. et MEVEL, J.-P. (2007). *Grand dictionnaire - Linguistique & Sciences du langage*. Editions Larousse, Paris.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Dubois *et al.*, 2002] DUBOIS, J., GIACOMO, M., GUESPIN, L., MARCELLESI, C., MARCELLESI, J.-B. et MÉVEL, J.-P. (2002). *Dictionnaire de linguistique*. Editions Larousse, Paris.
- [Dumez, 2012] DUMEZ, H. (2012). Qu'est-ce que l'abduction, et en quoi peut-elle avoir un rapport avec la recherche qualitative ? *Le Libellio d'Aegis*, vol. 8, n° 3, pages 3–9.
- [Dutreuil, 2018] DUTREUIL, S. (2018). James Lovelock's Gaia Hypothesis: "A New Look at Life on Earth" . . . for the Life and the Earth Sciences. *Dreamers, Visionaries, and Revolutionaries in the Life Sciences*, 2018, pages 272–287.
- [Feyel, 2018] FEYEL, G. (2018). Revues savantes : les collèges invisibles. *L'Europe - Encyclopédie historique*, pages 1934–1936. Actes Sud, Arles.
- [Fintel (von), 2006] FINTEL (VON), K. (2006). Modality and Language. *Encyclopedia of Philosophy, Second Edition (2006)*, vol. 10, pages 20–27. Macmillan Reference USA - Thomson Gale, Farmington Hills, Michigan, EUA.
- [Fraisie et Paroubek, 2015] FRAISSE, A. et PAROUBEK, P. (2015). Un modèle générique pour une annotation à granularité fine des opinions et des sentiments. *Atelier : Etudier le Web Politique : Regards croisés, May 2015, Lyon, France*.
- [Friedman *et al.*, 1998] FRIEDMAN, C., HRIPCSAK, G. et SHABLINSKY, I. (1998). An Evaluation of Natural Language Processing Methodologies. *AMIA Annual Symposium Proceedings*, 1998, pages 855–859.
- [Fuchs, 1996] FUCHS, C. (1996). *Les ambiguïtés du français*. Collection L'Essentiel français. Ophrys, Paris.
- [Funtowicz et Ravetz, 1990] FUNTOWICZ, S. O. et RAVETZ, J. R. (1990). *Uncertainty and Quality in Science for Policy*. Springer Netherlands, Dordrecht, Pays-Bas.
- [Fusco, 2013] FUSCO, G. (2013). L'incertitude dans les sciences géographiques : spécificités, schémas d'appréhension. *1^e Table Ronde « L'incertitude : regards pluridisciplinaires », Nice, 19 juin 2013, MSHS Sud-Est, Saint-Jean d'Angély 3, Projet « Faire Science avec l'Incertain », Axe 4 : « Territoires, systèmes techniques et usages sociaux ».*

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Fusco *et al.*, 2015] FUSCO, G., BERTONCELLO, F., CANDAU, J., EMMELM, K., HUET, T., LONGHI, C., POINAT, S., PRIMON, J.-I. et RINAUDO, C. (2015). Faire science avec l'incertitude : réflexions sur la production des connaissances en Sciences Humaines et Sociales.
- [Gandon *et al.*, 2012] GANDON, F., FARON-ZUCKER, C. et CORBY, O. (2012). *Le web sémantique - Comment lier les données et les schémas sur le web ?* Dunod, Malakoff, France.
- [Gardies, 1979] GARDIES, J.-L. (1979). *Essai sur la logique des modalités*. Presses Universitaires de France - PUF, Paris, France.
- [Gentilhomme, 1979] GENTILHOMME, Y. (1979). Microsystèmes linguistiques et langagiers - Fonctions heuristiques et didactiques - Introduction méthodologique. *Travaux du Centre de Recherches sémiologiques (CdRS), mars 1979, vol. n° 34 « Langue et discours (II) – Colloque Besançon-Neuchâtel. Neuchâtel 2-4 octobre 1978 », pages 1–31.*
- [Gentilhomme, 1985] GENTILHOMME, Y. (1985). *Essai d'approche microsystemique. Théorie et pratique - Application dans le domaine des sciences du langage*. Peter Lang, Berne, Suisse. Collection Sciences pour la communication.
- [Gentner, 2002] GENTNER, D. (2002). Analogy in Scientific Discovery: The Case of Johannes Kepler. *Model-Based Reasoning, 2002*, pages 21–39. Springer, Boston, Massachusetts, EUA.
- [Ginsparg, 2009] GINSPARG, P. (2009). Text in a Data-centric World. *Chapitre in [Hey et al., 2009, p. 185-191]*.
- [Giuliani, 2016] GIULIANI, G. (2016). Partage de données environnementales. *Techniques de l'ingénieur - Techniques d'analyse, février 2016, vol. TIP630WEB, réf. P4250 v1.*
- [Gosling, 2018] GOSLING, J. P. (2018). SHELF: The Sheffield Elicitation Framework. *Dias Luis C., Morton Alec et Quigley John (eds.), 18 novembre 2017, Elicitation - The Science and Art of Structuring Judgement, International Series in Operations Research & Management Science, vol. 261, chap. 4, pages 61–93.* Springer International Publishing, Cham, Suisse.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Goujon, 2009] GOUJON, B. (2009). Uncertainty Detection for Information Extraction. *International Conference RANLP 2009*, pages 118–122.
- [Grabar et al., 2018] GRABAR, N., CLAVEAU, V. et DALLOUX, C. (2018). CAS: French Corpus with Clinical Cases. *LOUHI 2018 - The Ninth International Workshop on Health Text Mining and Information Analysis, octobre 2018, Bruxelles, Belgique*, pages 1–7. Association for Computational Linguistics (ACL), Stroudsburg, Pennsylvanie, EUA.
- [Grossmann et Tutin, 2010] GROSSMANN, F. et TUTIN, A. (2010). Evidential markers in French scientific writing: the case of the French verb *voir*. *Diwald Gabriele et Smirnova Elena (ed.)(2010), Linguistic Realization of Evidentiality in European Languages ; Bosson Georg, Comrie Bernard et Matras Yaron (ed.), Empirical Approaches to Language Typology, 2010, n° 49*, pages 279–307. De Gruyter Mouton, Berlin, Allemagne.
- [Halliday, 1995a] HALLIDAY, M. A. K. (1995a). A recent view of "missteps" in linguistic theory. *Functions of Language, vol. 2, n° 2, Jan 1995*, pages 249–267.
- [Halliday, 1995b] HALLIDAY, M. A. K. (1995b). On Language in Relation to Fuzzy Logic and Intelligent Computing. Publié in [[Halliday, 2005](#)], p. 196–212.
- [Halliday, 2005] HALLIDAY, M. A. K. (2005). *Computational and Quantitative Studies*. Continuum, Londres, Royaume-Uni.
- [Halliday et Matthiessen, 2014] HALLIDAY, M. A. K. et MATTHIESSEN, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar*. Routledge, Abingdon, Royaume-Uni, 4^e édition.
- [Harman, 1965] HARMAN, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review, janvier 1965, vol. 74, n° 1*, pages 88–95.
- [Hart et al., 2019] HART, A., MAXIM, L., SIEGRIST, M., GOETZ (VON), N., CRUZ (DA), C., MERTEN, C., MOSBACH-SCHULZ, O., LAHANIAS, M., SMITH, A. et HARDY, A. (2019). Guidance on Communication of Uncertainty in Scientific Assessments. *EFSA Journal, janvier 2019, vol. 17, n° 1, e05520*, 73 p.

RÉFÉRENCES BIBLIOGRAPHIQUES

[Haute Autorité de Santé, 2007] HAUTE AUTORITÉ DE SANTÉ (2007). Evaluation de la qualité des sites e-santé et de la qualité de l'information de santé diffusée sur Internet (Revue de la littérature des outils d'évaluation). Rapport technique, Service qualité de l'information médicale - Haute autorité de santé (HAS), Saint-Denis La Plaine, France.

[Haute Autorité de Santé, 2013] HAUTE AUTORITÉ DE SANTÉ (2013). Niveau de preuve et gradation des recommandations de bonne pratique. Rapport technique, Service des bonnes pratiques professionnelles - Haute autorité de santé (HAS), Saint-Denis La Plaine, France.

[Herslund, 1989] HERSLUND, M. (1989). Modality: A presentation. *On modality. Papers from meetings and discussions in the linguistic circle of Copenhagen, Herslund, M. (ed.)(1989). København : Akademisk forlag, 7-16.* Cité par [Vold, 2008, p. 42].

[Hey et al., 2009] HEY, T., TANSLEY, S. et TOLLE, K. (2009). *The Fourth Paradigm - Data-Intensive Scientific Discovery.* Microsoft Research, Redmond, Washington, EUA.

[Higuera (de la), 2019] HIGUERA (DE LA), C. (2019). A report about Education, Training Teachers and Learning Artificial Intelligence: Overview of key issues. Rapport technique, Knowledge 4 All Foundation Ltd (K4A), Redhill (Surrey, Angleterre), Royaume-Uni.

[Hirsch, 2005] HIRSCH, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS, 15 novembre 2005, vol. 102, n° 46*, pages 16569–16572.

[Hirsch, 2020] HIRSCH, J. E. (2020). Superconductivity, What the H? The Emperor Has No Clothes. *Physics and Society, janvier 2020, vol. 49, n° 1*, pages 4–9.

[Holmes, 1982] HOLMES, J. (1982). Expressing Doubt and Certainty in English. *RELC Journal, décembre 1982, vol. 13, n° 2*, page 9–28.

[Holmes, 1984] HOLMES, J. (1984). Modifying illocutionary force. *Journal of Pragmatics, juin 1984, vol. 8, n° 3*, pages 345–365.

- [Hora, 1996] HORA, S. C. (1996). Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Treatment of Aleatory and Epistemic Uncertainty, Helton J.C. et Burmaster D.E. (ed.), Reliability Engineering & System Safety, novembre-décembre 1996, vol. 54, n° 2-3*, pages 217–223. Elsevier Science Limited, Oxford, Royaume-Uni.
- [Hora, 2016] HORA, S. C. (2016). Probability Elicitation. *Hájek Alan et Hitchcock Christopher (ed.), The Oxford Handbook of Probability and Philosophy, septembre 2016, partie V - Probabilistic Judgment and Its Applications, chap. 24*. Oxford University Press, Oxford, Royaume-Uni.
- [Hyland, 1996a] HYLAND, K. (1996a). Talking to the Academy: Forms of Hedging in Science Research Articles. *Written Communication, 1996, vol. 13, n° 2*, pages 251–281.
- [Hyland, 1996b] HYLAND, K. (1996b). Writing Without Conviction. Hedging in Science Research Articles. *Applied Linguistics, vol. 17 n° 4*, pages 433–454.
- [IPCC, 2005] IPCC (2005). Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties. Rapport technique, Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- [IPCC, 2018] IPCC (2018). Global Warming of 1.5 °C - Special Report. Rapport technique, Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- [Irwin et Mandel, 2019a] IRWIN, D. et MANDEL, D. R. (2019a). Improving information evaluation for intelligence production. *Intelligence and National Security, 2019, vol. 4, n° 4*, pages 503–525.
- [Irwin et Mandel, 2019b] IRWIN, D. et MANDEL, D. R. (2019b). Standards for Evaluating Source Reliability and Information Credibility in Intelligence Production. *D. R. Mandel (éd.), Assessment and Communication of Uncertainty in Intelligence to Support Decision Making: Final Report of Research Task Group S. Organisation du traité de l'Atlantique nord (OTAN), Bruxelles, Belgique*.
- [Jackiewicz et al., 2010] JACKIEWICZ, A., HUNSTON, S. et EL-BÈZE, M. (2010). Opinions, sentiments et jugements d'évaluation - Préface. *Traitement Automatique des Langues, 2010, vol. 51, n° 3*, pages 1–11.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Jacques, 2017a] JACQUES, M.-P. (2017a). La dynamique du texte, corpus, outils, analyses - Tome 1. Mémoire d'HDR.
- [Jacques, 2017b] JACQUES, M.-P. (2017b). La structuration textuelle en discours scientifique : comparaison oral / écrit. *Chimera. Romance Corpora and Linguistic Studies*, 2017, n° 1, pages 89–115.
- [JCGM, 2008] JCGM (2008). Evaluation des données de mesure - Guide pour l'expression de l'incertitude de mesure (alias « GUM »), version française. Rapport technique, Groupe de travail 1 du Comité commun pour les guides en métrologie (Working Group 1 of the Joint Committee for Guides in Metrology - JCGM/WG 1) du Bureau International des Poids et Mesures (BIPM), Sèvres, France.
- [JCGM, 2012] JCGM (2012). Vocabulaire international de métrologie - Concepts fondamentaux et généraux et termes associés (VIM), version française. Rapport technique 3^e édition, Comité commun pour les guides en métrologie (Joint Committee for Guides in Metrology - JCGM) du Bureau International des Poids et Mesures (BIPM), Sèvres, France.
- [Jean et al., 2017] JEAN, P.-A., HARISPE, S., BELLOT, P., RANWEZ, S. et MONTMAIN, J. (2017). Gérer l'incertitude lors de l'extraction de relations et lors de l'inférence de nouvelles connaissances. *Conférence en Recherche d'Informations et Applications - CORIA 2017*, mars 2017, Marseille, France.
- [Jones, 2014] JONES, M. D. (2014). Communicating Climate Change: Are Stories Better than "Just the Facts"? *Policy Studies Journal*, 2014, vol. 42, n° 4, pages 644–673.
- [Kant, 1781] KANT, E. (1781). *Critique de la raison pure (Kritik der reinen Vernunft)*. Librairie Félix Alcan, Paris. Traduction française par Tremesaygues A. et Pacaud B., 1905.
- [Kerdjoudj, 2015] KERDJOU DJ, F. (2015). *Gestion de l'incertitude dans le cadre d'une extraction de connaissances à partir de textes*. Thèse de doctorat.
- [Klyueva et al., 2018] KLYUEVA, N., VERNEROVÁ, A. et QASEMIZADEH, B. (2018). Querying Multi-word Expressions Annotation with CQL. *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16)*, Prague, Czech Republic, 23–24 janvier 2018, page 73–79.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Knight, 1921] KNIGHT, F. H. (1921). *Risk, uncertainty and profit*. Houghton Mifflin Company, EUA.
- [Kratzer, 2012] KRATZER, A. (2012). *Modals and Conditionals*. Oxford University Press, Oxford, Royaume-Uni.
- [Kronning, 2003] KRONNING, H. (2003). Modalité et évidentialité. *Birkelund Merete, Boysen Gerhard & Kjærsgaard Poul Søren (éd.)*(2003), « *Aspects de la modalité* », pages 131–152. Editions Max Niemeyer (groupe De Gruyter), Tübingen, Allemagne.
- [Kuhn, 1970] KUHN, T. S. (1962-1970). *The Structure of Scientific Revolutions - Second Edition, Enlarged*, volume 2 de *International Encyclopedia of Unified Science*. The University of Chicago Press, Chicago, EUA.
- [Kupka et Rusnok, 2019] KUPKA, J. et RUSNOK, P. (2019). Regression Ensemble with Linguistic Descriptions. *Advances in Intelligent Systems and Computing, 2019, vol. 832 - Uncertainty Modelling in Data Science*, pages 141–148.
- [Lakoff, 1973] LAKOFF, G. (1973). Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic*, vol. 2, pages 458–508.
- [Latour et Woolgar, 1986] LATOUR, B. et WOOLGAR, S. (1979-1986). *Laboratory Life - The Construction of Scientific Facts*. Princeton University Press, Princeton, New Jersey, EUA.
- [Le Querler, 1996] LE QUERLER, N. (1996). *Typologie des modalités*. Presses Universitaires de Caen, Caen.
- [Lefeuvre et al., 2014] LEFEUVRE, A., ANTOINE, J.-Y., SAVARY, A., SCHANG, E., ABOUDA, L., MAUREL, D. et ESHKOL, I. (2014). Annotation de la temporalité en corpus : contribution à l’amélioration de la norme TimeML. *21^e Conférence Traitement Automatique du Langage Naturel (TALN 2014)*, juillet 2014, Marseille, France.
- [Lilensten, 2018] LILENSTEN, J. (2018). *Les sens du mot science*. EDP Sciences, Les Ulis, France.
- [Lim et al., 2004] LIM, B., SPANGER-SIEGFRIED, E., BURTON, I., MALONE, E. L. et HUQ, S. (2004). *Adaptation Policy Frameworks for Climate Change*:

RÉFÉRENCES BIBLIOGRAPHIQUES

- Developing Strategies, Policies and Measures*. Press Syndicate of The University of Cambridge, Cambridge, Royaume-Uni.
- [Lindsay, 2011] LINDSAY, D. (2011). *Scientific Writing = Thinking in Words*. CSIRO Publishing, Collingwood, Australie.
- [Lipshitz et Strauss, 1997] LIPSHITZ, R. et STRAUSS, O. (1997). Coping with Uncertainty: A Naturalistic Decision-Making Analysis. *Organizational Behavior and Human Decision Processes*, février 1997, vol. 69, n° 2, pages 149–163.
- [Loucks et Beek (van), 2017] LOUCKS, D. P. et BEEK (VAN), E. (2017). *Water Resource Systems Planning and Management - An Introduction to Methods, Models, and Applications*. Springer Nature, Deltares and UNESCO-IHE, Springer International Publishing AG, Gewerbestrasse 11, 6330 Cham, Suisse.
- [Lyons, 1977] LYONS, J. (1977). *Semantics*, volume 2. Cambridge University Press, Cambridge, Royaume-Uni.
- [López Ferrero, 2005] LÓPEZ FERRERO, C. (2005). Funciones retóricas en la comunicación académica: Formas léxicas de modalidad y evidencialidad. *Signo y Seña*, avril 2005, n° 14, pages 115–139.
- [Makkaoui et al., 2012] MAKKAOUI, O., DESCLÉS, J. et DESCLÉS, J.-P. (2012). Evaluation and performance improvement of the BioExcom system for the automatic detection of speculation in biomedical texts. *In 3rd Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM)*.
- [Malhotra et al., 2013] MALHOTRA, A., YOUNESI, E., GURULINGAPPA, H. et HOFMANN-APITIUS, M. (2013). ‘HypothesisFinder:’ A Strategy for the Detection of Speculative Statements in Scientific Text. *PLOS Computational Biology*, juillet 2013, vol. 9, n° 7, pages 1–10.
- [Masson-Delmotte, 2020] MASSON-DELMOTTE, V. (2020). Réchauffement climatique : état des connaissances scientifiques, enjeux, risques et options d’action. *Géoscience*, 2020, vol. 352, n° 4-5, pages 251–277.
- [Mastrandrea et al., 2010] 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. et

RÉFÉRENCES BIBLIOGRAPHIQUES

- ZWIERS, F. W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Rapport technique, Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- [Mastrandrea *et al.*, 2011] MASTRANDREA, M. D., MACH, K. J., PLATTNER, G.-K., EDENHOFER, O., STOCKER, T. F., FIELD, C. B., EBI, K. L. et MATSCHOSS, P. R. (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Climatic Change*, 2011, vol. 108, n° 4, pages 675–691. Intergovernmental Panel on Climate Change (IPCC), Genève, Suisse.
- [Masó, 2017] MASÓ, Joan et Zabala, A. (2017). Testbed-12 Imagery Quality and Accuracy Engineering Report. Rapport technique, Joan Masó et Alaitz Zabala (éd.), Open Geospatial Consortium (OGC), Arlington, EUA.
- [Meunier, 1981] MEUNIER, A. (1981). Grammaires du français et modalités - Matériaux pour l'histoire d'une nébuleuse. *DRLAV*, 25, Université de Paris VIII, Paris. Cité par [Le Querler, 1996, p. 49].
- [Mikolov *et al.*, 2013] MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S. et DEAN, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in neural information processing systems*, vol. 26 - Neural Information Processing Systems Conference 2013, October 2013, pages 3111–3119.
- [Morante et Sporleder, 2012] MORANTE, R. et SPORLEDER, C. (2012). Modality and Negation: An Introduction to the Special Issue. *Computational Linguistics*, 2012, Vol. 38, n° 2.
- [Morin, 1980] MORIN, E. (1980). *La Vie de la Vie - La Méthode*, tome 2. Editions du Seuil, Paris.
- [Morin, 1999] MORIN, E. (1999). *Les sept savoirs nécessaires à l'éducation du futur*. Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), Paris.
- [Mosteller et Youtz, 1990] MOSTELLER, F. et YOUTZ, C. (1990). Quantifying Probabilistic Expressions. *Statistical Science*, 1990, vol. 5 n° 1, pages 2–12.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Motet, 2010] MOTET, G. (2010). Le concept de risque et son évolution. *Les Annales des Mines - Responsabilité & Environnement*, janvier 2010, n° 57 : « Faire face à l'incertitude ».
- [Mushin, 2001] MUSHIN, I. (2001). *Evidentiality and Epistemological Stance - Narrative retelling*. John Benjamins Publishing Company, Amsterdam, Pays-Bas. Jucker Andreas H. (ed.), *Pragmatics & Beyond New Series*, 2001, vol. 87.
- [Mélac, 2014] MÉLAC, E. (2014). *L'évidentialité en anglais - Approche contrastive à partir d'un corpus anglais-tibétain*. Thèse de doctorat.
- [Nabarette et Caniard, 2010] NABARETTE, H. et CANIARD, E. (2010). Sites santé et régulation. *Les Tribunes de la santé*, 2010, vol. 29, n° 4, pages 49–58. Presses de Sciences Po, Paris.
- [Newell et al., 2018] 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. et ZHANG, Z. (2018). The codata 2017 values of h , e , k , and N_A for the revision of the SI. *Metrologia*, 2018, vol. 55, n° 1, pages L13–L16.
- [Novák, 2007] NOVÁK, V. (2007). A Comprehensive Theory of Trichotomous Evaluative Linguistic expressions - Research report n° 71. Rapport technique, Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Ostrava, République tchèque.
- [Nunes et al., 2018] NUNES, J., BARBOSA, M., SILVA, L., GORGÔNIO, K., HYGGO, A. et ANGELO, P. (2018). Issues in the Probability Elicitation Process of Expert-Based Bayesian Networks. *Vizureanu Petrică (ed.), Enhanced Expert Systems*, 5 novembre 2018, chap. 2. IntechOpen, Londres, Royaume-Uni.
- [Nuyts, 2017] NUYTS, J. (2017). Evidentiality reconsidered. *Marín Arrese Juana Isabel, Haßler Gerda and Carretero Marta (ed.) (2017), Evidentiality Revisited*, pages 57–83. John Benjamins Publishing Company, Amsterdam, Pays-Bas.
- [Ollinger, 2014] OLLINGER, S. (2014). *Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français*. Thèse de doctorat.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [OTAN, 2003] OTAN (2003). Accord de normalisation STANAG 2511 - Comptes rendus de renseignement (Standardization Agreement (STANAG) 2511 - Intelligence Reports - Ed: 1). Rapport technique, Organisation du traité de l'Atlantique nord (OTAN), Bruxelles, Belgique. Cité par [Bureau de normalisation de l'OTAN, 2018, Annexe B], [Irwin et Mandel, 2019b, p. 7-1 à 7-4], [Irwin et Mandel, 2019a, § Overview of Current Methods] et [Blasch *et al.*, 2013, § III.B].
- [Peirce, 1958] PEIRCE, C. S. (1866-1958). *The Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, Massachusetts, EUA. Editeurs : Hartshorne Charles et Weiss Paul (1931-1935, vol. n° 1-6), Burks Arthur W. (1958, vol. n° 7-8). Réédition commune de 1994.
- [Pepper et Prytulak, 1974] PEPPER, S. et PRYTULAK, L. S. (1974). Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions. *Journal of Research in Personality*, 1974, vol. 8, pages 95–101.
- [Perkins, 1983] PERKINS, M. (1983). *Modal expressions in English*. Frances Pinter, Londres, Royaume Uni. Cité par [Vold, 2008, p. 42].
- [Peroni, 2014] PERONI, S. (2014). *Semantic Web Technologies and Legal Scholarly Publishing*. Springer, Cham, Suisse.
- [Pochet, 2015] POCHET, B. (2015). *Comprendre et maîtriser la littérature scientifique. Version du 2 août 2018*. Presses agronomiques de Gembloux, Gembloux, Belgique.
- [Poibeau, 2011] POIBEAU, T. (2011). *Traitement automatique du contenu textuel*. Hermès-sciences, Lavoisier, Paris.
- [Polanco, 1995] POLANCO, X. (1995). Aux sources de la scientométrie. « *Les sciences de l'information : bibliométrie, scientométrie, infométrie* » in *Solaris*, 1995, n° 2. Presses Universitaires de Rennes.
- [Polya, 1954] POLYA, G. (1954). *Mathematics and Plausible Reasoning - Vol. I. Induction and Analogy in Mathematics*. Princeton University Press, Princeton, New Jersey, EUA.
- [Polya, 1985] POLYA, G. (1985). *How to Solve It - A New Aspect of Mathematical Method - Second Edition*. Princeton University Press, Princeton, New Jersey, EUA.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Pontille, 2007] PONTILLE, D. (2007). Matérialité des écrits scientifiques et travail de frontières : le cas du format IMRAD. In *Hert Philippe et Paul-Cavallier Marcel (ed.)*(janvier 2007), « *Sciences et frontières : délimitations du savoir, objets et passages* », pages 229–253. Éditions Modulaires Européennes, Fernelmont, Belgique. Collection Echanges.
- [Prabhakaran et al., 2010] PRABHAKARAN, V., RAMBOW, O. et DIAB, M. T. (2010). Automatic Committed Belief Tagging. *Coling 2010, Beijing, August 2010: Poster Volume*, page 1014–1022.
- [Pustejovsky et al., 2003] PUSTEJOVSKY, J., CASTAÑO, J., INGRIA, R., SAURÍ, R., GAIZAUSKAS, R., SETZER, A. et KATZ, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. *Fifth International Workshop on Computational Semantics - IWCS-5, 15-17 janvier 2003, Tilburg University, Pays-Bas*.
- [Quigley et al., 2018] QUIGLEY, J., COLSON, A., ASPINALL, W. et COOKE, R. M. (2018). Elicitation in the Classical Model. *Dias Luis C., Morton Alec et Quigley John (eds.)*, 18 novembre 2017, *Elicitation - The Science and Art of Structuring Judgement, International Series in Operations Research & Management Science*, vol. 261, chap. 2, pages 15–36. Springer International Publishing, Cham, Suisse.
- [Raj et Raman, 2017] RAJ, P. et RAMAN, A. C. (2017). *The Internet of Things - Enabling Technologies, Platforms, and Use Cases*. CRC Press, Taylor & Francis Group, Boca Raton, Floride, EUA.
- [Rendsvig et Symons, 2019] RENDSVIG, R. et SYMONS, J. (2019). Epistemic Logic. *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*.
- [Rey, 2011] REY, F.-C. (2011). Promotion des bonnes pratiques d’internet pour la santé. Rapport technique, Fondation Health On The Net, Genève, Suisse. Rapport de stage Master 2.
- [Rey et al., 2018] REY, F.-C., BERTIN, M. et ATANASSOVA, I. (2018). Une étude de l’incertitude dans les textes scientifiques : vers la construction d’une ontologie. *Terminologie & Ontologie : Théories et Applications, 10 septembre 2020, Actes de la conférence TOTh 2018, Le Bourget du Lac – 7 & 8 juin 2018*, pages 229–242. Presses universitaires Savoie Mont Blanc, Chambéry, France. Collection Terminologica.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Richards et Schmidt, 2010] RICHARDS, J. C. et SCHMIDT, R. W. (2010). *Longman Dictionary of Language Teaching and Applied Linguistics, 4th ed.* Pearson Education Limited, Harlow, Royaume-Uni.
- [Rinaudo, 2015] RINAUDO, C. (2015). Le traitement de l'incertitude dans la relation d'enquête ethnographique en sciences sociales. *Incertitude et connaissances en SHS : production, diffusion, transfert, Jun 2014, Nice, France. 2015.*
- [Ripple *et al.*, 2017] RIPPLE, W. J., WOLF, C., NEWSOME, T. M., GALETTI, M., ALAMGIR, M., CRIST, E., MAHMOUD, M. I., LAURANCE, W. F. et « 15.364 SCIENTIFIQUES SIGNATAIRES ISSUS DE 184 PAYS », ... (2017). World Scientists' Warning to Humanity: A Second Notice. *BioScience*, novembre 2017, vol. 67, n° 12, pages 1026–1028.
- [Robert-Schwartz et Treiner, 2002] ROBERT-SCHWARTZ, C. et TREINER, J. (2002). Incertitude des mesures de grandeur. Rapport technique, Ministère de l'Éducation nationale, France.
- [Roche, 2005] ROCHE, C. (2005). Terminologie et ontologie. *Langages*, 2005, vol. 157, n° 1, pages 48–62.
- [Rubin *et al.*, 2006] RUBIN, V. L., LIDDY, E. D. et KANDO, N. (2006). Certainty identification in texts: Categorization model and manual tagging results. *Computing Attitude and Affect in Text: Theory and Applications*, chap. 7, pages 61–76.
- [Saint-Hilaire, 1839] SAINT-HILAIRE, J. B. (1839). *Logique d'Aristote - Premiers analytiques - Tome II.* Librairie de Ladrangé, Paris.
- [Saurí *et al.*, 2006] SAURÍ, R., VERHAGEN, M. et PUSTEJOVSKY, J. (2006). Annotating and recognizing event modality in text. *Proceedings of the 19th international FLAIRS conference, FLAIRS 2006, Melbourne Beach, Florida, May 11-13*, pages 333–338.
- [Saussure (de), 2012] SAUSSURE (DE), L. (2012). Modalité épistémique, évidentialité et dépendance contextuelle. *Langue française*, 2012, vol. 1, n° 173, pages 131–143. Armand Colin, Paris.
- [Schlaufer, 2016] SCHLAUFER, C. (2016). The Narrative Uses of Evidence. *Policy Studies Journal*, 2018, vol. 41, n° 1, pages 90–118.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Shackley et Wynne, 1996] SHACKLEY, S. et WYNNE, B. (1996). Representing Uncertainty in Global Climate Change Science and Policy: Boundary-Ordering Devices and Authority. *Science, Technology, & Human Values*, été 1996, vol. 21, n° 3, page 275–302.
- [Shane (Mc) et al., 2005] SHANE (MC), L. M., ALTMAN, D. G., SAUERBREI, W., TAUBE, S. E., GION, M. et CLARK, G. M. (2005). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). *Journal of the National Cancer Institute*, 17 août 2005, Vol. 97, n° 16, pages 1180–1184.
- [Shelton, 1998] SHELTON, D. (1998). Certitude et incertitude scientifiques. *Revue Juridique de l'Environnement*, numéro spécial - L'irréversibilité, pages 39–47.
- [Shotton, 2009] SHOTTON, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, avril 2009, n° 22, page 85–94.
- [Skitmore et Runeson, 1996] SKITMORE, M. et RUNESON, G. (1996). Criteria for evaluating research outcomes. In *Proceedings Australian Universities Building Educators Association Annual Conference, Royal Melbourne Institute of Technology, Melbourne*.
- [Sluijs (van der) et al., 2006a] SLUIJS (VAN DER), J. P., DOUGUET, J.-M., JANSSEN, P. H. M., O'CONNOR, M. et PETERSEN, A. C. (2006a). Tools to Assess Uncertainty in a Deliberative Perspective. A Catalogue. *Cahiers du C3ED*, juin 2006, n° 06.02 *Tools to Assess Uncertainty in a Deliberative Perspective. A Catalogue*, pages 2–35. Centre d'Economie et d'Ethique pour l'Environnement et le Développement, Université de Versailles Saint-Quentin-en-Yvelines, Guyancourt, France.
- [Sluijs (van der) et al., 2006b] SLUIJS (VAN DER), J. P., DOUGUET, J.-M., O'CONNOR, M., GUIMARAES PEREIRA, A., CORRAL QUINTANA, S., MAXIM, L. et RAVETZ, J. R. (2006b). Qualité de la Connaissance dans un processus délibératif. *Cahiers du C3ED*, février 2006, n° 06.03 *Evaluation de la Qualité de la Connaissance dans une Perspective Délibérative*, pages 1–18. Centre d'Economie et d'Ethique pour l'Environnement et le Développement, Université de Versailles Saint-Quentin-en-Yvelines, Guyancourt, France.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Sluijs (van der) *et al.*, 2008] SLUIJS (VAN DER), J. P., PETERSEN, A. C., JANSSEN, P. H. M., RISBEY, J. S. et RAVETZ, J. R. (2008). Exploring the quality of evidence for complex and contested policy decisions. *Environmental Research Letters*, 2008, vol. 3, n° 2, pages 1–9.
- [Smith *et al.*, 2018] SMITH, K. A., WILBY, R. L., BRODERICK, C., PRUDHOMME, C., MATTHEWS, T., HARRIGAN, S., et MURPHY, C. (2018). Navigating Cascades of Uncertainty - As Easy as ABC? Not Quite... *Extreme Events*, 2018, vol. 5, n° 1.
- [Sokolova et Lapalme, 2008] SOKOLOVA, M. et LAPALME, G. (2008). Verbs Speak Loud: Verb Categories in Learning Polarity and Strength of Opinions. Bergler S. (ed.) *Advances in Artificial Intelligence. Canadian AI 2008. Lecture Notes in Computer Science*, vol. 5032, page 320–331.
- [Solla Price (de), 1965] SOLLA PRICE (DE), D. J. (1965). Is Technology Historically Independent of Science? A Study in Statistical Historiography. *Technology and Culture*, 1965, vol. 6, n° 4, pages 553–568.
- [Sollaci et Pereira, 2004] SOLLACI, L. B. et PEREIRA, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the medical library association*, juillet 2004, vol. 92, n° 3, pages 364–367.
- [Sous-comité 26 du Comité Z39 de l'ANSI, 1979] SOUS-COMITÉ 26 DU COMITÉ Z39 DE L'ANSI (1979). ANSI Z39.16-1979 - American national standard for the preparation of scientific papers for written or oral presentation: approved January 9, 1979 - Revision of ANSI Z39.16-1972. Rapport technique, American National Standards Institute - ANSI, Washington, district de Columbia, EUA. Référence de série de la norme : American national standards on library work and documentation - ANSI Z39.16-1979. Référence du Comité Z39 de l'ANSI : « American National Standards Institute, Inc. (ANSI), Committee Z39 » « on Library Work, Documentation, and Related Publishing Practices » ou « on Library and Information Sciences and Related Publishing Practices » ; ce Comité Z39 étant devenu en 1984 le "National Information Standards Organization Z39" (NISO).
- [Sterken, 2011] STERKEN, C. (2011). Scientific Writing for Young Astronomers - Writing a Scientific Paper - I. The Writing Process. *EAS Publications Series*, 2011, n° 50, page 1–63.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Szarvas *et al.*, 2008] SZARVAS, G., VINCZE, V., FARKAS, R. et CSIRIK, J. (2008). The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. *BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, juin 2008, Columbus, Ohio, EUA, page 38–45.
- [Sztahó *et al.*, 2009] SZTAHÓ, D., NAGY, K. et VICSI, K. (2009). Automatic Sentence Modality Recognition in Children’s Speech, and Its Usage Potential in the Speech Therapy. *Lecture Notes in Computer Science* (éd. Esposito A. et Vích R.) - Subseries *Lecture Notes in Artificial Intelligence* (éd. Goebel R., Siekmann J. et Wahlster W.) - vol. LNAI 5641 « Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions - COST Action 2102 International Conference Prague, Czech Republic, October 15-18, 2008 - Revised Selected and Invited Papers », pages 266–275. Springer, Berlin, Allemagne.
- [Trask, 1999] TRASK, R. L. (1999). *Key concepts in language and linguistics*. Routledge, Londres, RU.
- [Usher et Strachan, 2013] USHER, W. et STRACHAN, N. (2013). An expert elicitation of climate, energy and economic uncertainties. *Energy Policy*, octobre 2013, vol. 61, page 811–821. Elsevier BV, Amsterdam, Pays-Bas.
- [Vernier, 2011] VERNIER, M. (2011). *Analyse à granularité fine de la subjectivité*. Thèse de doctorat.
- [Vernier *et al.*, 2009] VERNIER, M., MONCEAUX, L. et DAILLES, B. (2009). DEFT’09 : détection de la subjectivité et catégorisation de textes subjectifs par une approche mixte symbolique et statistique. *Atelier Défi Fouille de Textes (DEFT’09)*, Juin 2009, Paris, France, p.101-112.
- [Villers (de), 1801] VILLERS (DE), C. (1801). *Philosophie de Kant, ou Principes fondamentaux de la philosophie transcendental*. Collignon, Metz.
- [Vinzerich, 2007] VINZERICH, A. (2007). *La sémantique du possible : approche linguistique, logique et traitement informatique dans les textes*. Thèse de doctorat.
- [Viviani, 1994] VIVIANI, J.-L. (1994). Incertitude et rationalité. *Revue française d’économie*, 1994, vol. 9, n° 2, pages 105–146.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Vold, 2006] VOLD, E. T. (2006). Epistemic modality markers in research articles: a cross-linguistic and crossdisciplinary study. *International Journal of Applied Linguistics*, 2006, vol. 16, n° 1.
- [Vold, 2008] 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*. Thèse de doctorat.
- [W3C, 2008] W3C (2008). Uncertainty reasoning for the world wide web. Rapport technique, W3C Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) - W3C (MIT, ERCIM, Keio).
- [Willett, 1988] WILLETT, T. (1988). A Cross-Linguistic Survey of the Grammaticization of Evidentiality. *Studies in Language*, janvier 1988, vol. 12, n° 1, pages 51–97.
- [Williams *et al.*, 2009] WILLIAMS, M., CORNFORD, D., BASTIN, L. et PEBESMA, E. (2009). Uncertainty Markup Language (UnCertML). Document ref.: 08-122r2. Rapport technique, Open Geospatial Consortium Inc., Wayland, Massachusetts, EUA.
- [Willows *et al.*, 2003] WILLOWS, R., CONNELL, R., REYNARD, N. et MEADOWCROFT, I. (2003). Climate adaptation: Risk, uncertainty and decision-making. Rapport technique, UK Climate Impacts Programme (UKCIP), Oxford, Royaume-Uni.
- [Wu, 2011] WU, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, 5 novembre 2011, vol. 26, n° 10, page 1345–1349.
- [Yvon, 2007] YVON, F. (2007). *Une petite introduction au Traitement Automatique des Langues Naturelles*. École nationale supérieure des télécommunications - Télécom ParisTech, Paris, France. Version du 26/04/2007.
- [Zegarelli, 2007] ZEGARELLI, M. (2007). *Logic for dummies*. Wiley Publishing, Inc., Hoboken, New Jersey, EUA.

RÉFÉRENCES BIBLIOGRAPHIQUES

Liste des tableaux

2.1	Typologies de l'incertitude selon l'IPCC	14
2.2	Echelle de probabilité de l'IPCC	15
2.3	Double échelle de confiance de l'IPCC : dimensions à croiser . . .	15
2.4	Double échelle de confiance de l'IPCC : croisement des dimensions	16
2.5	Charte des degrés d'incertitude de Sherman Kent	17
2.6	Les huit principes de confiance du HONcode	22
3.1	Diversité de définitions de l'incertitude : exemples	26
3.2	Opérateurs verbaux finis selon Halliday et Matthiessen : extrait . .	44
3.3	Compléments adverbiaux modaux selon Halliday et Matthiessen .	45
3.4	Matrice modale implicite-explicite et subjectif-objectif selon Hal- liday et Matthiessen	45
3.5	Catégories d'orientations de hedges selon Hyland	58
3.6	Avantages comparatifs des types de modélisation des expressions subjectives selon Fraisse et Paroubek	61
4.1	Axe de véridicité de Saurí-Verhagen-Pustejovsky	78
6.1	Dénombrement des phrases du gold standard selon leur champ et corpus	143
8.1	Matrice de confusion (modèle explicatif)	164
8.2	Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec la première version des règles.	166
8.3	Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec les règles finales.	168
8.4	Résultats de l'expérimentation sur le corpus d'étude (200 phrases, partie 2).	169
8.5	Résultats de l'expérimentation sur le corpus d'étude (700 phrases).	170

LISTE DES TABLEAUX

8.6	Résultats de l'évaluation.	171
8.7	Résultats sur le gold standard.	172

Liste des schémas

1.1	Objectifs	3
3.1	Double dimension de la causalité dans la langue	41
3.2	Carré d'Aristote	47
3.3	Carré des modalités aléthiques	48
3.4	Une structure des modalités linguistiques : dimensions	51
3.5	Une structure des modalités linguistiques : exemples	51
4.1	L'approche microsystemique	83
4.2	Fonctionnement de la méthode de l'Exploration Contextuelle	85
5.1	Pile du web sémantique	99
5.2	Caractéristiques de l'incertitude scientifique	109
5.3	Ontologie de l'incertitude	110
6.1	Etapes principales de la méthodologie	112
6.2	Composition du Gold Standard	132
6.3	Ontologie de l'incertitude scientifique, avec les abréviations utilisées pour l'annotation	135
6.4	Comparaison des annotations des experts : aperçu du document CSV	138
6.5	Gold standard : aperçu du document CSV	139
6.6	Dénombrement des phrases du gold standard selon leur champ scientifique	141
6.7	Répartition proportionnelle des phrases du gold standard selon leur champ et corpus	142
7.1	Règles linguistiques : aperçu du document CSV	152

8.1	Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec la première version des règles : Précision, Rappel, F-mesure	166
8.2	Résultats de l'expérimentation sur le corpus d'étude (500 phrases, partie 1) avec les règles finales : Précision, Rappel, F-mesure . . .	168
8.3	Résultats de l'expérimentation sur le corpus d'étude (200 phrases, partie 2). : Précision, Rappel, F-mesure	169
8.4	Résultats de l'expérimentation sur le corpus d'étude (700 phrases) : Précision, Rappel, F-mesure	170
8.5	Résultats de l'évaluation : Précision, Rappel, F-mesure	171
8.6	Résultats sur le gold standard : Précision, Rappel, F-mesure	172
8.7	Précision, Rappel et F-mesure pour les différents corpus	173

Annexe A

Exemples de définitions

A.1 Exemples de définitions de la notion de modalité

- « Ensemble de formes permettant au locuteur d'indiquer la manière dont il envisage le contenu de son énoncé ». Source : Larousse « Dictionnaire de français » édition en ligne. Consulté le 24/11/2018 sur <https://www.larousse.fr/dictionnaires/francais/>
- « Ensemble des faits linguistiques (mode, forme assertive, interrogative ou injonctive de la phrase, adverbes ou auxiliaires modaux) traduisant l'attitude du sujet parlant par rapport à ce qu'il énonce ». Source : Centre National de Ressources Textuelles et Lexicales (CNRTL), consulté le 26/08/2019 sur <https://www.cnrtl.fr/definition/modalite> (§ *Linguistique*).
- Ensemble des formes exprimant l'attitude du locuteur par rapport au contenu de son énoncé. Source : [Le Querler, 1996].
- Modalité épistémique : modalité exprimant un degré de véracité scientifique de l'énoncé. Source : [Vold, 2006].
- « **modality**, **mode** (a classification of propositions on the basis of whether they claim necessity or possibility or impossibility) ». « **mood**, **mode**, **modality** (verb inflections that express how the action or state is conceived by the speaker) ». Source : *WordNet - A Lexical Database for En-*

glish, Princeton University, New Jersey, EUA. Consulté le 26/08/2019 sur <http://wordnetweb.princeton.edu/perl/webwn>.

- « **mood** n : a set of contrasts which are often shown by the form of the verb and which express the speaker's or writer's attitude to what is said or written ». Source : [Richards et Schmidt, 2010].
- « **modality** : The grammatical category associated with the expression of obligation, permission, prohibition, necessity, possibility and ability. It is by no means easy to separate *modality* from the more traditional category of **mood** , which expresses degree or kind of reality [...]. In practice, modality shades off imperceptibly into several other categories: *evidentiality* (the amount and nature of the evidence which you have for saying something), *modalization* (the probability or regularity of an occurrence), *modularity* (the degree of commitment or willingness involved on somebody's part), *hedging* (reducing your own commitment to what you are saying) and *vague language*. These extensions of modality have been particularly investigated within **Systemic Linguistics**. ». « **mood** : The grammatical category which expresses the degree or kind of reality attached to an utterance. Mood is not well developed as a grammatical category in English, and we mostly use words and constructions to express mood distinctions. In this way, we can assign varying degrees of certainty to an utterance [...] We can also assign different degrees of reality within conditions [...] ». Source : [Trask, 1999].
- « **Modality** (n.) : A term used in GRAMMATICAL and SEMANTIC analysis to refer to contrasts in MOOD signalled by the VERB and associated categories. In English, modal contrasts are primarily expressed by a subclass of AUXILIARY verbs, e.g. may, will, can. This subclass is symbolized as **M** in the PHRASE-STRUCTURE RULES of a GENERATIVE grammar. Modal verbs share a set of morphological and syntactic properties which distinguish them from the other auxiliaries, e.g. No -s, -ing or -en forms. In CASE grammar, **modality** refers to one of the two major CONSTITUENTS of a sentences's DEEP STRUCTURE, the other being PROPOSITION ». « **Mood** (n.) : A term used in the theoretical and descriptive study of SENTENCE/ CLAUSE types, and especially of the verbs they contain. Mood (**modality**, or **mode**) refers to a set of SYNTACTIC and SEMANTIC CONTRASTS signalled by alternative PARADIGMS of the verb, e.g. INDICATIVE (the UN-

MARKED form), SUBJUNCTIVE, IMPERATIVE. Semantically, a wide range of meanings is involved, especially attitudes on the part of the speaker towards the factual content of the utterance, e.g. uncertainty, definiteness, vagueness, possibility. Syntactically, these contrasts may be signalled by alternative INFLECTIONAL forms of a verb, or by using AUXILIARIES. English mainly uses **modal auxiliaries**, e.g. *may, can, shall, must*, but makes a little use of inflection (e.g. *If I were you* v. *I was...*). The semantic analysis of **modal verbs**, and the study of their distribution in everyday speech, is a topic which has attracted a great deal of attention in LINGUISTICS, and several classifications involving such notions as necessity, possibility, certainty, etc., have been proposed. The results of such studies have implications for fields other than linguistics; for example, theoretical modal distinctions involving such notions have been a major concern of logicians. See also ALETHIC, DEONTIC, EPISTEMIC ». Source : [Crystal, 2008].

- « **mood**¹ (noun) (Origin : Old English *mōd* (also in the senses ‘mind’ and ‘fierce courage’), of Germanic origin; related to Dutch *moed* and German *Mut*.) : **1.** A temporary state of mind or feeling. (**1.1** The atmosphere or pervading tone of something. **1.2** [*as modifier*] (especially of music) inducing or suggestive of a particular feeling or state of mind.) **2.** An angry, irritable, or sullen state of mind. **mood**² (noun) (Origin : Mid 16th century: variant of *mode*, influenced by *mood*.) : **1.** (Grammar) A category or form which indicates whether a verb expresses fact (indicative mood), command (imperative mood), question (interrogative mood), wish (optative mood), or conditionality (subjunctive mood). **2.** (Logic) Any of the valid forms into which each of the figures of a categorical syllogism may occur ». Source : « Oxford Dictionaries » version anglaise, Oxford University Press. Consulté le 24/11/2018 sur <https://www.oxforddictionaries.com>.

A.2 Exemples de définitions de la notion de doute

- Source : Centre National de Ressources Textuelles et Lexicales (CNRTL), consulté le 27/08/2019 sur <https://www.cnrtl.fr/definition/doute>
 - Sens général : « État naturel de l’esprit qui s’interroge, caractérisé à

des degrés différents soit par l'incertitude concernant l'existence ou la réalisation d'un fait, soit par l'hésitation sur la conduite à tenir, soit par la suspension du jugement entre deux propositions contradictoires. [...] Anton. *certitude* ».

- Philosophie : « *Doute philosophique, doute méthodique de Descartes.* Attitude du sujet pensant qui considère tout jugement sur tout objet de connaissance comme douteux afin de tendre vers la plus grande certitude possible, la certitude première étant celle du sujet pensant lui-même », « La méthode [cartésienne] est applicable à l'édification des sciences d'observation et d'expérimentation. Il faut procéder toujours par le **doute philosophique**, avec précaution, avec défiance ».
- « **doubt**, uncertainty, incertitude, dubiety, doubtfulness, dubiousness (the state of being unsure of something) », « **doubt**, dubiousness, doubtfulness, question (uncertainty about the truth or factuality or existence of something) "the dubiousness of his claim"; "there is no question about the validity of the enterprise" ». Source : *WordNet - A Lexical Database for English*, Princeton University, New Jersey, EUA. Consulté le 24/11/2018 sur <http://wordnetweb.princeton.edu/perl/webwn>.
- « **doubt** (mass noun / count noun) (Origin : Middle English: from Old French doute (noun), douter (verb), from Latin dubitare 'hesitate', from dubius 'doubtful' (see dubious).) : A feeling of uncertainty or lack of conviction. **doubt** (verb) : **1.** (with object) Feel uncertain about. **1.1** Question the truth or fact of (something). **1.2** Disbelieve or lack faith in (someone). **1.3** no object Feel uncertain, especially about one's religious beliefs. **2.** (archaic, with clause) Fear; be afraid. **Phrases** : « **beyond doubt** » : Allowing no uncertainty. « **in doubt** » : **1.** Open to question (**1.1** Feeling uncertain about something). « **no doubt** » : Used to indicate the speaker's firm belief that something is true (**1.1** Used to introduce a concession which is subsequently dismissed as unimportant or irrelevant). « **without (a) doubt** » : Indisputably ». Source : Oxford University Press (2018), « Oxford Dictionaries » version anglaise. Consulté le 24/11/2018 sur <https://www.oxforddictionaries.com>.
- « État de quelqu'un qui ne sait que croire, qui hésite à prendre parti », « Manque de certitude, soupçon, méfiance quant à la sincérité de quelqu'un, la véracité d'un fait, la réalisation de quelque chose (surtout pluriel) », « Po-

sition philosophique qui consiste à refuser quelque chose parmi des données jugées incertaines, par exemple les données des sens, et à refuser tout système philosophique ». Source : Larousse « Dictionnaire de français » édition en ligne. Consulté le 24/11/2018 sur <https://www.larousse.fr/dictionnaires/francais/>.

- Nom : (RU) « (a feeling of) not being certain about something, especially about how good or true it is », (EUA) « a feeling of not knowing what to believe or what to do, or the condition of being uncertain ». Verbe : (RU) « to not feel certain or confident about something or to think that something is not probable », (EUA) « to be uncertain about something or someone, or to have difficulty believing something ». Source : « Cambridge Dictionary » version « English » seul, Cambridge University Press (2018). Consulté le 27/08/2019 sur <https://dictionary.cambridge.org/dictionary/english/>

A.3 Exemples de définitions de la notion d'incertitude

- « Caractère d'imprécision d'une mesure, d'une conclusion. Synon. imprécision, relativité ». « *PHYS., MÉTROL.*, Quantité mesurant la limite de l'erreur éventuelle pouvant affecter une mesure » (attribué à : *Dew. Mes. 1973*). « Caractère imprécis, vague d'une perception, d'une image. Synon. *flou, imprécision*; anton. *netteté, précision* ». « Caractère imprévisible du résultat d'une action, d'une évolution ». « Impossibilité dans laquelle est une personne de connaître ou de prévoir un fait, un événement qui la concerne ; sentiment de précarité qui en résulte ». « État d'esprit d'une personne qui hésite à croire à la réalité d'un fait, à la vérité d'un jugement ou qui hésite à adopter et à maintenir une ligne de conduite. Synon. *embarras, doute, hésitation, indécision, irrésolution, perplexité*; anton. *assurance, détermination, résolution*. ». Source : Centre National de Ressources Textuelles et Lexicales (CNRTL), consulté le 26/08/2019 sur <https://www.cnrtl.fr/definition/incertitude> (§ *Linguistique*)..
- (Physique) Source : [JCGM, 2012, p. 17, 25-27] :
 - L'**incertitude de mesure** « est un paramètre non négatif caractérisant

la dispersion des valeurs attribuées à un mesurande » (un mesurande est la « grandeur que l'on veut mesurer »), « Le paramètre peut être [...] un écart-type appelé incertitude-type ».

- L'**incertitude définitionnelle** est la « composante de l'incertitude de mesure qui résulte de la quantité finie de détails dans la définition d'un mesurande ».
 - Le **bilan d'incertitude** (*uncertainty budget*) est la « formulation d'une incertitude de mesure et des composantes de cette incertitude, ainsi que de leur calcul et de leur combinaison ».
 - Une **évaluation de type A de l'incertitude** est une « d'une composante de l'incertitude de mesure par une analyse statistique des valeurs mesurées obtenues dans des conditions définies de mesurage »
 - Une **évaluation de type B de l'incertitude** est une « évaluation d'une composante de l'incertitude de mesure par d'autres moyens qu'une évaluation de type A de l'incertitude » (p. ex. : des valeurs publiées faisant autorité, la valeur d'un matériau de référence certifié, un certificat d'étalonnage, la classe d'exactitude d'un instrument de mesure vérifié, des limites déduites de l'expérience personnelle).
 - L'**incertitude cible** ou incertitude anticipée (*target measurement uncertainty* ou *target uncertainty*) est l'incertitude de mesure « spécifiée comme une limite supérieure et choisie d'après les usages prévus des résultats de mesure ».
- « uncertainty, uncertainness, precariousness (being unsettled or in doubt or dependent on chance) "the uncertainty of the outcome"; "the precariousness of his income" », « doubt, uncertainty, incertitude, dubiety, doubtfulness, dubiousness (the state of being unsure of something) ». Source : *WordNet - A Lexical Database for English*, Princeton University, New Jersey, EUA. Consulté le 24/11/2018 sur <http://wordnetweb.princeton.edu/perl/webwn>.
 - « **uncertainty** (noun) (plural : uncertainties) : 1. The state of being uncertain. 1.1 (usually **uncertainties**) [*count noun*] Something that is uncertain or that causes one to feel uncertain ». Source : « Oxford Dictionaries » version anglaise, Oxford University Press. Consulté le 24/11/2018 sur <https://www.oxforddictionaries.com>.

- Source : Larousse « Dictionnaire de français » édition en ligne. Consulté le 24/11/2018 sur <http://www.larousse.fr/dictionnaires> :
 - « Caractère de ce qui est incertain ».
 - « État de quelqu'un qui ne sait quel parti prendre, ou état plus ou moins préoccupant de quelqu'un qui est dans l'attente d'une chose incertaine : Être dans une profonde incertitude et incapable de se décider ».
 - Mathématiques : « Majorant de la valeur absolue de l'erreur Δ_a d'un nombre a ».
 - Métrologie : « Défaut du résultat d'une mesure qui s'écarte de la valeur vraie d'une quantité inconnue, aléatoire, dont on peut estimer l'ordre de grandeur mais non le signe ».
 - Physique : « Relations d'incertitude, relations d'inégalité énoncées par W. Heisenberg, qui traduisent l'impossibilité de mesurer simultanément la vitesse et la position d'une particule ». (Consulté le 24/11/2018).
- « a situation in which something is not known, or something that is not known or certain ». Source : « Cambridge Dictionary » version « English » seul, Cambridge University Press (2018). Consulté le 27/08/2019 sur <https://dictionary.cambridge.org/dictionary/english/>

A.4 Exemples de définitions autour de la notion d'incertitude liée au changement climatique

Ambiguity (climatique) : « A characteristic of a system or decision where there is only vague belief concerning the probabilities that certain states or outcomes may occur. A form of uncertainty ». Source : [Willows *et al.*, 2003, p. 111, § *Glossary*].

Scenario (climatique) : « A coherent, internally consistent and plausible description of a possible future state of the world, usually based on specific assumptions ». Source : [Willows *et al.*, 2003, p. 117, § *Glossary*].

'Subjective' probabilities (climatique) : « Probabilities and probability distributions are often used to express or summarise strength of belief. There is

continuing debate about the validity of this, however in pragmatic terms there are good reasons to accept the use of subjective probabilities if it helps inform a decisionmaking process. The following should be borne in mind: [...] Probabilities should be elicited from a number of people with similar appropriate experience, knowledge or expertise. [...] ». Source : [Willows *et al.*, 2003, p. 115, § *Glossary*].

Uncertainty (climatique) : « A characteristic of a system or decision where the probabilities that certain states or outcomes have occurred or may occur is not precisely known. A concept that reflects a lack of confidence about something, including forecasts. Decision-makers may have more or less certain knowledge of a risk ». Source : [Willows *et al.*, 2003, p. 118, § *Glossary*].

Uncertainty (climatique) : « is an expression of the degree to which a value (e.g., the future state of the climate system) is unknown ». Source : [Lim *et al.*, 2004, p. 250, § *Annex A: Glossary of Terms*].

A.5 Exemples de définitions de la notion de subjectivité en linguistique

subjective (linguistique) : « The meaning of an adjective is most typically a temporary or permanent state or condition : *big* [...]. Many adjectives express subjective perceptions, rather than objective facts : *interesting* [...]. A few adjectives express very unusual types of meaning: [...] *heavy* in *She's a heavy smoker* ». (Source : [Trask, 1999]).

subjectivité₁ (linguistique) : (ang. : *subjectivity* ; esp. : *subjetividad*). « On appelle subjectivité la présence du sujet parlant dans son discours ; ainsi, la subjectivité du discours se manifeste par les embrayeurs » [Dubois *et al.*, 2007]. Voir aussi *embrayeur*.

subjectivity, subjectiveness (linguistique) : « judgment based on individual personal impressions and feelings and opinions rather than external facts ». (Source : Princeton University *WordNet*[®] lexical database of English. URL : <http://wordnetweb.princeton.edu/perl/webwn>).

Annexe B

Liste d'expressions clefs liées au changement climatique en anglais

A1 Marker Scenario
A2 Marker Scenario
anthropogenic climate
anthropogenic warming
B1 Marker Scenario
B2 Marker Scenario
Bio-Carbon Fund Initiative for Sustainable Forest Landscapes
carbon capture and storage
Carbon Capture and Storage
Carbon Capture Use and Storage
Carbon credit
Carbon credits
Carbon Dioxide Information Center
Carbon Dioxide Information Centers
carbon dioxide removal
carbon dioxide removals
Carbon Emission Reduction Target
Carbon Emission Reduction Targets
Carbon Price Floor
Carbon Price Support
Carbon Price Supports
Carbon Reduction Commitment
Carbon Reduction Commitments

carbon sequestration
carbon sequestrations
Ci-Dev
Clean Development Mechanism
Clean Development Mechanisms
Clean Technology Fund
Clean Technology Funds
climate action
climate action mainstreaming
climate actions
Climate and Development Knowledge Network
Climate and Development Knowledge Networks
Climate and Energy Package
Climate and Energy Packages
climate change
climate changes
climate conference
climate conferences
climate credit
climate credits
climate crisis
climate debt
Climate Development Finance Facilities
Climate Development Finance Facility
climate diplomacy
climate engineering
climate finance
climate financing
climate forcer
climate forcers
climate forcing
climate fund
climate funds
Climate Innovation Centre
Climate Innovation Centres
climate intervention
climate interventions

Climate Investment Fund
Climate Investment Funds
Climate Investor One
climate policies
climate policy
Climate Public Private Partnership Programme
Climate Public Private Partnership Programmes
climate resilience
climate resiliences
climate resilient
climate response
climate responses
climate scenario
climate scenarios
climate sensitivity
climate simulation
climate simulations
climate summit
climate summits
Climate Technology Centre
Climate Technology Centres
climate warming
climate-related risk
climate-related risks
climate-related sensitivity
compatible carbon emissions
compatible CO2 emissions
compatible emissions
compatible fossil fuel emmissions
Conference of Parties - COP
Conference of Parties (COP)
Conference of the Parties to the UNFCCC
Conferences of Parties - COPs
Conferences of Parties (COPs)
COP 15
COP 17/CMP
COP 21

COP 21/CMP11
COP/CMP
COP15
COP17/CMP
COP21
COP21/CMP11
Coupled Model Intercomparison Project
Coupled Model Intercomparison Projects
cumulative carbon emissions
cumulative CO2 emissions
cumulative emissions of carbon
cumulative emissions of CO2
direct air capture
direct air captures
Emissions Reduction Unit
Emissions Reduction Units
Energy and Emissions Projection
Energy and Emissions Projections
Energy Efficiency Commitment
Energy Efficiency Commitments
enhanced greenhouse effect
enhanced greenhouse effects
ENVE Commission
equity fund Marguerite
equivalent carbon dioxide (CO2) emission
equivalent carbon dioxide (CO2) emissions
equivalent carbon dioxide emission
equivalent carbon dioxide emissions
EU Emissions Trading System
EU ETS
Extended Concentration Pathway
Extended Concentration Pathways
FCCC
Feed-in Tariffs
Forest Carbon Partnership Facility
Forest Governance, Markets and Climate
fugitive emissions

Global Climate Partnership Fund
Global Environmental Change
global warming
global warming potential
Green Bus Fund
Green Climate Fund
Green Climate Funds
greenhouse gases and aerosols
Group on global change, climate and biodiversity
halocarbon global warming potential
INDC
INDCs
intended nationally determined contribution
intended nationally determined contributions
Intergovernmental Panel on Climate Change
International Climate Fund
IPCC
Kyoto Protocol
Major Economies Forum on Energy and Climate
Marguerite Fund
migration influenced by environmental change
migration influenced by environmental changes
Million tonnes of carbon dioxide equivalent
Montreal Protocol
National Appropriate Mitigation Action
National Appropriate Mitigation Actions
Nationally Appropriate Mitigation Action
Nationally Appropriate Mitigation Actions
ocean acidification
ocean acidifications
ozone hole
ozone holes
Paris climate agreement
Pilot Programme for Climate Resilience
post-Kyoto agreement
Programme for Climate Resilience
QEWER

QEWERS

Quantified Economy-Wide Emission Reduction

Quantified Economy-Wide Emission Reductions

Reducing Emissions from Deforestation and Forest Degradation

Renewable Energy Directive

Renewable Energy Directives

Renewable Energy Package

Renewable Energy Packages

Renewable Heat Incentive

Renewable Heat Incentives

Renewable Transport Fuel Obligation

Renewable Transport Fuel Obligations

Renewables Obligation Certificate

Renewables Obligation Certificates

representative Concentration Pathway

representative Concentration Pathways

Scaling Up Renewable Energy Programme

Scaling Up Renewable Energy Programmes

Solar radiation management

Special Report on Emissions Scenarios

SRES emissions scenario

SRES emissions scenarios

SRES scenario

SRES scenarios

substances that deplete the ozone layer

sustainable transition

sustainable transitions

transient climate

UNFCCC

United Nations Climate Conference

United Nations Climate Conferences

United Nations Framework Convention on Climate Change

urban heat island effect

urban heat island effects

Woodland Carbon Code

World Bank's Carbon Initiative for Development

Annexe C

Mode d'emploi fourni aux experts pour l'annotation manuelle

Notice explicative de l'expertise, pour les stagiaires

Contexte de l'étude à laquelle vous participez

Vous allez participer, en tant qu'experts de la langue anglaise, à la classification de phrases selon le type des *expressions d'incertitude* qu'elles contiennent. (Ces phrases sont extraites d'un corpus d'articles scientifiques liés au changement climatique de la Terre).

Cette classification est une étape décisive pour la découverte de règles d'identification et classification de l'expression de l'incertitude dans les textes scientifiques.

Matériel dont vous disposez

Vous disposez d'une liste de phrases en anglais, fournies dans un fichier CSV qui s'ouvre sous forme d'un tableau avec un logiciel tableur (de la suite bureautique que vous utilisez) : *Calc* (LibreOffice ou OpenOffice) ou *Excel* (Microsoft Office), ou *Numbers* (Apple iWork).

Dans le tableau, chaque ligne sert à l'évaluation d'une phrase différente, comme le résume le tableau suivant :

A Référence du texte d'origine de la phrase.	B Numéro de la phrase dans le texte.	C La phrase à catégoriser.	D La même phrase en contexte : entourée de la précédente et la de la suivante.
...F766.tei.xml	160	In addition, the CH4 fluxes were found to be...	Phrase précédente. In addition... Phrase suivante.
...7EB0F.tei.xml	136	The use of excessive variables in...	Phrase précédente. The use of... Phrase suivante.
Réf. de texte.	N° de phrase.	Phrase à catégoriser.	Phrase précédente. Phrase à catégoriser. Phrase suivante.
Réf. de texte.	N° de phrase.	Phrase à catégoriser.	Phrase précédente. Phrase à catégoriser. Phrase suivante.
Réf. de texte.	N° de phrase.	Phrase à catégoriser.	Phrase précédente. Phrase à catégoriser. Phrase suivante.

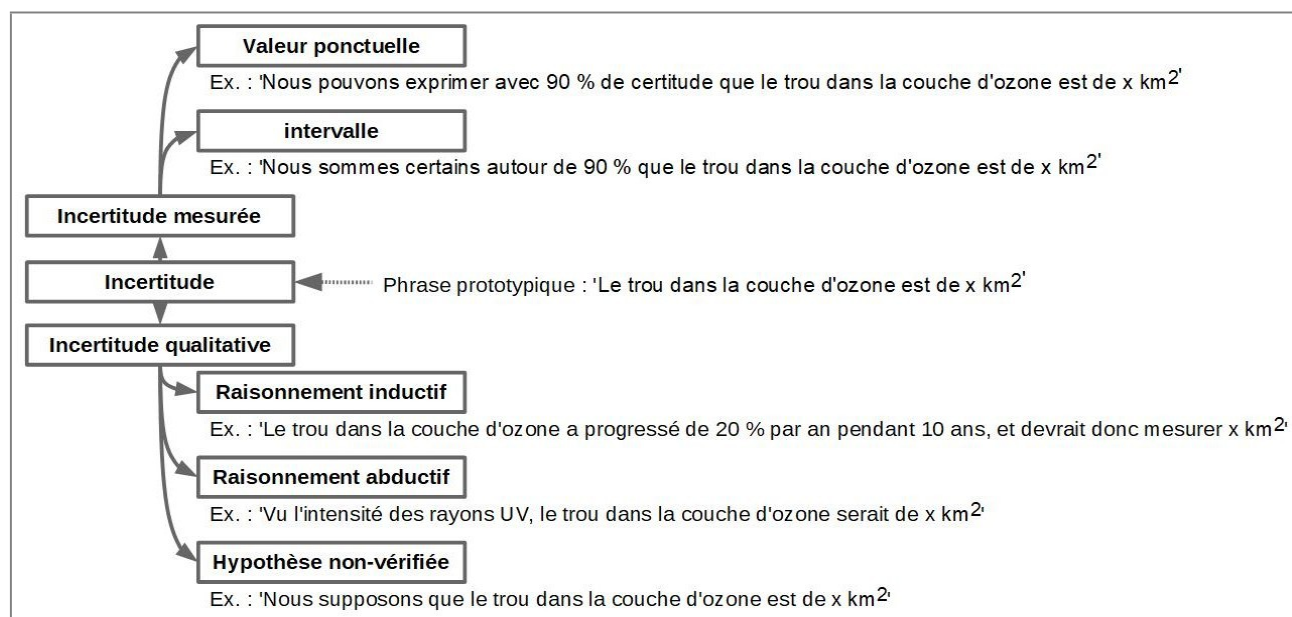
Exemples d'extraits de phrases dont le contenu exprimerait de l'incertitude (de possibles marqueurs d'une expression d'incertitude sur des faits scientifiques y sont soulignés) :

- « ... in a small local population of about 20 pairs near the southern coast of Finland ».
- « The rising of spring temperatures to a certain level often seems to be a key... ».
- « ... as well as the possible (but probably minor) "calendar effect"... ».
- « ... but there was no significant trend... ».
- « ... the link via effects of the winter weather on the timing of spring migration cannot be excluded... »

A coté de cette liste de phrases accompagnées de leurs contextes, vous disposez de 7 qualifications possibles de leur contenu. **Votre tâche est de reconnaître et indiquer la catégorie du contenu d'incertitude de chaque phrase.** Au final, vous fournissez un fichier CSV où les phrases sont qualifiées selon leur incertitude, à l'aide de 7 catégories de qualifications possibles. Voici une présentation des 7 catégories :

Types des expressions de l'incertitude pour leur classification

Voici ci-dessous un schéma indiquant les 5 types d'expression de l'incertitude scientifique :



A ces 5 types d'incertitude, s'ajoutent 2 autres classifications possibles du contenu des phrases : « aucune incertitude » et « classification non-décidable ».

Vous avez donc 7 possibilités pour qualifier le contenu d'incertitude exprimé dans les phrases soumises à votre jugement :

mvp « mesure de valeur ponctuelle »
mdi « mesure d'intervalle »
rin « raisonnement inductif »
rab « raisonnement abductif »
hnv « hypothèse non-vérifiée »

et

ai « aucune incertitude » (ex. : présentation d'un fait, raisonnement déductif, etc.)
nd « non-décidable » (catégorie à éviter autant que possible)

En présence de plusieurs catégories dans une même phrase, les noter séparées par un blanc. Ex. : « mvp mdi »

Explication sur les types de raisonnement en sciences

Déduction : on constate un enchaînement de causes certaines à effets certains.

L'énoncé (3^e phrase) est certain (si les prémisses¹ sont vraies).

Si l'on ne rencontre qu'un raisonnement déductif dans une phrase → **catégorie « ai »**

1. (Si la toundra se réchauffe, le pergélisol² dégèle.)
2. (Si le pergélisol dégèle, du méthane est émis dans l'atmosphère.)
3. La toundra se réchauffe, donc du méthane est émis dans l'atmosphère.

Induction : on généralise des observations certaines à des effets probables.

L'énoncé est probable. (effets = dans un futur) → **catégorie « rin »**

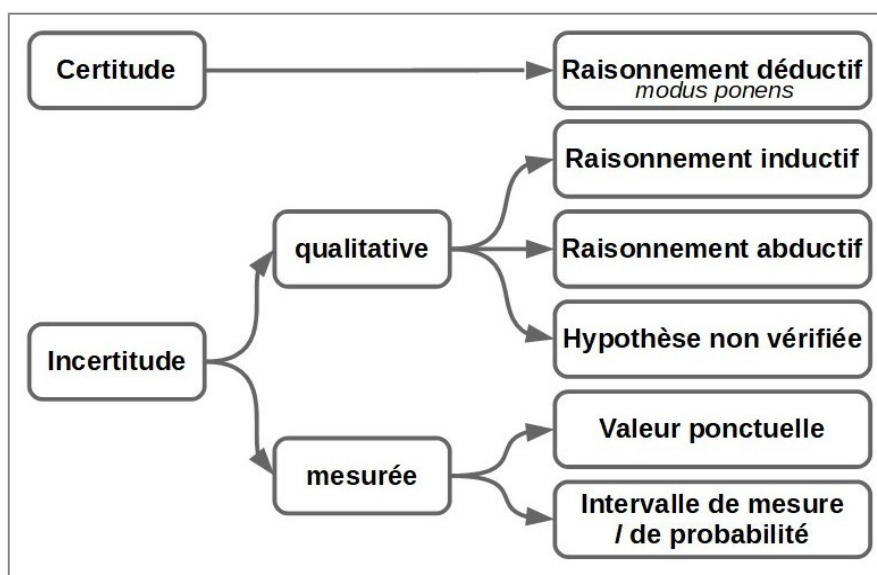
- La force des ouragans a augmenté chaque année les 10 dernières années, donc la force des ouragans augmentera cette prochaine année.

Abduction : partant d'un fait certain, on en infère des causes probables.

L'énoncé est plausible. (causes = dans un passé) → **catégorie « rab »**

- L'épaisseur des stries des arbres coupés sont régulières, donc il n'a pas gelé.

Schéma résumant :



1 Prémisse « désigne chacune des deux propositions d'un syllogisme d'où résulte la conclusion ; par extension, ce nom désigne aussi tout argument d'où découle une conséquence. ». Source : Académie française. URL (consultée le 10/04/2019) : <http://academie-francaise.fr/premisse-premices>

2 Equivalent anglais plus connu : *permafrost*.

Annexe D

Mode de calcul des indices moyens

Nombre de phrases annotées d'une catégorie d'incertitude d'un corpus : pour chaque corpus, le nombre total de phrases d'une catégorie¹ d'incertitude est VP+FN de cette catégorie². Le nombre de phrases annotées d'une catégorie d'incertitude d'un corpus est :

$$N_{catégorie/corpus}^{phrases} = VP_{catégorie/corpus} + FN_{catégorie/corpus}$$

Par exemple, le nombre de phrases *i-q-hnv* du gold standard est :

$$\begin{aligned} N_{i-q-hnv/standard}^{phrases} &= VP_{i-q-hnv/standard} + FN_{i-q-hnv/standard} \\ &= 258 + 150 = 408 \end{aligned}$$

Somme des phrases annotées de toutes les catégories d'incertitude d'un corpus : pour chaque corpus, le nombre total du cumul des phrases annotées de toutes les catégories d'incertitude du corpus est la somme du nombre $N_{catégorie}^{phrases}$ de phrases annotées de chaque catégorie. La somme des phrases annotées de toutes les catégories d'incertitude d'un corpus est :

$$S_{corpus}^{phrases} = N_{i/corpus}^{phrases} + N_{i-m/corpus}^{phrases} + N_{i-q/corpus}^{phrases} + N_{i-m-int/corpus}^{phrases} +$$

¹Une phrase peut être annotée avec plusieurs catégories d'incertitude, et donc être comptée dans plus d'une catégorie.

²Le nombre total des phrases qui ne comportent pas cette catégorie d'incertitude annoté manuellement est FP+VN de cette catégorie : ce nombre correspond aux phrases ayant uniquement des expressions d'autres catégories d'incertitude et aux phrases n'ayant aucun catégorie d'incertitude.

$$N_{i-m-val/corpus}^{phrases} + N_{i-q-rin/corpus}^{phrases} + N_{i-q-rab/corpus}^{phrases} + N_{i-q-hnv/corpus}^{phrases}$$

Par exemple, la somme des phrases annotées de toutes les catégories d'incertitude du gold standard est :

$$\begin{aligned} S_{standard}^{phrases} &= N_{i/standard}^{phrases} + N_{i-m/standard}^{phrases} + N_{i-q/standard}^{phrases} + N_{i-m-int/standard}^{phrases} + \\ &N_{i-m-val/standard}^{phrases} + N_{i-q-rin/standard}^{phrases} + N_{i-q-rab/standard}^{phrases} + N_{i-q-hnv/standard}^{phrases} \\ &= 842 + 196 + 662 + 148 + 20 + 96 + 168 + 408 = 2540 \end{aligned}$$

Indices moyens d'un corpus : les indices moyens (Précision³ moyenne ou Rappel moyen ou F-mesure moyenne) doivent être pondérés, car certaines phrases sont annotées avec plusieurs catégories d'incertitude. Un indice moyen est la somme des indices de même nature pondérés pour chaque catégorie d'incertitude du corpus, divisée par $S_{corpus}^{phrases}$. Par exemple, la Précision moyenne d'un corpus est la somme des Précisions pondérés pour chaque catégorie d'incertitude du corpus, divisée par $S_{corpus}^{phrases}$:

$$\begin{aligned} P_{corpus}^{moy} &= \left((Indice_{i/corpus} \times N_{i/corpus}^{phrases}) + (Indice_{i-m/corpus} \times N_{i-m/corpus}^{phrases}) + (Indice_{i-q/corpus} \times N_{i-q/corpus}^{phrases}) + (Indice_{i-m-int/corpus} \times N_{i-m-int/corpus}^{phrases}) \right. \\ &+ (Indice_{i-m-val/corpus} \times N_{i-m-val/corpus}^{phrases}) + (Indice_{i-q-rin/corpus} \times N_{i-q-rin/corpus}^{phrases}) + (Indice_{i-q-rab/corpus} \times N_{i-q-rab/corpus}^{phrases}) \\ &\left. + (Indice_{i-q-hnv/corpus} \times N_{i-q-hnv/corpus}^{phrases}) \right) \div S_{corpus}^{phrases} \end{aligned}$$

Par exemple, la Précision moyenne pour le gold standard est :

$$\begin{aligned} P_{standard}^{moy} &= \left((Précision_{i/standard} \times N_{i/standard}^{phrases}) + (Précision_{i-m/standard} \times N_{i-m/standard}^{phrases}) + (Précision_{i-q/standard} \times N_{i-q/standard}^{phrases}) \right. \\ &+ (Précision_{i-m-int/standard} \times N_{i-m-int/standard}^{phrases}) + (Précision_{i-m-val/standard} \times N_{i-m-val/standard}^{phrases}) \\ &+ (Précision_{i-q-rin/standard} \times N_{i-q-rin/standard}^{phrases}) + (Précision_{i-q-rab/standard} \times N_{i-q-rab/standard}^{phrases}) \\ &\left. + (Précision_{i-q-hnv/standard} \times N_{i-q-hnv/standard}^{phrases}) \right) \div S_{standard}^{phrases} \end{aligned}$$

³Nota bene : dans les résultats du corpus d'évaluation, l'absence de vrais positives et de fausses positives pour la catégorie *i-m-val* empêche de calculer la Précision et la F-mesure pour cette catégorie. La solution adoptée pour obtenir la Précision moyenne et la F-mesure moyenne du corpus est de les calculer sur le sous-ensemble du corpus composé par les autres catégories.

$$= \left((0,981 \times 842) + (0,831 \times 196) + (0,965 \times 662) + (0,829 \times 148) + \right. \\ \left. (1,000 \times 20) + (0,837 \times 96) + (0,979 \times 168) + (0,914 \times 408) \right) \div 2540 = 0,940$$

Annexe E

Algorithme d'annotation : exécution des règles linguistiques sur des corpus de phrases

```

1  # coding: utf8
2
3  # =====
4  # ADIS - Annotateur de degrés d'incertitude scientifique.
5  #
6  # Copyright : Centre Tesnière du CRIT de l'Université de Franche-Comté 2021.
7  # =====
8  """
9  Prend en entrée :
10     - un corpus de phrases :
11         "phrases.csv"
12     - une liste de regles linguistiques de degrés de l'incertitude :
13         "regles.csv"
14
15  Renvoie en sortie :
16     - la liste d'annotation sur le degré d'incertitude de chaque phrase :
17         "resultat_annotations.csv"
18     - les indicateurs de Précision et de Rappel :
19         "resultat-eval.txt"
20  """
21
22  import re
23  import csv
24  import io
25
26  #-----
27  # Classes et fonctions
28  #-----
29
30  def traduireELing2EReg(exprLing):
31      """Traduit une expression linguistique en une expression régulière.
32      1. Prend en entrée une expression linguistique.
33      2. Retourne l'expression régulière qui lui correspond."""
34
35      if exprLing == False:
36          return False
37
38      if exprLing == "":
39          return False
40
41      # Enlever d'éventuels espaces au début et fin de l'exprLing
42      exprLing = exprLing.strip()
43      # "^" -> "\b" pour prise en compte des début ou fin de mots, par ex. les points
44      exprLing = exprLing.replace("^", "\\b")
45      # "?" -> "\?" pour prise en compte des caractères '?' en fin de phrase
46      exprLing = exprLing.replace("?", "\\?")
47      # (...) -> (...) ?
48      exprLing = exprLing.replace(")", ")?")
49      # [...]/.../... -> (...|...|...)
50      exprLing = exprLing.replace("[", "(")
51      exprLing = exprLing.replace("]", ")")
52      exprLing = exprLing.replace("/", "|")
53      # _ -> [A-Za-z\\-]+
54      exprLing = exprLing.replace("_", "[A-Za-z\\-]+")
55      # "#" -> n'importe quel nombre anglais ou français, positif ou négatif
56      # Ex. : 12.4 ou 12,4 ou +230 ou 65.214,58 ou -153,485,234.30
57      exprLing = exprLing.replace("#", "[+\\-]?[0-9][0-9,.]*")
58      # ... -> .+
59      # Nota bene : "..." est un seul caractère, différent de "..."
60      exprLing = exprLing.replace("...", "...")
61      exprLing = exprLing.replace("...", ".+")
62      # (.+)? suivi d'un espace -> enlever l'espace
63      exprLing = exprLing.replace("(.)? ", "(.)?") # ???
64      # (.+)? precede d'un espace -> enlever l'espace
65      exprLing = exprLing.replace(" (.+)?", "(.)?") # ???
66
67      return exprLing      #Essayé : return "\b"+ exprLing +"\b" # == exprReg
68
69  class Annotation :

```

```

70     def __init__(self, cat, regle):
71         self.cat = cat
72         self.regle = regle
73
74     class Regle :
75         """Une regle linguistique comporte :
76         - une id composée de 3 lettres d'abréviation de la catégorie de degré
77         d'incertitude correspondante, suivie d'un numéro,
78         séparés par un "_".
79         Ex. : "HNV_0002"
80         - une expression linguistique positive :
81         elle décrit les expressions d'incertitude pour pouvoir
82         les détecter dans un texte.
83         Ex. : HNV_0002_ELin = "...".
84         - une expression linguistique négative :
85         elle décrit les expressions qui font exception aux précédentes.
86         Elle détecte des expressions qui ne doivent pas être confondues.
87         Ex. : HNV_0002_ELinNeg = "...".
88         - le nom de la catégorie
89         """
90
91     def __init__(self, id, cat, expr, expr_neg=False) :
92         self.id = id
93         self.expr = expr
94         self.expr_neg = expr_neg
95         self.cat = cat
96         self.traduireEnEReg()
97
98     # methode qui ajoute les expr. reg. dans l'objet
99     def traduireEnEReg(self):
100         self.ereg = traduireELing2EReg(self.expr)
101         self.ereg_neg = traduireELing2EReg(self.expr_neg)
102
103     # methode qui verifie si la phrase "match" la regle
104     def verifierPhrase(self, phrase) :
105         if (self.ereg_neg == False) :
106             if (re.search(self.ereg, phrase)) :
107                 return True
108             else:
109                 return False
110         else :
111             if (re.search(self.ereg_neg, phrase)):
112                 return False
113             else :
114                 if (re.search(self.ereg, phrase)) :
115                     return True
116                 else : return False
117
118     def afficher(self):
119         return self.id + " " + self.cat + " --> " + self.expr + " | " +
120         str(self.expr_neg)
121
122     def afficher_ereg(self):
123         return self.ereg + " | " + str(self.ereg_neg)
124
125     class ListeDeRegles:
126         """Une liste d'objets de type Regle.
127         Les regles servent à détecter les catégories d'incertitude,
128         en matchant les phrases qui les contiennent."""
129
130         # Création de la liste à partir d'un fichier CSV :
131         def __init__(self, filename):
132             self.lst = []
133             with open(filename, encoding='utf-8') as csvfile:
134                 readerF = csv.reader(csvfile, delimiter=',', quotechar='"')
135                 for row in readerF:
136                     if len(row) == 0 :
137                         continue

```

```

138         id = row[0].strip()
139         if (id[0] == "#") :
140             continue
141         cat = row[1].strip()
142         expr = row[2].strip().strip('')
143         neg = False
144         if (len(row)>3):
145             neg = row[3].strip().strip('')
146         f = Regle(id, cat, expr, neg)
147         self.lst.append(f)
148
149     # Méthode produisant la liste des étiquettes qui correspondent à une phrase.
150     def annoterPhrase(self, p):
151         lst_annots = []
152
153         for f in self.lst:
154             if ( f.verifierPhrase(p) ):
155                 lst_annots.append( Annotation(f.cat, f) )
156
157         return lst_annots
158
159     # Ajout des categories selon la hierarchie :
160     def ajoutHierarchie(categories):
161         if ("i-m-int" in categories) and ("i-m" not in categories) :
162             categories.append("i-m")
163         if ("i-m-val" in categories) and ("i-m" not in categories) :
164             categories.append("i-m")
165         if ("i-q-rin" in categories) and ("i-q" not in categories) :
166             categories.append("i-q")
167         if ("i-q-rab" in categories) and ("i-q" not in categories) :
168             categories.append("i-q")
169         if ("i-q-hnv" in categories) and ("i-q" not in categories) :
170             categories.append("i-q")
171         if ("i-m" in categories) and ("i" not in categories) : categories.append("i")
172         if ("i-q" in categories) and ("i" not in categories) : categories.append("i")
173         return categories
174
175     # =====
176     # PROGRAMME PRINCIPAL
177     # =====
178
179     fichier_phrases = "phrases.csv"
180     fichier_regles = "regles.csv"
181     fichier_resultat = "resultat_annotations.csv"
182
183     # Lecture des phrases :
184     phrases = []
185     annotations_prevues = []
186
187     with open(fichier_phrases, encoding='utf-8') as csvfile:
188         readerP = csv.reader(csvfile, delimiter=',', quotechar='')
189         for row in readerP:
190             if len(row) == 0 :
191                 continue
192             phrases.append(row[0].strip().strip('').strip())
193             # row[1] fonctionne avec 2e colonne dans phrases.csv
194             # Le string de row[1] du genre "i-q" ou "i", "i-q" est transformé en
195             liste
196             listRaw1 = []
197             if row[1].count("") == 2 :
198                 listRaw1.append(row[1].strip('').strip(''))
199             elif row[1].count("") > 2 :
200                 listRaw1 = list(eval(row[1])) # eval() seul donnerait un tuple
201             annotations_prevues.append(listRaw1)
202
203     # Ajout des categories selon la hierarchie :
204     for a in annotations_prevues :
205         a = ajoutHierarchie(a)

```

```

201 # Lecture des regles :
202 listeRegles = ListeDeRegles(fichier_regles)
203
204 # Annotation des phrases et export du resultat dans un fichier csv
205 f_resultat = io.open(fichier_resultat, "w", encoding="utf-8")
206 f_resultat.write("num_phrase,phrase,annotations_cat_trouvees,regles_ayant_trouve_cat,")
207 f_resultat.write("annotations_cat_attendues,correspondance_entre_cat_trouvee_et_attendues\n")
208
209 cnt = 1
210 annotations_phrases = []
211 for p in phrases:
212     print ("Phrase " + str(cnt))
213     annotations = listeRegles.annoterPhrase(p)
214
215     categories = []
216     for a in annotations :
217         if (a.cat not in categories) :
218             categories.append(a.cat)
219     trace_annotations = []
220     for a in annotations :
221         trace_annotations.append(a.regle.id + "-" + a.regle.cat)
222     # -----
223     # Ajout des categories selon la hierarchie :
224     categories = ajoutHierarchie(categories)
225     annotations_phrases.append(categories)
226
227     f_resultat.write(str(cnt) + ", \"" + p + "\", \"" + str(categories) + "\", \"")
228     f_resultat.write(str(trace_annotations) + "\", \"\" +
229 str(annotations_prevues[cnt-1]) + "\"\n")
230     cnt += 1
231 f_resultat.close()

```


Annexe F

Règles d'annotation de l'incertitude

ID	Catégorie	Partie positive	Partie négative
98	i-q-hnv	[[C/c]onsequence(s)/[R/r]esult(s)] [might/could] be [the/that/a(n)] ...	
222	i-q-hnv	[[may/might/could] be/is/are/[remain(s)/appear(s)/show(s) signs to be]] (_) [unlikely/improbable/dubious/questionable/not [likely/credible/verisimilar/presumable]]	
137	i-q-hnv	[[may/might/could](be)/is/are/will be] unlikely to ...	
215	i-q-hnv	[[may/might/could/would] be/is/are] ([more/less/now])likely to	the possibility that ... may be ... means
114	i-q-hnv	[A/a]lthough ... [suggest(s)/[is/are] suggesting] that ... [may/might/could] ...	[A/a]lthough ... [is/are/do(es)][not/n['/']t] suggest(ing) that ... [may/might/could] ...
238	i-q-hnv	[appears/seems/[creates/gives] the impression/shows signs] to be^	
237	i-q-hnv	[B/b]efore any (_)conclusion can be drawn	
172	i-q-hnv	[believe/consider/accredit/think/trust/credit/account/esteem/view] that... [may/might/could] be^	... [do not/don['/']t] [believe/consider/accredit/think/trust/credit/account/esteem/view] that ... [may/might/could] be ...
89	i-q-hnv	[E/e]xperiments(under way) [[may/might/could] be able to/[might/could]] [demonstrate/show/prove] ...	
175	i-q-hnv	[F/f]urther investigation is required	[N/n]o further investigation is required
201	i-q-hnv	[finding/result](s) [provide(s) hope/([are/is])suggest(ing)] ...	
51	i-q-hnv	[However/Nevertheless/Instead], ... [may/might/could] be ..., [since/due to/because]	
49	i-q-hnv	[However/Nevertheless/Instead], ... [may/might/could](_) be(_)_er	[However/Nevertheless/Instead], ... [may/might/could] not be(_)_er
44	i-q-hnv	[However/Nevertheless/Instead], ... suggest([ing/s]) that ... [may/might/could] be ...	[However/Nevertheless/Instead], the ... [do(es) not/do(es)n['/']t] suggest that ... [may/might/could] be ...
48	i-q-hnv	[However/Nevertheless/Instead], this [assumption/belief/expectation/guess/hypothesis/inference/(pre)supposition/theory/postulate/idea] [may/might/could] be ...	
42	i-q-hnv	[I/i]f ... [can/could] be (_)([at]tested/evidenced/prove[d/n]/show[n/ed]/demonstrated/checked(out)/verified/established/confirmed/validated/borne out/measured/identified/authenticated/witnessed/corroborated/backed up/vindicated](...), [then/therefore/accordingly/thus/consequently] ... [may/might/could] be ...	
221	i-q-hnv	[I/i]f ... [may/might/could/would] [probably/pehaps/possibly/plausibly/presumably/seemingly/most likely] be _ed	
140	i-q-hnv	[I/i]f ... we [could/may/might](...) determine [how/(in)what/when/why/where/who/(in)which]	[I/i]f ... we [could/may][not/n['/']t](...) determine [how/(in)what/when/why/where/who/(in)which]
1	i-q-hnv	[I/i]f ..., ... may be(not) ...	[Thus, ... may/[I/i]f ..., (...)decision(s) [may/can] be/[I/i]f ..., but(...) may be ... if/then(...) may be(_)(,) if .../may be _ed(...) in order to/If there is ..., ... may be _ed^]
117	i-q-hnv	[I/i]t [can/could/may/might] be assumed that ... [could/might] be ...	[I/i]t [can/could/may/might] be assumed that ... [could/might] be(...) responsible for^
106	i-q-hnv	[imply(e)s/[is/are] implying] that [the/a(n)] ... [may/might/could] ...	[imply(e)s/[is/are] implying] that [the/a(n)] ... [may/could][not/n['/']t] ...
99	i-q-hnv	[indicat(e)/show/point/express(e)/attest/denot(e)/demonstrat(e)/impl([i/y])(e)/mean][s/ing/ed/d](out)(_) that there [might/could] be ...	
155	i-q-hnv	[indicated/showed] that ... [may/might/could] be(_)_ed with [indicated/showed] that ... [may/might/could] be _er [[related/linked/attributed]/[caused/explained] by] ...

192	i-q-hnv	[Is/Are] ... ?	
243	i-q-hnv	[is/are] [a(n) (_)possibility/possible]	[[is/are] [a(n) (_)possibility/possible] ... [[were/was]/ could be from/lead(s) to^]/[S/s]o/[T/t]hus) it is possible that .../[I/i]t is possible that this occurred/... [will/ would/may/might/can/could/shall/should] ... future/ Since ... [is/are/was/were/as been/have been/[may/ might/can/could/should] be] ..., it is possible that ...]
107	i-q-hnv	[is/are] [believed/considered] to be ...	
93	i-q-hnv	[is/are] believed to ..., but ...	
92	i-q-hnv	[is/are] believed to be [_ ing^/_ed^/made/bro- ken/gone/left/lost/seen]	
94	i-q-hnv	[is/are] believed to have _ed	
224	i-q-hnv	[is/are] likely to result in a(n) ...	
209	i-q-hnv	[may/might/could] ... in determining ...	[... [may/could][not/n[']t] ... in determining .../ provide(s) (_)(_)[hint/clue/indication/suggestion/ink- ling](s) that ... [may^/might/could/would]]
2	i-q-hnv	[might/could] be (_)[_ ing^/_ed^/made/ broken/gone/left/lost/seen]	[[might/could] be [called/expected/defined/gained/ex- plained by/biased by/linked to^/related to^/associated(_) with^/caused by^/_ (_)(_)directed/... because/the most likely cause(s)]/which [might/could] be^/[T/t]he/ [A/a](n)/[O/o]ne/[A/a]nother)(_)(_)([possible/plau- sible/conceivable/credible/probable/creditable/presu- mable/supposable]) [factor/reason/explanation/ motive/motivation/cause/mechanism/causation/de- terminant/origin/provenance](s)(...) [might/could] be ...]
199	i-q-hnv	[may/might/could] be affected by ...	[Thus,/illustrate(s) how] ... may
214	i-q-hnv	[might/could] be expected to [might/could] be expected to ... [in the future, as/ due to] ...
248	i-q-hnv	[might/could/can] result^	[can/could/might] result from ...
122	i-q-hnv	[might/could/can] be a(n) ...	
253	i-q-hnv	[might/could/can] be ... if ...	[[W/w]hen ... [might/could/can] be ... if .../[T/t]he/[A/a] (n)/[O/o]ne/[A/a]nother)(_)(_)([possible/plausible/ conceivable/credible/probable/creditable/ presumable/supposable/likely]) [factor/reason/expla- nation/motive/motivation/cause/causation/determi- nant/origin/provenance](s)(...) [might/could/can] be ... if ...]
206	i-q-hnv	[may/might/could/can] be regarded as ([highly/poorly/rather/more/less]) [susceptible/ prone] to ...	
158	i-q-hnv	[may/might/could/can] potentially ...	
47	i-q-hnv	[R/r]esult(s) [provide(s)/[is/are] providing] [hints/clues] that ... [may/might/could]	provide(s) (_)(_)[hint/clue/indication/suggestion/ink- ling](s) that ... [may^/might/could/would]
40	i-q-hnv	[R/r]esult(s) suggest(s) that the... [may/might/ could]	
103	i-q-hnv	[suggest(s)/[is/are]suggesting] ... [that/which] [may/might/could]	suggest(s) ... that [may/might/could] be driven by ...
45	i-q-hnv	[suggest/propos(e)/advanc(e)]([s/ing]) that ... [may/might/could] ...	[[do(es) not/do(es)n[']t] suggest that .../Given the ..., ... suggest that ... [may/might/could]/ suggest([s/ing]) that ... [may/might/could] [be [due to/ [related/linked] [to/with]/based on/caused by]/result in^]/suggest(s) that ... factor(s)(, or ...) [may/might/ could] be at work/[may/might/could] [be/have/come from] [this/taht/an()other] (_)(_)(_)[cause/origin/ source]]
183	i-q-hnv	[suggest/propos(e)/advanc(e)]([s/ing/ed/d]) a possible ...	
63	i-q-hnv	[T/t]he evaluated results(...) [may/might/could] be _er than ...	
50	i-q-hnv	[T/t]he potential impact of ... [may/might/could] be	

3	i-q-hnv	[T/t]here [[may/could/might] be/are/is]([still/ however]) ([a(n)/the/some/more/less/a few]) [uncertain[y/ies]/risk(s) tha[t/n]]	
46	i-q-hnv	[T/t]here is [the possibility/no certainty](...) that ...	
31	i-q-hnv	[T/t]herefore(.) ... [might/could] be ...	[Therefore,(...) there will be .../Therefore, ... [might/could] be ... for .../[might/could] be from ...]
124	i-q-hnv	[technique(s)/result(s)/conclusion(s)] (...)may be questioned	
210	i-q-hnv	[that/which] [may/might/could] be [_ed^/_ing^]	[[provide(s)/for] ... [that/which] [may/might/could] be [_ing^/_ed^]/may be [_ingly/interpreted as/used to]/ collected(...) on ... that may be/[A/a/[T/t]he](...) me- tric(...) is ..., which may be expressed/[P/p]ossible examples of ... [include/contain/comprise/ encompass/are] .../... [is/are] _ed [to/by/from/at/on/ under/before/behind/in/out of/] ... that may ... might ...]
250	i-q-hnv	[W/w]e [guess/suppose/presuppose/presume/ conjecture/hypothesise/speculate](true) that ...	
5	i-q-hnv	[W/w]e believe ...	[[[W/w]e believe ... tend([s/ing]) to ...]/[A/a]s we be- lieve ...]/[Although ... we believe]/[Based on ... results, we believe]/[[One/one/Some/some/Several/several] of/Among(st)/among(st)] the main reasons(.) we be- lieve]
32	i-q-hnv	[W/w]e hypothesise that ...	
96	i-q-hnv	[W/w]hile ... [could/might] ..., ... [may/might/ could]	[W/w]hile ... [could/might][not/n[']t] ..., ... [may/could] [not/n[']t]
216	i-q-hnv	[will/would/shall/should/could] [determine/ show/inform] [whether/if] [[the/a/an] .../it/there] [are/is/were]	
205	i-q-hnv	^[(remaining)possibility is/it [is/[may/might/ can/could] be] (_)possible/there [exists/is](_) the possibility](_) that ... [may/might/could]	[might/could] be (originated)from ...
154	i-q-hnv	^[B/b]ut the ... suggest(s)(that) this [may/ might/can/could/] be evident ...	
59	i-q-hnv	^[I/i]t [might/could] be that ... will	
249	i-q-hnv	^[I/i]t is ([however/nevertheless/though]) ([very/quite/rather]) unlikely	
130	i-q-hnv	^[I/i]t is possible that the(...) value for ...	
95	i-q-hnv	According to ..., ... [may/might/could]	According to ..., ... [may/could][not/n[']t]
4	i-q-hnv	and (_)[may/might/could] be [relevant/important] to ...	
142	i-q-hnv	Because ... [may/might/could] be very different from ...	
143	i-q-hnv	believe that ... [could/may/might] [neverthe- less/however] ...	[do not/don[']t] believe that ...
150	i-q-hnv	Given [predicted/foreseen] ..., ... [may/might/ could] ...	
33	i-q-hnv	However, ... [may/might/could] be [[under/ over]estimated/(un)significant]	
152	i-q-hnv	In this context, the ... [might/could] be ... when ...	
231	i-q-hnv	require(s) further investigation(s)	do(es)[not/n[']t] require(s) further investigation(s)
233	i-q-hnv	result(s) [indicate/suggest/show/point](s) that ... [might/could]	
251	i-q-hnv	[suspect([s/ing]) that/^[is/are] thought to^]	
173	i-q-hnv	the possibility that... [may/might/could]	[[T/t]he possibility that ... [may/might/could] ... means/introduces the possibility that ... [may/might/ could]]
208	i-q-hnv	whether ... [may/might/can/could] be ...	refer(s) to whether ... [may/might/can/could] be ...

254	i-q-hnv	... may be ...	<p>[[may be (_)(_)(_)(_)][abandoned(_) when^/addressed(_) by^/explained(_) by^/biased(_) by^/driven(_) by^/favoured(_) by^/caused(_) by^/at work/applied/expected/plausible/_ explained by^/based(_) on^/computed/divided/considered/linked(_) to^/associated(_) with^/determined/defined(_) as^/defended(_) as^/emitted up(_) to^/related(_) to^/du_(_) to^/attributed[able/ed](_) to^/estimated/expressed(_)in^/evaluated/originated(_) from/gathered(_) from/interpreted(_) as^/modeled/realised/removed/represented(_) as^/rewritten/selected/seen(_) as^/specified(_) by/shown(_) as^/taken(_) as^/thought of(_) as^/understood(_) as^/used]]/[may be ... but ... can be^/may be contributory/may be a reflection of^/may be either ... or .../By implication, ... because ... may be _ed/Be-cause ... may be .../depending on .../where ... may be .../the possibility that ... may be ... means/[A/a]though ... may be [_ed/_er]/can be ... for ... that may be _ed/it is well-known that ... may be .../stipulates ... that ... may be set^/refers to whether ... may be reversed/only ... may be _ing/can be viewed as ... that may be .../Where(...,) ... may be .../illustrat_ that ... may be .../which may be ... under ... conditions/as may be expected().../to estimate what ... may become .../While ... may be ruled .../While ... may believe that ..., .../whereas ... may be^/Data may be assumed to be from .../how ... may be used [to/by] .../illustrate how ... may be _ed .../No ... may be _ed^/the number of times ... may be reused .../More details on ... may be found in .../One reason that ... may be... is because .../[The/A(n)] ... may be ... as long as .../to the exclusion of ... that may be^/suggest([s/ing]) ... may be a(n)(_)(_) [mechanism/function]/[over/about] the ... [they/we] may be _ to/around #()% may be/[[T/t]he/[A/a](n)/[O/o]ne/[A/a]nother](_)(_) [possible/plausible/conceivable/credible/probable/creditable/presumable/supposable/likely]] [factor/reason/explanation/motive/motivation/cause/causation/determinant/origin/provenance](s)(...) may be .../[that/which] may be (_) [related/expressed in^]/may be responsible for .../may be [the/a(n)/one/another](_)(_) [possible/plausible/conceivable/credible/probable/creditable/presumable/supposable/likely]] [factor/reason/explanation/motive/motivation/cause/causation/determinant/origin/provenance](s)/[may be/might be/could be/seem(s)](_)(_) [(im)possible/probable/likely] to(be) achieve(d)/Given the ..., ... suggest that ... may be .../[T/t]herefore, ... may be [a(n)/the/another] _ (_) [technique/solution/answer/procedure/process/approach/system/tactic/strategy/way/method(ology)] for .../Therefore, ... may be disconnected/[l/i]n the future(,) ... may _/effect(s) may be that .../[l/i]t/factor/reason/explanation/motive/motivation/cause/mechanism/causation/determinant/origin/provenance](s) may be that [a(n)/the] ...]]</p>
-----	---------	----------------	--

255	i-m-val	<p>[... may be # .../^[a(n)/the] #()[%/percent] (of) [(un)certainity/(im)probability/(non)confidence/ reliability/credibility/(dis)trust(worthiness)/plau- sibility/possibility/(un)likelihood/(un)likeliness/ (in)certitude/doubt(lessness)/positiveness/sur- eness/surety/indubitableness/assuredness/ conviction/chance]/[a(n)/the] [(un)certainity/ (im)probability/(non)confidence/reliability/cre- dibility/(dis)trust(worthiness)/plausibility/possi- bility/(un)likelihood/(un)likeliness/(in)certitude/ doubt(lessness)/positiveness/sureness/surety/ indubitableness/assuredness/conviction/ chance]([degree/level/coefficient/factor] [of/at] #[degree/level/coefficient/factor] of [(un)certainity/(im)probability/(non)confidence/ reliability/credibility/(dis)trust(worthiness)/plau- sibility/possibility/(un)likelihood/(un)likeliness/ (in)certitude/doubt(lessness)/positiveness/sur- eness/surety/indubitableness/assuredness/ conviction/chance] [of/at] #[^()(%/percent)]/ ^within([a/the] #()(%/percent)/[estimate/ guesstimate/evaluate](s) that ... [comprise/ contain/compose/constitute/form](s) #()(%/ percent)]</p>	<p>[[[A/a]]R/r]ound /A/a]bout /A/a]pproximately /O/ o]ver /B/b]y /N/n]etx to /N/n]igh /c[C/c]ose to /a/A/ a]proaching /c[C/i]rca /a[A/a]most /r[R/o]ughly /n[N/ n]arly /±()#()% of [(un)certainity/(im)probability/ (non)confidence/reliability/credibility/ (dis)trust(worthiness)/plausibility/possibility/ (un)likelihood/(un)likeliness/(in)certitude/ doubt(lessness)/positiveness/sureness/surety/indubi- tableness/assuredness/conviction/chance]</p>
256	i-q-rab	<p>[[might/could /seem([s/ing]) to] be (originated(_)from .../[may/might/could/ seem([s/ing]) to^] be (_)(_)(_) [explained(_)by^/biased(_)by^/driven(_) by^/at work/favoured(_)by/caused(_)by/as- sociated(_)with^/linked(_)to^/related(_) to^/due(_)to^/attribut[able/ed](_)to^/contribu- tory/a reflection of^] .../[T/t]he/[A/a](n)/[O/ o]ne/[A/a]nother](_)(_)([possible/plausible/ conceivable/credible/probable/creditable/pre- sumable/supposable] [factor/reason/explana- tion/motive/motivation/cause/mechanism/cau- sation/determinant/origin/provenance](s) (...) [may/might/could/seem([s/ing]) to] be .../ [that/which] [may/might/could/seem([s/ing]) to] be (_)related/[may/might/could/seem([s/ing]) to] be responsible for .../[might/could/seem([s/ ing]) to] be explained by/[may/might/could/ seem([s/ing]) to] be [the/a(n)/one/another](_) (_)([possible/plausible/conceivable/credible/ probable/creditable/presumable/supposable/li- kely] [factor/reason/explanation/motive/moti- vation/cause/causation/determinant/origin/pro- venance](s)/[is/are/seem(s)/appear(s)](_)(_) plausible/suggest([s/ing]) ... may be a(n)(_) (_) [mechanism/function]/[[l/i]t/factor/reason/ explanation/motive/motivation/cause/mecha- nism/causation/determinant/origin/provenance] (s) [may/might/could] be that [a(n)/the] .../ [may/might/could] be plausible/[may/might/ could] relate to^/provide(s) (_)(_) [hint/clue/in- dication/suggestion/inkling](s) that ... [may^/ might/could/would]/[may/might/could] [be/ have/come from] [this/taht/an()other] (_)(_)(_) [cause/origin/source]/[S/s]everal/[M/m]any/ A/a](n) ... number of] [factors/causes] may be(...) responsible for .../suggesting that ... did ...]</p>	<p>[stipulate(s) (...)that ... may be .../... [factor/reason/ explanation/motive/motivation/cause/mechanism/cau- sation/determinant/origin/provenance](s) (...)that ... may be ... [is/are] because .../^is [explained(_)by^/ biased(_)by^/driven(_)by^/at work/favoured(_)by/ caused(_)by/associated(_)with^/linked(_)to^/rela- ted(_)to^/due(_)to^/attribut[able/ed](_)to^/contribu- tory/a reflection of^] .../may be shown as follows/ ^so ... that ... may be [explained(_)by^/biased(_) by^/driven(_)by^/at work/favoured(_)by/caused(_) by/associated(_)with^/linked(_)to^/related(_)to^/ due(_)to^/attribut[able/ed]/expected to^] .../ [whether/[A/a]lthough] ... may be^]</p>

257	i-m-int	[[([A/a])[R/r]ound /([A/a]bout /([A/a]pproximately /([O/o]ver /([B/b]y /([N/n]etx to / [N/n]igh /c[C/c]lose to /a[A/a]proaching /c[C/] i)rca /a[A/a]most /r[R/o]ughly /n[N/n]arly /± ()#/#between #()(...) and #/estimate of #/[[- more/less] than/above/under] #]	estimate of [a/the] ...
258	i-q-rin	[[may be/might be/could be/seem(s)]()() [(im)possible/probable/likely] to(be) achieve(d)/Given the ..., ... suggest that ... [may/might/can/could/will/would] .../[T/ t]herefore, ... [may/might/could] be [a(n)/the/ another] _() [technique/solution/answer/pro- cedure/process/approach/system/tactic/strate- gy/way/method(ology)] for .../[may/might/ could] be expected/[I/i]n the future(.) ... [may/might/could] _/effect(s) [may/might/ could/should] be that .../would lead to [a(n)/ the]^[may/might/could] [become/generate/ produce/cause/create/engender/induce/result in/provoque/give [birth/rise] to]/[may/might] not [generate/produce/cause/create/engender/in- duce/result in/provoque/give [birth/rise] to]/^it is [possible/probable/likely] that ... [would/ lead(s) to/become]/[S/s]ince ... ^it is [possible/probable/likely] that/Consequently, ... might be .../that(..., and) might be ()_ for]	[^was^(...), as may be expected/rather than()()() [that/which] [might/could] be expected/[is/are] [that/ what]/what] ... [may/might/could] become]
259	i-m	estimate of [a/the] ...	

Annexe G

Corpus de phrases annotées - *gold standard*

Ici nous présentons le corpus annoté en trois parties : corpus d'étude (parties 1 et 2), et corpus d'évaluation. Dans le corpus d'étude, trois phrases ont été enlevées parce qu'elles étaient des doublons.

G.1 Corpus d'étude - partie 1 (500 phrases)

ID	Catégories	Phrase
1	'i-q-hnv'	Small-scale fires which mimic the patchiness of the formerly uncleared landscape at a smaller scale within reserves may be a management option for plant populations in isolated fragments.
2	'ai'	By implication, this places an increasing onus on businesses with development projects to “do the right thing” because government capacity may be limited (Zubair et al. 2011).
3	'i-q-hnv'	A baseline may be adjusted after the project has been in place for some years, if a GHG program dictates the validity period and the conditions that trigger review or revision of baseline driver variable assumptions.
4	'ai'	In selecting the environmental issues to consider, broader stakeholders can be engaged (and may be legally required to depending on local legislative requirements).
5	'i-q-hnv'	Basic links may be made at this stage between the issues, potential project activities, likely impact and potential alternatives to be considered.
6	'i-q-hnv'	If the risks associated with the project are considered too high, the final decision may be not to proceed.
7	'i-q-hnv'	So while project monitoring may be a normal part of business practice, it seems this is not translating into the EIA process for businesses, or governments, as needed.
8	'i-q-hnv'	Conversely, government bodies operate diverse non-related projects, and review projects which may be strikingly different.
9	'ai'	This, we believe, is therefore an aspect worthy of future research and is captured by Proposition 2.
10	'ai'	We have aimed to address this shortfall by developing an EIA that accommodates government and legal requirements, but which also captures the reality of business practices (or what they might be, ideally).
11	'i-q-hnv'	Hence, RegCM output may be useful to detect 3–6-month (monthly to seasonal) length moderate drought events over a heavy rainfall region likely Bangladesh.
12	'ai'	The final departure from traditional EIAs emerges in our proposed business EIA as we believe that the feed-forward loop is a critical element concerning businesses.
13	'i-q-hnv'	Ecosystems The NECT may be divided into three sections of vegetation patterns and edaphic types from west to east (Editorial Committee for Vegetation of China, 1980; Zhang et al., 1997), i.e.
14	'i-q-hnv'	Although the spatial resolution will lead to many errors and inconsistencies in local or regional scale, for instance the land elevation of <1 m may be considered sea by SRTM and has no data, it is satisfactory for determining vulnerable coastal areas on the national scale.
15	'i-q-hnv'	Coastal slope for example is largely temporally static, land use does change but quite slowly whereas sea level rise may be more dynamic.
16	'ai'	Although this is only a relative quantifiable method, which cannot directly express particular physical processes and effects, it can be used to identify the regions where the various effects of sea-level rise may be the greatest.
17	'i-q-hnv'	may be significant in environments where O ₂ is not limiting (Patureau et al. 2000) because many bacteria isolated from soils and sediment (including <i>Pseudomonas</i> , <i>Aeromonas</i> and <i>Moraxella</i> genera) are capable of nitrate respiration in the presence of O ₂ (Carter et al. 1995; Patureau et al. 2000).
18	'ai'	However, the possibility that some of the heterotrophic nitrification may be inhibited by C ₂ H ₂ means it is more likely to be underestimated than overestimated.
19	'ai'	Although root growth may be limited, a pot experiment was chosen for the present study due to the fact that the field-scale application of ¹⁵ N-labeled fertilizer contaminates soil with ¹⁵ N and thus does not allow a series of experiments to be performed.
20	'i-q-hnv'	Even though, detection of drought case or frequency may be helpful for long-term planning purposes with some limitation regarding detection of extreme drought cases.
21	'i-q-hnv'	There is a close relationship between pan evaporation and actual evaporation but they can not simply replace each other, and even they may be completely adverse to each other, especially in semi-arid and arid areas.
22	'ai'	Since species distributions also reflect interspecific interactions between prey and predator populations, these potential changes in predator–prey relationships can be relevant for understanding species distribution shifts that may be observed with climate change (Davis et al. 1998).
23	'i-q-hnv'	Moreover, most of the fish species studied can be classified as specialist feeders and adaptations may be clearly necessary if changes to benthic biological community continue in the north Bering Sea.
24	'i-q-hnv'	Dryland tillage and dry seeding avoid the preparation of seedling nursery, which may be important methane sources and data on methane fluxes from it are lacking.
25	'i-q-hnv'	Dry-seeded rice may be a mitigation option in the temperate zone where water is limited.
26	'ai'	In fact, it is well-known that some plant species and certain vegetation types may be able to reach and survive along the nival belt, close to the glacial forelands and fronts (Erschbamer 2007).
27	'i-q-hnv'	Prospects for regulation and legislation are low for the entire range of fields, even if regulations may be developed for various sectors of emerging technologies as the science and applications are better defined.
28	'i-q-hnv'	The most familiar example is a computer operating system, on which plug-and-play software programs may be designed to run; sometimes many of those programs will not run well on a substantially redesigned version of that operating system, and radical improvement of the system could require redesign of each and every piece of software.
29	'ai'	One oversimplification that dominated past thinking was that public involvement in decisions about technological development was required only when unusual harm might be done, such as health hazards if certain potentially harmful particulate matter were released into the environment in substantial quantities.
30	'i-q-hnv'	Whether this case can withstand scrutiny or not (Baumol 1986; De Long 1988; Baumol and Wolff 1988), it may be prudent to find as many ways as we can to enable the whole set of stakeholders to participate in understanding as well as advancing a public convergence agenda that benefits as many stakeholders as possible, as much as possible, and as equally as possible (ETAG 2005).

31	'i-q-hnv'	It is important to determine the most appropriate technology; simple water pumps may be more useful than advanced filtration technologies (Musaazi 2012).
32	'i-q-hnv'	These studies may be a useful reference for researching typhoon landfall in the Yangtze River Delta.
33	'i-q-hnv'	First, in reference to modern typhoon weather phenomena and disaster characteristics, this study collected and classified records on strong winds, heavy rains, tidal surges, floods, and tsunamis that may be induced by typhoons.
34	'i-q-hnv'	Supposing we can find regional disaster chains about details of damage caused by a typhoon strike from records at the same time, a typhoon strike in this area may be inferred.
35	'i-q-hnv'	Thus, if a north or north-east wind is blowing, it may be affected by a typhoon.
36	'i-q-rab'	This result shows that fewer typhoon strikes than in normal years may be the reason for the aridity in the Yangtze River Delta.
37	'ai'	The STL procedure is an iterative process, which may be interpreted as a frequency filter directly applicable to non-stationary time series including harmonic components (Cleveland et al. 1990).
38	'i-q-hnv'	There is also a well-developed warm anomaly near 50°S, 120°W in both EOFs, which may be familiar to people working on the South Pacific (e.g. Kidson and Renwick 2002; Kwok and Comiso 2002).
39	'ai'	This exact correspondence between the respective first SST EOF patterns on one side and the first SST SVD patterns on the other side may be taken as
40	'ai'	Specifically, the anomalous cyclonic wind anomalies over the equatorial and tropical South Indian Ocean may be interpreted as a Rossby-type response to enhanced condensational heating over the maritime continent and the western equatorial Pacific (Lau and Nath 2000, 2003).
41	'i-q-hnv', 'i-q-rin', 'i-q-rab'	Now, these wind anomalies may be important for ENSO variability because the anomalous westerlies over the western Pacific may generate oceanic downwelling Kelvin waves which propagate eastward and may be responsible for the El Niño onset observed a few months after the occurrence of the positive dipole events in the South Atlantic and Indian Oceans, respectively (see Figs. 9, 11).
42	'i-q-hnv'	However, even when SSTs are high, as in the western Pacific warm pool, convection may be suppressed (or enhanced) by other factors such as upper level subsidence (ascendancy) generated by remote forcing (Lau et al. 1997).
43	'ai'	For the early austral summer period, the leading SLP and SST EOF or SVD modes in the three ocean basins of the Southern Hemisphere are significantly related to each other and may be considered as the regional components of a global circumpolar mode of coupled ocean-atmosphere variability.
44	'i-q-hnv'	We found that westerly wind anomalies over the western equatorial Pacific involved in El Niño onsets may be induced by the large-scale motion arising from differential heating of the atmosphere between the Pacific warm pool and the tropical Indian and Atlantic Oceans during late boreal winter and spring and not only the local SST (Lau et al. 1997).
45	'i-q-hnv'	However, for policy-makers, an aggregate environmental performance indicator may be preferable as it avoids any risk of information overload by presenting the information in a format tailored to decision-making" (Jollands et al. 2003).
46	'i-q-hnv'	It may be observed that Eqs.
47	'i-q-hnv'	Third, concerning the data limitation and seemingly little contribution to environment, some materials (e.g., aluminum, copper, limestone, tile, rubber, and paint) and stages (e.g., design stage) are omitted in the analyzed cycle, while proper design and material selection may be also critical to minimize energy consumption and CO2 emission.
48	'ai'	The underlying phenomenon might be called the "efficiency paradox," because it represents a case in which business firms, which are often presumed (or taken axiomatically) to be economically efficient, make decisions that do not maximize profits.
49	'ai'	Social policy designed to improve energy efficiency might be better directed at improving profitability by, for example, controlling health care or personnel turnover costs.
50	'ai'	(This might be called the "strong profit maximization assumption.") Hypotheses or conjectures cannot be scientific unless they are vulnerable to falsification by evidence (Popper, 1934, 1962).
51	'i-q-hnv'	The argument that potential energy-saving profits exist but are not the best unexploited opportunities available to firms may be true, but a prescription of policy passivity does not follow.
52	'i-q-hnv'	Similarly, the presence or absence of utility rebates is likely to influence the decisions of firms that might be eligible for the rebates.
53	'ai'	If uncertainty about energy prices, equipment costs, and technical change is felt uniformly throughout the market for lighting upgrades (which is a reasonable assumption), the prediction of uniformity of the returns is unchanged, although the expected level of the returns might be somewhat higher than it would be without taking the option component into account.
54	'i-q-rin'	While some discrepancies between reported IRRs and paybacks might be expected, it is anomalous that the calculated IRR (computed using a single lifetime for all projects) explains the payback so much better than the reported IRR.
55	'ai'	cost components) may be thought of as indicating whether companies measured the transactions costs associating with this type of investment.
56	'ai'	Interaction between design and allocation choices The interaction between the trading design options and the possible allocation methods raises an additional distributive issue with equity considerations for decision-makers, because the choice between (a combination of) grandfathering and auctioning may depend on the choice between the di!
57	'ai'	A mitigation project may be defined as a planned set of activities that are bounded by specific geographic and temporal boundaries and an identifiable set of institutional arrangements (Brown et al. 2000b) that generate potentially tradable GHG emissions offsets.

58	'ai'	The clean development mechanism (CDM) ² , for example, limits the accounting of carbon pools for certain types of forestation projects to those within a project boundary and stipulates three different ways that project-specific baseline scenarios may be set: (1) existing or historical, as applicable, changes in carbon stocks in the carbon pools within the project boundary, (2) changes in carbon stocks in the carbon pools within the project boundary from a land use that represents an economically attractive course of action, taking into account barriers to investment, and (3) changes in carbon stocks in the pools within the project boundary from the most likely land use at the time the project starts.
59	'ai'	Finally, permanence or reversibility refers to whether the net emissions reductions attributable to the project may be reversed in the future due to natural or manmade causes, or convergence of the actual baseline with the project case emissions.
60	'i-q-hnv'	Subsequent approaches recognized that land use change rates may depend on socioeconomic and other drivers that may be used to set rates for a much larger area.
61	'i-q-hnv'	For CDM, propose new or use approved baseline methodology Transaction costs High, could decline if methods standardized Transparency Low, but may be high if program requires it, e.g.
62	'i-q-hnv'	Low where multiple projects exist Medium, but may be high if program requires it Likely
63	'i-q-hnv'	A baseline may be adjusted after the project has been in place for some years, if a GHG program dictates the validity period and the conditions that trigger review or revision of baseline driver variable assumptions.
64	'i-q-hnv'	Adjustable baselines may be preferred from a GHG program's perspective in order to permit more accurate accounting of a project's carbon benefits over its life, but if they pose too high a risk for project developers, few projects may be implemented.
65	'ai'	Advantages: Limited spatial data requirements may be gathered from government statistics and maps Disadvantages: Temporal period and spatial zone not investigated
66	'ai'	The LUCS and GEOMOD methods may be applied to both smaller (project) scale lands and more broadly at the regional level.
67	'ai'	The latter two approaches may be used either at the project scale or a broader regional scale.
68	'ai'	4.1 Summary classification of methods by mitigation option Based on experience to date estimating regional baselines, Table 4 shows a summary classification for three mitigation options—avoided deforestation, forestation, and forest management—where similar data and methods may be used to estimate land use change and carbon stock change rates for setting an emissions rate.
69	'i-q-hnv'	Biomass carbon stock estimates may be available from forest inventory or other plot measurement data, or estimated by Holdridge-type forest classification, and soil carbon may be available in inventory data or estimated using spatial covariation with respect to soil texture and drainage (Davidson 1995).
70	'ai'	Above-ground biomass/BGB may be estimated using forest inventory data or measurements by forest SOC: sampling or partial covariation approach may be used to estimate SOC Example: Davidson (1995) Carbon density based on measurement of above-ground biomass and SOC of a stratified sample of land cover and use Examples: Tipper and de Jong (1998)
71	'ai'	Historical time trends or a model that relates changes in management practices to driving variables may be used to classify the rate of change of practices.
72	'i-q-hnv'	One potential land management baseline approach may be to identify the suite of management practices in use by land and management conditions, in existing datasets or through landowner or expert surveys, and estimate a baseline utilizing differential implementation of these practices over the land base (see Sect. 5.1.1 for an example).
73	'ai'	The FAC model is applied at the national or regional level, whereas the other two LUCS and GEOMOD may be applied at the project or neighboring area level.
74	'i-q-hnv'	Others (e.g., 15.7+ m3 productivity in that type) have far higher standard errors (42.1 in this case) and thus may be less useful or require additional variables.
75	'i-q-hnv'	An alternative might be to set a limit on the magnitude or percentage change in the values of key driver variables for a given confidence interval—a zone within which they could float, but would need to be revised if they exceeded it.
76	'ai'	For land use change and forestry projects, the appropriate metric usually is the annual change in carbon stock per unit land area, which may be expressed in t C sequestered or emissions avoided per hectare of land per year (t C/ha/yr).
77	'ai'	These may be specified by a mitigation program targeted by a project, and/or based on the available data.
78	'ai'	If there is no clear trend, an average rate of past trends may be used.
79	'i-q-hnv','i-m-val'	The minimum future time period of validity for the baseline will be dictated by GHG mitigation program rules and by the intervals over which data become available, which may be 5 years or more for deforestation reduction projects that rely on remote sensing imagery.
80	'ai'	Project future baseline emissions rate: A future baseline emissions rate may be estimated using a single average historical value or simple extrapolation of the time trend.
81	'ai'	Where more sophisticated modeling has been done, the rate may be based on key driver variables.
82	'ai'	A SRB emissions rate may be determined using historical data or near-future estimates of land use and carbon density change.
83	'i-q-hnv'	A break due to a prolonged drought may not indicate a permanent change in the trend, but one caused by a technological breakthrough or change in land use practice might be an indicator of a new emerging trend.
84	'ai'	Atmos Environ 43:2571–2578 Papen H, von Berg R (1998) A most probable number method (MPN) for the estimation of cell numbers of heterotrophic nitrifying bacteria in soil.
85	'i-q-hnv'	shortage may be selected for (Metcalfe and Monaghan 2001).

86	'i-q-hnv'	Micro-evolution in response to climate change may be constrained by the specific lack of genetic variation for the trait under selection (Kellermann et al. 2009) or by the disagreement between the multivariate genetic correlation structure of traits and the predominant direction of multivariate selection (Walsh and Blows 2009).
87	'i-q-hnv'	Since the expression of genetic variation may be low specifically in situations of adverse environment and strong selection (Wilson et al. 2006), the need for phenotypic plasticity may be the greatest in exactly those conditions when it is suppressed.
88	'i-q-hnv'	Abstract The Kyoto Protocol introduces the possibility that changes in carbon stock on agricultural and forest land and soils may be counted against countries' commitments to reduce their greenhouse gas emissions.
89	'i-q-hnv'	These latter considerations for including a subset of the source can be influenced by a range of criteria aside from monitoring/verification criteria. An important criterion may be the environmental impact of different agricultural practices that sequester carbon. Advocates of elective measures to control greenhouse gas emissions worry that including previously uncounted carbon sinks related to Article 3.4 activities during the "first commitment period (2008-2012) will allow parties to take less stringent measures to limit emissions in other sectors.
90	'i-q-hnv'	Of these approaches, monitoring based exclusively on micro-meteorological sampling (1) may be immediately dismissed as too expensive at the scale required for a national inventory.
91	'i-q-hnv'	sets may be only \$1.5-\$3.0 billion assuming a value of \$50-\$100/tonne carbon applying to 30 Mt C sequestered.
92	'i-q-hnv'	erent practices may be difficult to prove or disprove at the level of a single farm or single hectare.
93	'i-q-hnv'	Further research may be needed to better understand the relative value of residues and crops.
94	'ai'	Although the potential area for organic soil disturbance in North America may be smaller than for conservation activities, unless changes to these soils are considered within a carbon assessment, an important source of carbon dioxide emissions would be overlooked.
95	'i-q-hnv'	Limitation on Article 3.4 credits Both cover crops and conservation tillage have become increasingly popular among US farmers, but neither practice is believed to have reached saturation, so a sizable carbon sequestration potential may exist.
96	'i-q-hnv'	In Canada as well, the relative potential for additional carbon accumulation may be greater than in the United States, because no-till is practiced on less than 10% of Canadian farmland (Dumanski et al., 1994).
97	'i-q-hnv'	For example, some activities, such as biomass yield achievement, may be inherently profitable but not merit extra compensation.
98	'ai'	clay content of the deeper soils most probably accounted for this observation, as soil gas diffusion is more strongly controlled by texture than by moisture content in finertextured soils (DÃ¶rr et al. 1993).
99	'ai'	oxidation in activity flasks, indicates that only very low levels of CH ₄ production may be occurring.
100	'ai'	Predictor variables may be either numerical or categorical.
101	'ai'	Some of these may be addressed by matching demands for mitigation (for activities and locations where adaptive capacity will be exceeded) with the ability to supply that demand through localised mitigative capacity by means of globally integrated mechanisms.
102	'i-q-hnv'	A single pass may be sufficient for understanding individual impacts, but if the aim is to avoid widespread critical damage, then outcomes with, and without, risk treatment options need to be assessed.
103	'ai'	Assessing the benefits of climate policy over time is difficult but the following points are relevant to managing climate risk with respect to action and response: • Investments in mitigating greenhouse gases (including carbon sequestration) will take some time to be realised in terms of the benefits of avoided damage; • Investments in mitigation made in one location will not be realised in terms of benefits of avoided damage in that location but will contribute to benefits across the globe; • The benefits of adaptation will be felt locally but depend on when anticipated climate change/variability occurs; • Co-benefits of actions may be realised on quite different timescales to the climate-related benefits.
104	'ai'	This can be viewed as the range of capacities that may be present over a region, a nation or the world.
105	'i-q-hnv'	The mitigative capacity of the regional agriculture and agricultural soils is unexplored but may be substantial when compared to emissions.
106	'ai'	Biodiversity benefits are high in riparian zones, areas of former box-ironbark forest, mallee and native perennial grassland and grassy woodlands, where sequestration rates and storage values may be much lower.
107	'i-q-hnv'	• Reduced water supply • the purchase of water rights may be required before plantations can be established, to compensate for the loss of water elsewhere in the catchment.
108	'i-q-hnv'	Most policy affecting adaptation and mitigation is likely to be • top down •, instituted at state, provincial or national level, while actions facilitated by those policies may be much more local.
109	'ai'	However, many of the power needs in northern Victoria may be addressed by niche generators involving both renewable energy and cogeneration.
110	'i-q-hnv'	We have conversations between Confucius and his followers in which there is nothing definite further than a commonplace moral put in the form of good, sound doctrine, which may be found as well expressed and better, in every place and amongst every people.
111	'ai'	James's review of Confucius and his followers's geographical thoughts still illustrates that a cultural barrier may be more important than a language barrier in modern times (James, 1972, pp.71-79).
112	'ai'	Also subject to criticism is the declaration • the central concern of this report is to assess how the array of approaches and techniques of the geographical sciences might be most effectively deployed in the effort to address major social and environmental questions • (NRC, 2010a, p. x).
113	'i-q-hnv'	In the West, Eliot Hurst's • geographical nihilism • may be one extreme • geography has neither an existence nor a future (Hurst, 1985).

114	'ai'	Although the actual WTD limit on vascular plantgrowing hummocks might be greater, the estimated values are not expected to vary much because the suppression of the growth in <i>S. papillosum</i> begins
115	'i-q-hnv'	These issues point towards the need to measure carbon, nitrogen and GHG fluxes above croplands and in the long term in order to (1) evaluate and compare the components of the nitrogen and carbon fluxes over the years and crop rotations, (2) estimate the uncertainty in the different components of the carbon and nitrogen fluxes and discuss them, and (3) estimate the carbon, nitrogen and GHG balances of these ecosystems, where N ₂ O may be a large contributor.
116	'ai'	Due to its sensor head heat capacity, the infrared gas analyzer (Li7500) was shown to produce errors in the CO ₂ flux, which may be large under cold conditions and can be corrected for, as an additional WPL (Webb Pearson Leuning) correction following Burba et al. (2008).
117	'ai'	The energy balance deficit is most certainly a translation of these drawbacks.
118	'i-q-hnv'	In the crop rotation studied here, with no leguminous, N ₂ fixation via symbiotic fixation is expected to be small (there may be a small contribution from weeds).
119	'i-q-hnv'	On the longer term, Cannell and Thornley (1998) suggested that the grassland relative response to elevated CO ₂ may be better in N-poor ecosystems than N-rich ones, but Reich et al. (2006) proved the opposite.
120	'i-q-hnv'	Taken together, our results indicate that the lowered soil labile C and N pools may be induced by the increased soil CO ₂ efflux.
121	'i-q-hnv'	Additionally, the temperature sensitivity of soil respiration may be less in soils with low carbon content (Luo et al. 2001).
122	'i-q-hnv'	All in all, responses of labile pools of SOM to experimental warming may depend strongly on ecosystem type or forest management practice.
123	'i-q-hnv'	The normal temperature range of the air passing through the air handlers is 21–29 °C. Because the rain forest atmosphere is connected with the atmospheres in the adjacent desert, savanna, and ocean biomes, there may be some loss of atmospheric humidity into the surrounding biomes (recently installed curtains can now be used to isolate the rain forest atmosphere from adjacent ecosystems).
124	'i-q-hnv'	However, the reader is advised that values of T _a shown later may be prone to absolute errors on the order of a few degree Celsius.
125	'ai'	At the ABRACOS Reserva Jaru site, for instance, monthly rainfall may be less than 10 mm per month in the dry season, but at other times of the year, rainfall at this site (and at the other ABRACOS sites) can be several hundreds of millimeters.
126	'i-q-hnv'	Recognizing and quantifying this equivalence in solar radiation regimes may be helpful in selecting the type of study that can be carried out in the B2C with most real-world relevance.
127	'ai'	To the best of our knowledge, N ₂ O production has not been investigated in mesic low arctic tundra soils during spring, when environmental conditions might be most favourable for denitrification.
128	'ai'	Nutrient addition beyond the upper limit to microbial N immobilization in early spring may be removed from the system, by facultative denitrifiers, or via physical means as leachate flows to the mineral soil or overland surface flows during spring thaw.
129	'ai'	In countries with relatively large changes in land area under cereals, production was, as may be expected, correlated to the land area under cereals (Table 4).
130	'ai'	reported that a reasonable average diffusivity for a drying clay soil might be 100 cm ² day ⁻¹ .
131	'i-q-hnv'	Summer fallow is not practiced under the current climatic conditions of the study area but may be adopted in the case of a decrease in rainfall by 46% projected by RCM.
132	'i-q-hnv'	Reasons for low number of identified global extinctions are suggested here and include success in protecting many endangered species, poor monitoring of most of the rest of species and their level of threat, extinction debt where forests have been lost but species still survive, that regrowth forests may be important in retaining “old growth” species, fewer co-extinctions of species than expected, and large differences in the vulnerability of different taxa to extinction threats.
133	'ai'	Hubbell et al. (2008) estimated the number, relative abundance and range size of tree species in the Brazilian Amazon using Hubbell’s neutral theory (Hubbell 2001) and then examined the optimistic and pessimistic scenarios of Laurance and co-workers to estimate what proportion of species may become extinct.
134	'i-q-hnv'	Although the way such lists are compiled means that their use in such extrapolations may be questionable (Possingham et al. 2002) this work provides further demonstrations that extinction rates vary considerably between different groups.
135	'ai'	In summary, climate envelope models provide an upper limit on the number of extinctions that might be possible rather than the number that might be expected.
136	'i-q-hnv'	Poleward and elevational shifts of tropical species may be hampered in anthropogenically modified and fragmented landscapes and by the lack of some key seed-dispersing animals previously lost through overhunting.
137	'ai'	Amazonian population growth Because deforestation rates may be influenced by the size of the Amazonian population, we obtained census data for the Brazilian Legal Amazon.
138	'ai'	Although the correlation matrix might have arbitrary structure, for orthogonal or close to orthogonal basis functions we believe I is a reasonable choice.
139	'ai'	Any real valued random field on a sphere that has finite variance and has realizations which are square integrable over the surface of the sphere may be represented as an infinite series of spherical harmonics (e.g. Jones 1963).
140	'ai'	While the economic activities may be ruled only by the costs and prices expressed in money values, the impacts on the environment and human health are of very different nature starting at one extreme with the loss of a human life and ending with the destruction of a beautiful landscape.
141	'i-q-hnv'	Simulations of coral bleaching events under future climate change have also revealed a geographic variability in required thermal adaptation and imply that coral reefs in some regions, like Micronesia and western Polynesia, may be particularly vulnerable to global warming (Donner et al. 2005).

142	'i-q-hnv'	Because of the complex mixture of freshwater and marine NPP, the simple Traverse $\delta^{18}O$ marine index (MI) is of limited value for recording fluctuation in Holocene sea level within the Black Sea, although it may be valid for megacycles of marine transgression and regression during the Pliocene–Pleistocene.
143	'ai'	Data may be assumed to be from several families of probability distributions, including the normal, binomial, Poisson, negative binomial, or gamma distribution, many of which better fit the nonnormal error structures of most ecological data (Guisan et al., 2002).
144	'i-q-rab'	Of particular relevance to the focus on temperature and salinity responses to MOC changes, Cunningham and Alderson (2007) diagnose the vertical structure of changes in temperature and salinity at 24.5° N and discuss how these changes may be linked to MOC slowdown reported by Bryden et al. (2005).
145	'i-q-hnv'	Disadvantages of these methods may be that other phenomena project on the signature (e.g., north–south gradient or $\delta^{18}O$ interbasin balances) and the derived indices are partly different from MOC trends.
146	'i-q-rin'	It is conceivable that more high-resolution proxy data in sensitive regions may become available in order to trace past MOC variations and trends.
147	'ai'	Afandi et al. (2010), for the middle Egypt area under A2B2 climate change scenarios, in which the adoption of correct irrigation scheduling may be used to reduce the negative impact on heat yields due to higher temperature and lower rainfall.
148	'i-q-hnv'	These results also highlighted that summer crops, including tomato, may be strongly affected by climate change that in this very important agricultural area of Italy.
149	'ai'	temperature may be considered the primary cause of significant losses in tomato yield and quality (Dane et al. 1991).
150	'ai'	Conclusions This simulation study has provided details relating to the responses of crops under climate change and also how management practices may be used to maximize the crop yield optimizing the application quantities and the time of application during the crop cycle.
151	'i-q-hnv'	It appears there may be a delay to increased microbial N ₂ O production in response to changing soil environmental variables, principally soil WFPS, and this should be considered when investigating or predicting N ₂ O emissions in pasture systems under changing climate conditions.
152	'i-q-hnv'	However, under environmental stress, this balance may be perturbed and the production of ROS can be enhanced.
153	'ai'	Here, we illustrate how the elderly may be affected by disasters by breaking the review's findings into four categories based on the stages of a disaster: preparation, occurrence of the hazardous event, response, and recovery and reconstruction.
154	'i-q-hnv'	During the occurrence of a hazardous event, life-support equipment, such as oxygen generators or electric wheelchairs, may be dysfunctional from breakdowns in lifelines (Fernandez et al. 2002; California Department of Aging 2004).
155	'i-q-hnv'	As a result, uninstitutionalized elderly with very small social networks may not know how to react without external assistance or may simply deny their vulnerability, whereas those living in eldercare institutions may be less vulnerable, thanks to the shifting of responsibility for care from the individual to the institution.
156	'i-q-hnv'	These locations are home to the most socially vulnerable elderly who—although they may be spared from hurricane hazards caused by weak and moderate storms—can potentially suffer disproportionately when a stronger hurricane hits the area.
157	'ai'	The input parameters were based on the most probable interpretation of each radar image to be validated.
158	'i-q-hnv'	The hyperbolic returns came from a profile slanted towards the short axis of the glacier, which indicates that the profile may be part of a transverse profile of a water body or river.
159	'i-q-hnv'	hyperbola may be part of englacial river flow.
160	'i-q-hnv'	It is almost certainly related to the characteristics of the ice body itself, and the simulation shows that the scattering phenomenon is highly dependent on the englacial small-scale water distribution.
161	'ai'	The cases show that this logic may be defended as at least being equally reasonable to the emission-effect calculations that form the core of Risk Assessment and Life-cycle Impact Assessment (LCIA).
162	'ai'	On the basis of a review of case-studies, an attempt was made to formulate a number of characteristics of knowledge claims that make them vulnerable for flaming (TuKKER, 1999a) 7, s The results, reflected by Table 2.2 suggest that when the problem is simple, it most probably can be solved with robust knowledge via a traditional decision making approach.
163	'i-q-hnv'	Policy scientists see this as an important indication that a controversy may be rooted in framing differences rather than 'scientific' uncertainty.
164	'ai'	In terms of management solutions, naturally occurring substances may be emitted up to a level that depends on the natural background.
165	'i-q-hnv'	the practitioner acknowledges the fact that (s)he and/or his/her method may be biased.
166	'ai'	No part of the issue may be reproduced by a n y mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher.
167	'i-q-hnv'	If temperature increases occur mainly in the summer, the arrowhead scale may be under control by increased Aphytis densities, but this may be balanced by declining Coccobius densities.
168	'ai'	Where data availability allows, baseline periods extending further to the past may be used for some purposes, particularly when focusing on the statistics of extremes, which are difficult to estimate from short time series.
169	'i-m'	(6), (8) and (9) is that, in rare circumstances, the ratios in these equations may become negative or approach infinity.

170	'ai'	Appendix: Brier score and the continuous ranked probability score Consider a binary event E. For example, E might be defined to occur when the mean temperature of a given month in a given place exceeds a given threshold τ . The Brier score B (Brier 1950; Wilks 1995) for this threshold value is defined as $B = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
171	'i-q-hnv'	Promiscuous larvae allow species to associate with Symbiodinium appropriate for potentially novel environments that may be experienced following dispersal.
172	'ai'	Vertical transmission may benefit the host because it guarantees transmission of a symbiont that was successful in the parent colony.
173	'i-m-int','i-q-rab'	carbon and nitrogen profiles, and below average values since approximately 1950 may be associated with increased atmospheric N loading or reduced productivity.
174	'ai'	The use of trade and company names is for the benefit of reader; such use does not constitute an official endorsement or approval of any service or product by the USDA to the exclusion of others that may be suitable.
175	'ai'	Based on the above results, we believe that the FORFLUX model adequately predicted ecosystem-atmosphere exchange of water vapor and O ₃ for the nine sampling days at the study site.
176	'ai'	We have found that biophysical models might be useful when assessing eddy flux measurements made in rough terrain.
177	'i-q-hnv'	As observed by Hao et al. (2004), the lower (Control $r = 0.385^*$; $n = 33$) and non-significant (DDGS $r = 0.128$; $n = 32$) correlation coefficients between the rate of O ₂ consumption and N ₂ O emission indicate the relationships between the two parameters are complicated and may be non-linear.
178	'i-q-hnv'	the injectivity may be used to characterize the flow & stance of naturally formed geothermal index reservoir reeks to liquid production and predict the deliverability of geothermal fluids from slim holes.
179	'ai'	Potential CO ₂ e emission reductions could be gained from container designs that reduce or eliminate polypropylene mass, or by increasing the number of times containers may be reused in operations, as the nursery already reuses containers.
180	'i-q-hnv'	Substituting a renewable energy source for propane may be an option for reducing energy emissions along with improving insulation of the heated space.
181	'ai'	In conclusion, we find that air mass types have independent effects on mortality for both the cold and warm season and may be used to predict weather-related adverse health effects.
182	'ai'	More details on the methods and data used may be found in Kassomenos et al. (2003).
183	'i-q-rin'	While the predicted warmer air temperatures in the central South Island of New Zealand (Ministry for the Environment, 2008) may result in warmer lake surface waters and stronger stratification (e.g., greater stability, shallower thermocline, an earlier onset of stratification) in monomictic New Zealand lakes, the potential exists for higher wind forcing and increased (cold) river flows in winter and spring to offset the effects of air temperature warming on the thermal structure.
184	'ai'	This assumption might be conservative (i.e., optimistic), but it has been shown to provide good estimates of species loss levels for mountain areas (Engler et al. 2009).
185	'i-q-hnv'	Our models projected that many of the vegetation units analyzed in this study may be threatened by climate change.
186	'i-q-rin'	The extinction of some of the alpine and subalpine species would lead to a reduction in genetic diversity on the Iberian Peninsula, although it may be expected to survive in other European areas (Benito Garzoñán et al. 2008).
187	'ai'	It has previously been shown that many waterfowl species might benefit from occupation of agricultural habitats, as this often result in forage of high energetic profitability and expansion of suitable habitats (Lovvorn and Baldwin 1996; van Eerden et al. 1996; Therkildsen and Madsen 2000; Fox et al. 2005).
188	'ai'	In this regard, Shashua-Bar et al. (2010a) have suggested that the reconciliation of these conflicting ecosystem services might best be realized through landscape design modalities consisting of shade trees alone, while Zhou et al. (2011) have found that the composition and configuration of vegetated landscape has significant impacts on UHI mitigation.
189	'i-q-rab'	Although surface air temperatures (SATs) or surface heat fluxes rather than T850 may be more directly thermodynamically related to sea-ice and oceanic fields, the surface data from NCEP/DOE reanalysis is quite closely dependent on the model parameterization, and does not very well match the observations of SAT and surface heat flux (Kalnay et al., 1996; Kanamitsu et al., 2002).
190	'i-q-hnv'	A natural question associated with the importance of the autumn preconditioning of the Okhotsk Sea studied in this paper may be whether a similar relation is observed in other seasonal sea-ice regions.
191	'ai'	Within the scientific culture, the most valid and reliable method of fixing belief is, of course, that of science; within the non-scientific culture, however, other methods of fixing belief may be considered just as valid and reliable.
192	'ai'	One reason that efforts at science communication may be ineffective, especially on controversial issues, is because they are often based on a 'deficit model' of public understanding of science.
193	'i-q-hnv'	He replied: "We used to think that we could publish speculative research which advanced interesting new ideas which may be wrong, but which were important to provoke debate and discussion. We don't think that now [We don't seem able to have a rational conversation in a public space about difficult, controversial issues, without people drawing a conclusion which could be very, very adverse." (A shot of reality 2010) Despite the fact that Horton, in the same interview, describes the controversy as a "system failure," involving media, government, and the scientific community, this quotation underscores the implicit assumption that the science should be able to speak for itself but cannot because the media distort the science and the public continue to hold on to irrational fears.
194	'ai'	While some scientists may believe that providing a science-illiterate public with more information will turn them into a science-literate and science-appreciative public, the relationship is not quite so simple.

195	'i-q-hnv'	Scientists TM claims that "there is no evidence that MMR vaccine causes autism" and that "autism occurs after MMR vaccine at the same rate that it occurs in children who did not receive the vaccine" are clearly generated out of respect for the scientific method but may be incomprehensible to a public that is unaware of or unconcerned with issues regarding scientific methodology (Offit and Coffin 2003).
196	'ai'	For issues on which they have little or no personal experience, however, media consumers may depend on information presented by mass media to form an opinion.
197	'ai'	Like most HFEs, a general mechanism of tropospheric degradation mechanism of HFE-125 may be shown as follows (Scheme 1): The above scheme is based on the fact that the initial attack of photochemically produced OH radical on HFE-125 leads to the formation of haloalkyl radical (CF ₃ OCF ₂ •).
198	'i-q-hnv'	From the statistical point of view, different models may give similar good fits to data, but results may be contradictory in terms of flow direction.
199	'i-q-hnv'	A downward concave thermal profile may be accounted for by downward flow or horizontal leakage of cold water, whereas an opposite shape can be the result of upward movements or a warm horizontal flow.
200	'i-q-hnv'	Climatic changes may distort the temperature log with the same style of groundwater flow, and the simultaneous assessment of the relative magnitude of groundwater and climatic perturbations may be a difficult task (see e.g., Taniguchi et al. 1999; Bodri and Cermak 2005; Ferguson 2005).
201	'ai'	There being in 1952 no satisfactory theory to deal with irreversibility, this might be taken as a practical man's injunction not to use a tool where it had no business, and indeed it was; but the line of argument was not put to rest with the subsequent development of option value theory.
202	'ai'	, the probability distribution of the future values of that which might be irretrievably lost was assumed known.
203	'ai'	In the problem treated, the set of states of the world (i.e., the values that might be taken on by each random viable) is known, but there is no probability distribution over states; the set of states for any random variable, for any given action, when ordered from least to most desirable, is bounded; and the payoffs depend only on states of the world.
204	'ai'	The action may be thought of as breaching the SMS, and the subsequent certainty of extinction as simplification convenient for placing two possible games against nature in the same framework.
205	'ai'	The SMS suggested by Randall is a true compromise because it specifically denies preservation such "status: the SMS may be abandoned when its cost is intolerably high.
206	'ai'	While the static model may be understood as applying to a world of only flow variables, the dynamic model is focused on thresholds in state variables.
207	'ai'	Our results may be seen as an attempt to reduce the scope for argument between proponents and opponents of SMSs as a policy tool in that we present simple conditions for when a SMS is consistent with utilitarian calculus.
208	'ai'	6 In this case, the problem may be modeled in a static framework, since it is not truly dynamic (Samuelson 1947).
209	'ai'	We present a simple model to illustrate how the optimality of a SMS may be determined.
210	'ai'	The good x_2 may be thought of as an aggregate of other goods.
211	'i-q-hnv'	The intensity $\hat{I}(x)$ is assumed everywhere continuous and increasing ¹⁰ except possibly when $\dot{\hat{I}}(x(t)) = 0$ where the derivative of $\hat{I}(\hat{A})$ may be zero.
212	'i-q-hnv'	6.8 is may be envisioned as a sequence of problems, where one solves a deterministic optimal control problem whenever $\dot{\hat{I}}(\hat{A}) < 0$ and a piecewise deterministic problem whenever $\dot{\hat{I}}(\hat{A}) > 0$.
213	'i-q-hnv'	$\hat{A}^-(10) \geq T, x + g(x) \hat{I}, \hat{A}^+ \geq T, x + \hat{I} \geq x + T$ The central insight from (9) to (10) is that there may be a jump in the co-state variable when $x(t)$ crosses from one side of $\dot{\hat{I}}(\hat{A}) = 0$ to the other.
214	'ai'	As always, the co-state variable may be thought of as the shadow cost in terms of current and future welfare incurred when an additional unit of the state variable is made freely available.
215	'i-q-hnv'	An infinite number of such jumps may be allowed, and will, in certain circumstances, be required if a path is to observe necessary conditions.
216	'ai'	This may be considered a simple model of carbon control emissions when there is a risk of a disintegration of the Western Antarctic Ice Sheet as discussed in the introduction.
217	'ai'	After all, the catastrophic event has already happened so there ¹² This particular chosen form for $\hat{I}(x)$ may be considered a Taylor expansion around $x = x_l$, of an underlying risk structure.
218	'ai'	$\hat{A}^c(0) = (r + \hat{I})\hat{A}\mu x + \hat{I} > 1 (u(\hat{A}^+ \hat{I} - x_l))^2 + (19) 2r$ 13 A co-state variable for $a(t)$ also may be computed, but does not affect the solution and has no important economic interpretation.
219	'i-q-hnv'	If a regulator is able to form a probability distribution over possible values of A , then a SMS may be optimal if $E(A) \neq \hat{A}$ — even if $A < \hat{A}$ — with some probability.
220	'ai'	or we must at least be able to form meaningful probability distributions over the values they may be able to take.
221	'ai'	The choice of the quadratic function implies that $\hat{I}(x)$ may be considered a Taylor expansion of the underlying risk structure.
222	'i-q-hnv'	The algebraic procedure $\hat{E}(x)$ may be quite complicated, but once this task has been needed to translate (x) to achieved, the results from above follow directly.
223	'ai'	31, 38 it is clear that the conditions in (31) and (38) may be rewritten $p(T, \hat{A}^+) = p(T, +) + \hat{I} \dot{\hat{I}}(x(t))$
224	'ai'	From our study, this haplotype was the most common strain of this Trebouxia species in the Ross Sea Region and may be selected to the current or geologically recent climatic conditions.
225	'ai'	Possible examples of why customers choose not to alter their beliefs include holding on to socially or morally superior beliefs, not having to change lifestyle to reflect the beliefs, avoid having to form new costly beliefs and holding onto beliefs that may be used more efficiently with time.
226	'i-q-hnv'	Buying green electricity may be an investment to develop, improve or reinforce their corporate image.

227	'i-q-rin'	The issue of whether the electricity that actually comes down the wires is from renewable sources may become important, however.
228	'i-q-hnv'	So, a high initial willingness to pay for renewable electricity may be accompanied by a high preference to free ride, possibly higher than in the low information case.
229	'ai'	Because of the special circumstances in the renewable electricity market, in particular, inelastic supply in the short run and inelastic consumer beliefs associated with environmental quality and free riding, its successful development may depend critically on the initial demand.
230	'i-q-hnv'	In contrast to traditional simulators, under sufficiently simplifying assumptions much less computationally demanding models may be developed.
231	'ai'	These aspects correspond to our initial caveats (4) and (7) listed at the outset and may be evaluated through a comparison of the observed, NCEP and HadCM2SUL data.
232	'i-q-hnv'	Any discrepancy between the downscaling and GCM may be an artefact of the downscaling method per se, inaccurate or inappropriate GCM predictors, an inaccurate GCM surface climate, or a combination of all four.
233	'i-q-hnv'	However, from a scientific/sustainability point of view, it may be preferred to use the 500-year time horizon, as in this case, additional relevant impacts are captured.
234	'i-q-hnv'	The user must, however, be conscious that the estimated effect of PM may be an indicator of the overall effect of the air pollution rather than based on a proven causeâ€effect relationship for PM.
235	'ai'	The models may be continental in scale or even national as long as the region they cover is of an adequate heterogeneity.
236	'i-q-rin'	The European continent is so large and heterogeneous that the European-based characterization factors may be expected to serve as reasonable proxies also for other continents.
237	'ai'	This implies (and may be parsimoniously explained by) a widening of the Ferrel cell, reminiscent of the pattern seen in Kidston and Vallis (2010) when studying climatological eddy-driven jet position and width.
238	'i-q-rin','i-m-int'	Hence an uncertainty of around 5% may be expected in this study.
239	'i-m'	The higher and sustained turbulence at this level is almost certainly associated with the movement of air between the air handlers that are located below the canopy in the B2C rain forest biome.
240	'i-m-int'	The thermal tolerance threshold for <i>A. salinus</i> development might be around 25â29Â°C.
241	'i-q-hnv'	If the fire interval is too short, populations may be killed before they have produced enough seeds to replace themselves.
242	'i-q-hnv'	As a consequence, many fragments are long unburned, and intervals may be approaching the upper limits for some fire recruiting non-sprouting species to persist.
243	'i-q-hnv'	While in our model we assume seed survival is 100% in the first 2 years and 33% thereafter, there may be some uncertainty about this, as in studies of other plant taxa (not <i>Verticordia</i>) it has been found that loss of seed to predators can occur in the soil (Burne et al. 2003).
244	'i-q-hnv'	needs to be considered for <i>V. fimbrialepis</i> and may be relevant to species in other parts of the world.
245	'i-q-hnv'	If the process of conducting IAs and EIAs are highly procedural (Loiselle 2004), and involve a protracted governmental approval process (Creasey and Beswick 2004) then, as Montanez-Cartaxo (2004) suggests, the best scientific understanding and technical experience about environmental protection may be thwarted by government attitudes, customs, and procedures.
246	'i-q-rin'	Given this context, a developer is then confronted with the decision of either positively or negatively influencing the project approval process by including (or overlooking) issues which may result in the project proposal being rejected (or approved).
247	'i-q-hnv'	Oâ™Faircheallaigh (2010) suggests that developers (businesses), in wanting their project to be approved, â€are likely to ignore or downplay negative impacts or risks, and to exaggerate potential project benefits, and in particular may be prone to exaggerate the economic benefits that usually constitute the major justification for large industrial projectsâ™ (p. 21).
248	'i-q-hnv'	There are several points of departure from the typical EIAs that we propose during Parameter Setting: first, we believe this step is more likely to be viewed within the realm of Risk Assessment for a business, and that the business will determine whether the risks of the project (in terms of environmental impact and potential mitigation requirements) are worth further investment of time and resources.
249	'i-q-hnv'	Therefore, in our study the rate of forest coverage from 1981 to 2002 is estimated as 13.54%, which may be higher than that of previous studies.
250	'i-q-hnv'	These latter features may be crucial to an understanding of liana success in dry environments and whether all liana species will be favoured or only those adapted to drier environments.
251	'i-q-hnv'	These findings are consistent with a relatively short food chain for benthic fish that are for the most part specialized feeders with narrow preferences for food and may be affected by changes in benthic prey distributions.
252	'i-q-hnv'	Therefore, passive CH ₄ transport by molecular diffusion may be the dominant process in <i>P. arundinacea</i> as water levels in the area seldom increase above the peat surface at the investigated field site.
253	'i-q-hnv'	However, as only 80% of the PAR passes through the Plexiglas, C sequestration may be underestimated.
254	'i-q-hnv'	As the convergence of the requisite technologies of spaceflight became possible, communities began to form around the vision as a shared goal that might be achieved through the specific technology of liquid-fuel rockets.
255	'i-q-hnv'	The high correlations between the leading EOFs in the South Atlantic and Indian Oceans, particularly during FM, suggest that, rather than thinking of these leading EOFs as different modes of SST variability, it may be more appropriate to think of them as a single mode of variability (Fauchereau et al. 2003; Hermes and Reason 2005).
256	'i-q-hnv'	In deed, the 400 and 200-hPa temperature correlation patterns suggest that the enhancement process for the local convection and westerly surface wind events over the western Pacific may be triggered by this huge temperature response in the upper atmosphere remotely forced by the large-scale subsidence over the Atlantic and Indian Oceans.

257	'i-q-hnv'	5 Conclusions and discussions The present study suggests that extra-tropical atmospheric forcing of ENSO may be more important than previously thought.
258	'i-q-hnv'	Furthermore, our results suggest that the enhancement process for the local convection over the western Pacific may be triggered by a "cold Kelvin" wave at middle and upper levels in the atmosphere remotely forced by the large-scale subsidence over the Atlantic and Indian Oceans.
259	'i-q-hnv'	Secondly, if such a positive relationship can be scientifically attested, then there may be less need for further governmental intervention in relation to environmental regulations, and some existing regulations could probably be relaxed (Telle 2006).
260	'i-q-hnv'	However, the very limited existing data suggest that lichen growth rate may be a good indicator of climate change in Antarctica.
261	'i-q-hnv'	Sancho et al. (2007) suggest that lichen growth rate might be one of the more sensitive indicators of climate change in Antarctica.
262	'i-q-hnv'	There is the possibility, because all of the lichen population was not surveyed, that larger thalli might be present; however, we suspect that this is unlikely as an attempt was made to select the largest lichens at the beginning of the measurements.
263	'i-q-rab'	However, the environment is not extreme for the lichens in terms of temperature or dryness, and the results provide hints that snow quantity and duration might be the key environmental factors in determining the opportunity for lichen colonisation and survival.
264	'i-q-hnv'	Nevertheless, the observed flux underestimation might be even larger when the observed flux reduction due to collar insertion in most CSC cover-box studies (and the subsequent loss in root-derived fluxes, partic-
265	'i-q-hnv'	The potential impact of climate change on the structure and function of subalpine coniferous forests such as in the western Sichuan region may be of great importance to the global carbon balance (Wang et al. 2003a, 2003b).
266	'i-q-hnv'	However, this problem may be less important, since relative humidity was not a limiting factor for the growth of subalpine conifer species in this region.
267	'i-q-hnv'	Competitive and adaptive relationships between the conifers may be altered as a result of response differences to warming manipulation.
268	'i-q-hnv'	It might be that differences between the nutrient treatments will only become visible in a next growing season.
269	'i-q-hnv'	Bogs with thick peat deposits may be more P-limited, because of the distance to the mineral subsoil (Bedford et al. 1999), while tundra ecosystems are considered to be more N-limited (Shaver and Chapin 1980, 1995).
270	'i-q-rab'	This may be caused mainly by the different climates in these two areas, which means the cold climate in the northeast need much more electricity for heating than in southeast.
271	'i-q-hnv'	This situation is obvious in the USA, which may be resulted from its high technology of production and demolition.
272	'i-q-hnv'	Hence, the evaluated results in the studied building may be lower than that in fact.
273	'i-q-hnv'	In the uppermost Eiberg section, the reduction in size of P. triassica and its decrease in abundance broadly occur together with major positive $\delta^{18}O$ shifts which might be interpreted as cooling pulses, sea-water chemical changes and/or decreasing pH, as also inferred for benthic foraminifers (see previous paragraph).
274	'i-q-hnv'	Leakage refers to the possibility that net emissions reductions of a project may be lower (or higher) because of increased (decreased) emissions elsewhere that are attributable to the project.
275	'i-q-rin'	Some classes like oak-hickory in Mississippi (productivity below 15.6 m3) have standard error (a precision metric) that is low (6.0–9.1), and might be useful for setting a baseline value for that forest type in that location.
276	'i-q-hnv'	In the stable case, the choice of the number of years will not have much impact on the SRB emissions rate; the steady trend case suggests the use of fewer years to estimate the future emissions rate since including data from too many years back would tend to overestimate the rate; and the scattering case suggests that a greater number of years' worth of data may be necessary to obtain an average representative of the range of changes.
277	'i-q-rab'	The significant correlation between soil EC and N ₂ O _d (Table 3 and Fig. 3b) suggests that rising salinity, which typically accompanies soil acidification (Cao et al. 2004), may be another cause of higher N ₂ O emission.
278	'i-q-hnv'	Lead investigators in the USA (Paustian, 1999) and Canada (McKonkey, 1999) agree that a convincing understanding of the uncertainties involved in estimating carbon flux that considers the many variables of soil type, crop type, management, and climate is not yet available, but they hold that experiments under way may be able to demonstrate an improved modeling capability.
279	'i-q-hnv'	ered to farmers related to their actual carbon sequestration achievements, means of verification and reporting at the level of the farmer may be needed, but indications suggest that the cost of such monitoring and reporting would be excessive.
280	'i-q-hnv'	Carbon sequestration options compared 3.1. Environmental impacts 3.1.1. Soil conservation practices Many agricultural practices can enhance carbon accumulation in soils, but conservation tillage and cover crops may be the most important categories that also o!
281	'i-q-hnv'	Land area under cover crops in the USA is believed to be increasing
282	'i-q-hnv'	Worldwide, the methane sink is believed to average 38 Mt methane each year (798 Mt carbon dioxide equivalent, 218 Mt carbon equivalent), but the possible future reduction in uptake related to agricultural practices may
283	'i-q-hnv'	Importantly, the increase in crop/residue ratios is believed to have slowed in some temperate regions and crop yield growth increasingly means an increase in total biomass (Sinclair, 1998; Hay, 1995).
284	'i-q-hnv'	According to a number of investigations, carbon accumulation under cover cropping may be double that under conservation tillage on an area basis.

285	'i-q-hnv','i-q-rin'	(In actuality, the sequestration trend line may not be so steep if trends in activities other than conservation tillage were considered.) Nonetheless, it is possible that the level of sequestration that would exceed this trend would be quite low.
286	'i-q-hnv'	While these limitation approaches could exclude either a large majority or a small minority of emissions in the USA, the implications may be quite di!
287	'i-q-hnv'	In the long run, the worldwide potential for carbon sequestration in agricultural soils is signi"cant, particularly in the former Soviet Union and in developing regions, although the potential may be minor in parts of Europe. Potential future revenues for agricultural soil sequestration could be sizable in aggregate, if not per land unit. Assuming an emissions permit/trading price of \$50 per tonne (carbon equivalent), applying to annual sequestration of 25 Mt carbon in the USA, the value of agricultural soil carbon would be \$1.25 billion/yr (\$5/h over 40 million hectares). This revenue potential compares modestly with the \$17 billion commodity payments the US Department of Agriculture paid farmers in 1999. Given that parties could o!set sizable emissions in other sources, through o!sets on agricultural soils, if parties agree to include agricultural soils in Kyoto Protocol inventories, it may be necessary to limit credits for the "rst commitment period (2008}2012).
288	'i-q-hnv'	Consequences might be the loss of biodiversity of associated floras and faunas, including the genetic diversity of the species populations (Thompson 1999).
289	'i-q-hnv'	On the whole, deviance explained by the models was relatively low, indicating that there might be other important factors influencing richness pattern not included in this study.
290	'i-q-rab'	An important factor limiting decomposition along this study transect might be the observed low pH values (pH range of 2.4â€"4.4, with lower pH values at higher altitudinal sites).
291	'i-q-hnv'	Recent evidence suggests substantial changes in vegetation including increasing shrub cover and density in the Arctic over the past three decades that may be in response to climate change.
292	'i-q-hnv'	Recent evidence suggests substantial increases in shrub cover and density in the Arctic over the past three decades that may be in response to climate change (Devi et al. 2008; Goetz et al. 2005; Sturm et al. 2001; Tape et al. 2006).
293	'i-q-hnv'	This result implies that the effect of rising temperature on mass balance may be more significant than the influence of increasing precipitation.
294	'i-q-hnv'	This study is believed to be the first attempt to quantitatively evaluate large scale dynamical sea surface temperatures forecasts on seasonal timescales in a reef system with the view for future application in coral bleaching management.
295	'i-q-rab'	This may be simply a statistical artefact due to a small sample.
296	'i-q-rab'	Alternatively, it may be an increased skill due to remote forcing.
297	'i-q-hnv'	This increases the possibility that the remaining â€"unavoidableâ€" climate change can be adapted to, and that the sum of those adaptations may be positive.
298	'i-q-hnv'	Although AMO signal is correlated with the tropical Atlantic SSTG, previous studies suggest that it may be physically independent from the SST changes in the Table 2 Number of years with statistically significant at 5% level positive (PP) and negative (NP) phases for the SST variability indices considered here DJF
299	'i-q-hnv'	Surface oscillation may be an important self-preservation mechanism of mire ecosystems against water-table fluctuations because water stress in Sphagnum is lower when water levels are low (Price 2003), and Sphagnum can escape form submergence when water levels are high.
300	'i-q-hnv'	Wet deposition of organic nitrogen may be a large contribution to total wet deposition of nitrogen.
301	'i-q-hnv'	Moreover, the 15N dilution technique itself (used in these N-fixation studies) may be questioned, since it does not account for other gaseous exchanges which can alter the 15N dilution (NH3, NO2) (Loubet 2000, p. 177).
302	'i-q-hnv'	However, intermittent shading may be significant when considering time-average physiological behavior because individual leaves high in the forest canopy in B2C spend part of the time in bright sunlight and part of the time in shade.
303	'i-q-hnv'	[Note: it is possible that the higher-measured value for albedo in B2C shown in February in Fig.
304	'i-q-hnv'	Arctic tundra soil microbial biomass and organic soil structure may be less sensitive to air temperature freezeâ€"thaw cycles than in temperate agricultural and forest soils (Lipson and Monson 1998; Lipson et al. 2000; Grogan et al. 2004; Walker et al. 2006; Henry 2007; Matzner and Borken 2008).
305	'i-q-hnv'	Instead, arctic tundra soils have very low growing season pools of nitrate and low rates of net nitrification (Giblin et al. 1991; Nadelhoffer et al. 1992) suggesting that although environmental conditions may be favourable for N2O production at thaw, low NO3 availability may restrict denitrification.
306	'i-q-hnv'	In temperate, alpine and sub-arctic systems soil microbes may act as intermediate retainers of the spring flush of N before plant N uptake begins (Zak et al. 1990; Jaeger et al. 1999; Grogan and Jonasson 2003), theoretically promoting tight internal soil cycling of N. Our results suggest that this microbial mechanism of springtime ecosystem nutrient retention may be important, at least in fertilized tundra.
307	'i-q-hnv','i-q-rab'	Our early spring results suggest that this N loss in fertilized tundra is unlikely to be explained by N2O production, and that the missing N may be more likely attributed to N2 or leachate loss.
308	'i-q-hnv'	They argued that if this reflected a true difference in the vulnerability of birds to invertebrates, we might be able to use this â€"relative extinction rateâ€" to determine how many invertebrate species have gone globally extinct.
309	'i-q-hnv'	A large effort to reduce illegal activities in Amazonia was recently initiated by Brazilâ€™s national environmental agency (IBAMA), but its effectiveness may be limited because the planned sweeps were trumpeted in local newspapers prior to being implemented (e.g. Anon. 2001a).
310	'i-q-hnv'	Because future climate may be very different from the observational record a primary tool for assessing changes are large computer models that simulate the Earthâ€™s climate system under different circumstances.
311	'i-q-hnv'	, D N], are not directly interesting we believe that they can nevertheless be useful for diagnostic purposes.

312	'i-q-hnv'	We have indicated that dinoflagellates and other NPP may be more sensitive to changes in nutrients than in salinity.
313	'i-q-hnv'	This abundance of males may be necessary to increase the probability of mating and serves to decrease the latency times (Devreker et al., 2009) in the Laguna Honda population.
314	'i-q-hnv'	Given predicted increases in global temperature, the lifespan of ectotherms may be substantially shortened in the future (Munch & Salinas, 2009).
315	'i-q-hnv'	In this context, the interaction between mortality and developmental variability might be an important mechanism for population persistence (Lopez, 1991) when it faced a scenario of global warming.
316	'i-q-hnv'	Increased productivity, through P fertilizer and legume management, did not significantly increase N ₂ O emissions, or reduce CH ₄ uptake, during this 4 week measurement period, but the lack of an N ₂ O response to rainfall in the No P pasture suggests this may be evident over a
317	'i-q-hnv'	Cross correlation and lag analysis indicated that N ₂ O emissions in the High P pasture may be better correlated with soil environmental variables measured 1 or 2 days earlier.
318	'i-q-hnv'	However, this may be an oversimplification, as the effects from ammonia fertiliser addition can outlast the enhanced
319	'i-q-hnv'	In the Tibetan Plateau of China, enhanced UV-B and temperature are prominent factors and our study may be of significance in understanding the regeneration behavior of subalpine coniferous forests under future climate change.
320	'i-q-hnv'	The map suggests that although the elderly in most eldercare institutions may be spared from a weak or moderate storm surge, they could potentially suffer extensive repercussions when stronger storm surge inundates the area because 48 out of 90 eldercare institutions are located within the Categories 4&5 storm-surge-risk zone.
321	'i-q-hnv'	For those living farther inland, it may be more important to ensure equal access to resources and services during emergencies and nondisaster times by relating policies to larger development planning for the community.
322	'i-q-rab'	Several factors may be responsible for the observed changes in body size.
323	'i-q-hnv'	They simply believe that frightening ignorance may be at stake, and hence seek a more comprehensive evaluation of signs of possible danger.
324	'i-q-hnv'	In terms of problem analysis they are sensitive to the possibility that concentrations that are currently regarded as 'safe' may be severely in error.
325	'i-q-hnv'	This study suggests that cyclic population dynamics of the arrowhead scale are driven by the dynamics of a plant trait that may be triggered by climatic disturbance.
326	'i-q-hnv'	The difference in the population levels between the groves may be ascribed to those in the initial population sizes of the parasitoids and in climate conditions; however, further investigation is required before any definitive conclusion can be drawn.
327	'i-q-hnv'	This issue may be particularly important for precipitation, which appears to be generally more sensitive to the nature and distribution of radiative forcing than temperature is (Hegerl et al. 2007).
328	'i-q-hnv'	The cost of reducing activity or investing in morphological defences could be higher for organisms exposed to severe time constraints, and they may be selected to accept a higher predation risk in order to speed up growth (Dmitriew 2011).
329	'i-q-hnv'	Symbiodinium C2 is believed to be a light-loving symbiont (Ulstrup and van Oppen 2003), and the waters around Magnetic Island are, on occasion, highly turbid.
330	'i-q-hnv','i-q-rab'	Biogenic silica and organic carbon accumulation in the sediments suggest a possible decline in lake production during the 20th century that may be associated with changes in the river discharge regime.
331	'i-q-hnv'	Nitrogen limitation could favour the growth of cyanobacteria, which are isotopically light because they fix nitrogen directly from the atmosphere (d ₁₅ N = 0) (Brenner et al. 1999), as might be the case under reduced runoff conditions and catchment N input.
332	'i-q-rab'	shows that at least for some species the gain in energy intake might be associated with an elevated cost in energy expenditure, and that this cost is mainly associated with higher disturbance frequencies causing birds to spend more time in energetically expensive flight.
333	'i-q-hnv'	Large cold SST anomalies are located in the northern Okhotsk Sea, except around the coast of northeastern Sakhalin, which may be affected by oceanic conditions such as the East Sakhalin Current, which the western boundary current in the Okhotsk Sea (Ohshima et al., 2002).
334	'i-q-hnv'	These findings provide hope that the existing gap between scientific and nonscientific cultures that has helped to exacerbate scientific controversies may be bridgeable.
335	'i-q-hnv'	5.2 Mixing Endogenous and Exogenous Risk For some problems where a SMS may be a policy option, it may be the case that there is underlying risk that is independent of human activity.
336	'i-q-hnv'	For instance, there may be a risk that a population of a threatened species may collapse even in the absence of human interference.
337	'i-q-hnv'	It may be the case that the true optimal path is very close to a SMS, even if it can be established that the SMS is not optimal.
338	'i-q-hnv'	In these more extreme environments, it is possible that lichens will utilise local photobionts, which may also be better adapted to the local environment, rather than restricting themselves to a specific photobiont (Fernaldez-Mendoza et al. 2011).
339	'i-q-hnv'	However, one possibility is that the higher haplotype diversity may be a residual effect of local photobionts accumulating and
340	'i-q-hnv'	In terms of pure optics, monostromatic and optically translucent thalli of delicate eulittoral species might be regarded as highly susceptible to UV damage.
341	'i-q-hnv'	in disease levels should be interpreted with caution, as it is possible that mortality of oysters post-flood contributed to the decline in disease levels if oysters with higher weighted prevalence died at a faster rate than those with lower weighted prevalence (La Peyre et al. 2003).
342	'i-q-hnv'	After many years of considering that strict controls are the only solution, many observers are starting to ask whether increased competition and liberalisation within energy markets might be a less-burdensome policy for reducing pollution.

343	'i-q-hnv'	The following section highlights two factors â€” price differential between sources of electricity and information provision â€” that may be important in determining the level of demand for renewable electricity once the market is liberalised in 1998 and, because of inelasticities on the supply side, the development of the market.
344	'i-q-hnv'	about electricity generation may be increasingly more influential in determining demand than information from Government sources.
345	'i-q-hnv'	Table 4 shows that the proportion of all consumers explicitly choosing to â€”keep down the amount of electricity and fuel usedâ€” appears to be on the rise; as may be the proportion taking actions to improve the environment.
346	'i-q-hnv'	If so, extending NFFO for further rounds may be of considerable service for the future of the renewable electricity market.
347	'i-q-hnv'	If a potential 20 million customersâ€™ beliefs about the value of renewable electricity and free riding form at the same time and remain more-or-less fixed thereafter, and if the appropriate incentives are created, it is possible that demand will be initially too large and later not rise in-line with generating capacityâ€™s growth.
348	'i-q-hnv'	In this regard, we assume that wellbore leakage occurs predominantly along the exterior of the casing, through a zone that consists of well cement that may be degraded or poorly emplaced and the host rock that may be damaged to drilling action.
349	'i-q-hnv'	However, there is no certainty that such predictor variables will comply with the remaining caveats to downscaling.
350	'i-q-hnv','i-m-int'	Within a global perspective, treeline position seems to coincide with a mean temperature in the growing season ranging between 6 and 7C (Kolrner 2003), suggesting that growth processes might be strongly limited below this threshold.
351	'i-q-hnv','i-m-int'	Our results and other authorsâ€™ observations (Elster et al., 1960; Rokneddine, 2004; Jim��nez-Melero et al., 2007) suggest that the thermal tolerance threshold for A. salinus development might be around 25��29��C.
352	'i-q-hnv','i-q','i'	Thus, results suggest that convective gas flow is not the primary transport mechanism although mature P. arundinacea may be capable of building slight pressure differentials resulting in slight CH4 transport during flowering and increased O2 transport to the rhizosphere during the second growth peak and the late growth season.
353	'i-q-hnv','i'	It must be stressed that there may be some uncertainties in these determinations.
354	'i-q-rab'	It might indicate that the net carbon budgets of forest ecosystems in the northeastern China were more sensitive to an increase of air temperature than a change of precipitation.
355	'i-q-rab'	One of the main reasons we believe is the difference in vegetation types (phenology), age, and succession stages between the forests between these two regions.
356	'i-q-rab'	reasons we believe for the higher carbon stock is that large areas of evergreen coniferous forests found in the northeastern China tend to provide carbon sinks all year round.
357	'i-q-rab'	Increasing nitrogen and phosphorus deposition, climate changes, and the regrowth of forests and afforestation, might be possible mechanisms as well.
358	'i-q-rab'	Thus, it is possible that concentrations of up to 400 ��g g���1 in our experiment were inhibitive for reduction of N2O to N2 at 70% WFPS.
359	'i-q-rab'	Taken together, therefore, it is possible that in shoots the reduced DM and N content as well as N concentration with elevated [CO2] alone could be from faster and favorable translocation of photosynthates and N from shoots to grains.
360	'i-q-rab'	Lianas may be favoured by low rainfall: evidence from Ghana Michael D. Swaine �� John Grace
361	'i-q-rab'	Both trends in lianas, one temporal, the other spatial, may be related to water availability, though parallel changes in canopy density (disturbance) may be contributory.
362	'i-q-rab'	The reasons for the high sensitivity to rainfall may be that lianas have an especially effective water-transport system, with deep roots, large xylem vessels and a mechanism to avoid cavitation of water.
363	'i-q-rab'	It appears, then, that there are two processes which may be
364	'i-q-rab'	However, the Bering Sea is also undergoing a northward biogeographical shift that may be climate related (Overland and Stabeno 2004; Grebmeier et al. 2006; Bluhm and Gradinger 2008; Grebmeier 2012).
365	'i-q-rab'	This finding may be related to size limitations of younger Arctic cod preying on benthic amphipods relative to small copepods, or a lower capability for younger Arctic cod to migrate vertically as adult predators can.
366	'i-q-rab'	Another factor influencing our findings may be the presence of the persistent cold pool (-1.8 to -1.4 ��C) of bottom water south of SLI where the Bering flounder were primarily collected during our sampling (Cui et al. 2009).
367	'i-q-rab'	Including the water level in a multiple regression analyses did not improve the significance, most probably due to a high autocorrelation between temperature and water level.
368	'i-q-rab'	The different observations may be due to the different climate zone and soil condition.
369	'i-q-rab'	It is also possible that the instrumental record may be incomplete or inaccurate due to the sparse observation network in the early instrumental period, as stated by Kam-biu Liu (2001).
370	'i-q-rab'	In turn, these circulation anomalies enhance the seasonal flows over the equatorial and South subtropical Indian Ocean and may be responsible for the appearance of the initial cold SST anomalies over the eastern Indian Ocean due to the enhanced upward surface latent heat flux in relation to increased surface wind speed in both areas and reduced incoming shortwave radiation associated with enhanced convection in the eastern part of the Indian basin (Fig. 13a).
371	'i-q-rab'	Whereas a proportion of this inefficiency may be attributable to farm management, another part may be due to factors that are beyond the control of the farm manager.
372	'i-q-rab'	The contradictory results may be due to a variety of reasons, including seedling age, duration of treatment, seed sources, and environmental and genetic differences.

373	'i-q-rab'	One possible reason may be that the effects of light by itself on seedling growth were much greater than those of warmed temperature under
374	'i-q-hnv'	Carbon allocation may be more important than carbon acquisition (photosynthesis) in the acclimation to environmental changes (Domisch et al. 2002).
375	'i-q-rab'	This may be due to chronic photoinhibition and damage to PSII in the <i>A. faxoniana</i> seedlings caused by full light (Mitchell and Arnott 1995).
376	'i-q-rab'	Enhanced nutrient limitation for the graminoid species in the mixtures might be explained by the lack of response of the grasses to the nutrient applications.
377	'i-q-rab'	This linear correlation and the very light values of $\delta^{18}O$ might be related to meteoric diagenesis (Marshall 1992).
378	'i-q-rab'	The short-term, lighter $\delta^{13}C_{carb}$ values correlate with increased TOC values, and this might be related to partial recycling of OM.
379	'i-q-rab'	It might be related to a temperature decrease (i.e., cooling pulse) and/or seawater chemistry change.
380	'i-q-rab'	This negative excursion observed both in the organic and inorganic carbon record has been observed in several sections, from different environmental settings, and may indicate a perturbation of the global carbon cycle in coincidence with the end-Triassic mass extinction (McRoberts et al. 1997; Ward et al. 2001; Palfy et al. 2001; Guex et al. 2004; Galli et al. 2005; Ruhl et al. 2009).
381	'i-q-rab'	At Tiefengraben, Kendelbach and Schlossgraben the absence of hyaline foraminifera and the occurrence of Trochammina (perhaps enhanced by selective dissolution, as mentioned previously) at the end-Triassic may indicate stressful conditions (Figs. 7c, 8, 9b).
382	'i-q-rab'	These "pre-perturbation" benthic-planktonic linkages might indicate a well-structured trophic chain and good recycling of nutrients (as also attested by type II organic matter), in an oxygenated and mixed water column.
383	'i-q-rab'	Alternatively, Prinsiosphaera's extinction might be linked to changes in sea-water chemistry.
384	'i-q-rab'	In the same way, the average change of "calcispheres" diameter might indicate undercalcified specimens linked to the cooling episode, and maybe also to low pH conditions.
385	'i-q-rab'	Volcanogenic SO ₂ emissions might be responsible of the establishment of the major cooling episode, seawater chemistry changes and/or low pH conditions at the end of the Triassic stage.
386	'i-q-rab'	Table 4 Classification of regional baseline methods, by land use, for estimating land use change and carbon stock change Quantifying land use change (examples) Avoided deforestation Undisturbed Model that relates rate of forests deforestation to predisposing and driver variables, logging may be the most likely cause of deforestation.
387	'i-q-rab'	Model that relates rate of deforestation to predisposing and driver variables, agriculture and fuelwood demand may be the most likely causes of deforestation Examples: Brown et al.
388	'i-q-rab'	Inflexion or break points may be caused by policy changes such as resettlement policies for forest lands in Kalimantan, Indonesia in the 1980s, the passing of a forest conservation act in India in 1980, or of a logging ban in China in 1999.
389	'i-q-rab'	These changes were most probably due to the stimulatory effect of ammonium-based fertiliser application on nitrification potential (Xue et al. 2006; Zhong et al. 2007).
390	'i-q-rab'	The increase of N ₂ O emission during heterotrophic nitrification may be attributed to the increase in heterotrophic nitrification rate.
391	'i-q-rab'	The reason for this suppression may be food limitation due to the mistiming that acts through both parental feeding and nestling competition.
392	'i-q-rab'	E Lower residue yield (may be due to fertilizer inputs or genetic improvements) E Change to crop types with lower biomass levels E Lower lignin content crops E Longer fallow
393	'i-q-rab'	For instance, reduced carbon accumulation in a given year may be related to an unusually warm fallow period rather than to a failure to implement carbon sequestration practices.
394	'i-q-rab'	However, this lack of agreement may be due to differing sampling and incubation protocols between the two studies.
395	'i-q-rab'	This observation might be explained by the water-energy-hypothesis (Rosenzweig and Abramsky 1993), since the drier, lower part of the island within the habitat of succulent scrub perennial plant richness was found to positively correlate with mean annual precipitation (Otto et al. 2001).
396	'i-q-rab'	This chemical complexity can be inherent to the original plant material, arise from newly formed microbial residues, or may result from condensation and complexation
397	'i-q-rab'	Our wet sedge soils also had high N mineralization potentials, which might be because of the combination of a high-quality organic matter input (dead sedges) and the very high soil moisture levels for much of the summer in sedge communities, preventing the decomposition of this high-quality organic matter.
398	'i-q-rab'	The high N mineralization potential in the tall birch soil may be partly attributed to higher nutrient availability in tall birch patches due to nutrient influxes in stream and seep flow (Chapin and Shaver 1996).
399	'i-q-rab'	The model does not appear to capture these relatively high anomalies, which may be possibly due to local
400	'i-q-rab'	While this lack of integration may be due to good reasons, mainly to do with not mixing science and policy, it remains problematic for risk management, where it is generally recommended that the steps in risk management be assimilated into a single (though possibly many layered and diverse) process (McCarty and Power 2000; AS/NZS 2004).
401	'i-q-rab'	At present a whole cohort of Forest Red Gum, <i>Eucalyptus blakeleyi</i> , is dying throughout central and eastern Victoria, ostensibly through insect attack (Gibbons and Boak 2002) but a warming climate and persistent drought conditions may be an underlying cause.
402	'i-q-rab'	Thus, it can be assumed that the combined effect of temperature, food supply and food composition might be responsible for the high diversity in the warmer set-ups.

403	'i-q-rab'	With the onset of the spring bloom, <i>L. oviformis</i> was displaced by <i>B. comatum</i> , and since data from the spring bloom period showed high grazing rates of ciliates preying upon bacteria, this might indicate a bacterivore feeding habit of <i>B. comatum</i> .
404	'i-q-rab'	Water soluble organic nitrogen was also found in significant amounts in the Amazon basin and may be related to biomass burning (Zhang et al. 2002).
405	'i-q-rab'	These differences may be explained by the more favourable warmth and radiation conditions in the southern part of Bavaria (BayFORKLIM 1996).
406	'i-q-rab'	In the present case, the increased soil CO ₂ efflux mentioned above may be linked directly to the loss of labile C. Additionally, some studies have reported that warming has reduced the root biomass in boreal soils (Bergner et al. 2004; Bronson et al. 2008).
407	'i-q-rab'	The warming-induced decline in labile C and N pools may be linked to the enhanced soil surface CO ₂ efflux.
408	'i-q-rab'	It is likely that the night-time difference is the result of poor relative calibration between two different types of Ta sensors, while the greater discrepancy during the day may be due to solar heating of the metal screen in which the Ta /RH probe is housed.
409	'i-q-rab'	This early spring period may be associated with nitrogen (N) and carbon (C) loss from soils as leachate or as trace gases, due to the detrimental impact of soil freeze-thaw cycles and a developing active layer on soil microorganisms.
410	'i-q-rab'	Strong positive associations with soil ammonium and nitrate pools suggest that N ₂ O production may be due to both nitrification and denitrification in the organic soils we investigated.
411	'i-q-rab'	This lack of extinction in the Atlantic forests and elsewhere where there has been extensive forest conversion may be due both to the often fragmented nature of this conversion, that it can take many years for extinction inducing factors to take effect (Diamond 1972; Tilman et al. 1994; Brooks and Balmford 1996; Brooks et al. 1999), that biodiversity hotspots too have their own hotspots (Hugall et al. 2002; Carnaval et al. 2005) and that hotspots of endemism do not necessarily coincide with hotspots of threat (Orme et al. 2005).
412	'i-q-rab'	Additionally, chronologies based on density banding alone may be biased by the occasional missing or doubling of high-density bands within some years (double high-density band: dHDB) seeming to be related to thermal stress, bleaching and diseases (Barnes and Lough 1993).
413	'i-q-rab'	One possible explanation might be that dHDBs are rarely identified by studies if no proxy data are compared to X-radiographs.
414	'i-q-rab'	Such a biogeographical distribution may indicate a longer-term cooling or warming to have taken place upon global change during growth, causing SSTs to shift away from an initially more equilibrate site of larval settlement.
415	'i-q-rab'	Occasional Caspian Sea taxa such as <i>Spiniferites cruciformis</i> type C of Marret et al. (2004) (Fig. 4-22), <i>Caspidinium rugosum</i> (Fig. 4-19) and <i>Impagidinium caspiense</i> (Fig. 4-20) may indicate periodic flooding from the Caspian-Aral Sea basins.
416	'i-q-rab'	These differences might be due to the different diets used or to a positive correlation between female's size and clutch size, such as this author found in this species (Rokneddine, 2004b).
417	'i-q-rab'	The striking distribution in pulses of these times when they are plotted against their cumulative proportions (Fig. 3) might be due to the selection of particular phenotypes.
418	'i-q-rab'	1 Introduction Animals are exposed to various environmental factors that may influence their body size, and many studies have documented changes in body size of birds and mammals that might be related to changed environment. Communicated by Roland Brandl Y. Yom-Tov (&)† S. Yom-Tov Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel e-mail: yomtov@post.
419	'i-q-rab'	These data may indicate a trend of an increase in sea temperature along the southern part of the Norwegian coast.
420	'i-q-rab'	The irregular linear feature and funnel shape may be caused by irregular melting within the ice.
421	'i-q-rab'	The dense chaotic return zone may be closely related to the density or temperature of the ice body.
422	'i-q-rab'	9b shows a decrease in the dielectric constant with depth, which may result from the increasing content of debris and sediment in cracks near the glacier surface and indicating that the amount of debris and sediment in cracks decreased with increasing glacier thickness.
423	'i-q-rab'	13b shows several continuous 1-2 signal pulses, which may result from the increasing content of debris and sediment in cracks near the glacier surface, and the waveform gradually fades after 150 ns.
424	'i-q-rab'	reflection (i.e., the location approaches the bed interface and the width is large), we infer that the signal may be a reflection of a subglacial slow-speed channel.
425	'i-q-hnv'	The GPR survey results suggest there may be intense ice body melting in some areas of the glacier.
426	'i-q-hnv'	Thus, it appears that bottom-up forces may be more important than top-down control by the parasitoids in the post-transient phase of this system.
427	'i-q-rab'	The increase in <i>Coccobius</i> densities may be due to the predicted rise in the winter temperatures that overrides the rainfall effect, while that of the <i>Aphytis</i> densities may be due to the predicted increase in summer temperatures as well as that of the host scale population.
428	'i-q-rab'	This may be caused by the deposition of carbonates in the glacial period as shown below.
429	'i-q-rab'	Although the SCEs-as are the main products of phytoplankton Chl-a, a little difference between Chl-a and SCEs-a patterns may be attributed to the complex degradation processes of SCEs-a.
430	'i-q-rab'	22 cal kyr BP-present may be attributed to the stagnation of lake water.
431	'i-q-rab'	It allows almost up to the us to use the observed evolution of GÅ°tÅž present, before substituting this with model simulations, in which the rate of global warming might be biased by model errors (see Sect. 3.2 below).
432	'i-q-rab'	Its rarity at Magnetic Island suggests geographical partitioning of <i>Symbiodinium</i> communities that may be driven by environmental conditions.
433	'i-q-rab'	While each of these changes is suggestive of warmer climatic conditions, differences in timing suggest that additional local or regional factors, or multiple forcings may be at work.

434	'i-q-rab','i-q-hnv'	Since there is no indication of sediment disturbance in either the varve record or 137Cs profile, and an upward migration of 137Cs is highly unlikely, it is plausible that the 210Pb results underestimated the sediment's true age in the uppermost 3 cm.
435	'i-q-rab'	This may be due to differential scavenging and post depositional remobilization processes, which can influence the distribution of radio nuclides in arctic lake sediments (Wolfe et al. 2004).
436	'i-q-rab','i-q-rin'	Because N limitation does not appear to be a factor in Sanagak Lake, diatom and d15N responses to atmospheric N loading may be expected to be minor due to relatively high watershed nutrient fluxes and relative N and P nutrient levels.
437	'i-q-rab'	In contrast, isothermic lakes displayed more subtle changes, suggesting enhanced thermal stratification may be an important mechanism for expanding planktonic communities (Sorvari et al. 2002; Rul'hland and Smol 2005).
438	'i-q-rab'	The prolonged dominance of Aulacoseira taxa in the planktonic assemblage prior to the 1980s may be attributed to the fact that certain species in this genus are common in oligotrophic waters and often abundant in large, deep lakes with sufficient wind-induced turbulence to sustain these relatively heavy taxa in the photic zone (e.g., Lemmen et al. 1988; Wilson et al. 1993; Wolfe 1994; Sorvari and Korhola 1998; Rul'hland et al. 2003b; Rul'hland and Smol 2005).
439	'i-q-rab'	This might indicate that the ecosystem is storing and releasing energy on a scale of several days.
440	'i-q-rab'	Some of these discrepancies may be due to meteorological input data (especially wind data) recorded on the shore not being representative of actual meteorological conditions over the lake.
441	'i-q-rab'	The contrast between the predictions of more rapid warming in some monomictic lakes compared to our lakes may be due to differences in the seasonal distribution of atmospheric warming, differences in lake mixing regimes, changes in inflow volumes and temperatures not considered in
442	'i-q-rab'	It is worth noting that the wintertime Z700 anomalies over northern Eurasia are much smaller than those in late autumn (Fig. 6), which may be due to the strengthening of the Aleutian Low in winter.
443	'i-q-rab'	However, the level of selectivity is variable among species and may be related to the ability of some (e.g. B. frigida) to colonise a wider range of habitats.
444	'i-q-hnv'	Thus, potentially stimulating effects of N addition on N mineralization from decomposing litter may be counteracted by interactions with secondary compounds in that litter, leading to chemical immobilization of the added N (e.g. Pastor et al. 1987; Hobbie and Vitousek 2000).
445	'i-q-rab'	Thus, an important reason for the observed interspecific variation in the response of litter decomposition to N additions might be that the phenolics fractions differed among species.
446	'i-q-rab','i-m-int'	On the contrary, the denitrification of nitrate might be responsible for the second peak, which occurred approximately 40 days after rice transplantation.
447	'i-q-rab','i-q-hnv'	This may indicate an advantage of direct measurement of 15N-N2O rather than reliance on inhibitors that may also affect denitrification.
448	'i-q-rab','i-q-hnv'	Several studies have suggested that this seasonal phase locking of the SST anomalies in the Southern Hemisphere may be attributed to the shallowing of the mixed layer during summertime in such way that the latent heat flux and shortwave radiation anomalies influence the SST anomalies more efficiently during this season (Suzuki et al. 2004; Huang and Shukla 2006, 2007; Morioka et al. 2009).
449	'i-q-rab','i-q-hnv'	The upward trend in ZW3 occurrence during recent decades may be partly an artifact due to the introduction into the reanalyses of satellite information in the late 1970s, but parts of this signal seems also robust (Gillett and Thomson 2003).
450	'i-q-rab','i-q-hnv'	However, this assumption might be a methodological artefact (i.e. limited time points over a long period are likely to result in a smooth linear regression, whereas a higher sample frequency might show a different curve shape, i.e. an asymptotic increase with an initially steeper increase in headspace concentrations) and thus could be misleading.
451	'i-q-rab','i-q-hnv'	These different conclusions may be due to differences in models, location and scale, and require further investigation.
452	'i-q-rab','i-q-hnv'	Thus, it is possible that these latter types are more specific to the local situation and may better reflect the associated health effects.
453	'i-q-rab','i-q-hnv'	This may be caused by mixing of the ascending water with colder fluids from the relatively shallower carbonatic aquifer, probably occurring throughout the layer thickness.
454	'i-q-rab','i'	Part of these differences might relate to different rates of adaptation to occupy these relatively new habitats, but the possibility that they might be caused by species-specific differences in perceived predation risk and vulnerability to disturbance means that we should be cautious in applying general predictions.
455	'i-q-rin'	Given the challenges that governments face in dealing effectively with the pace and complexity of scientific innovation at present, much less the longterm impacts, global issues such as climate change suggest that the redesign of governance may be the most critical endeavor for the long-term survival of human civilization.
456	'i-q-hnv'	Dorrepaal et al. (2005) found that there are large differences in litter quality between functional groups of plants and therefore concluded that on the long-term changes in relative abundance of plant functional groups might be more important than the direct impacts of environmental change, like N deposition.
457	'i-q-rin'	Such optimization of end-of-life materials disposal may become increasingly important in the future.
458	'i-q-rin'	Flow regimes and water depths in the Delta might be expected to change in the future, as snowmelt arrives earlier and sea level rises.
459	'i-q-rin'	Renewable Energy the renewable energy target of 20% by 2010 and 80% by 2020 may be impossible to achieve.
460	'i-q-rin'	Therefore, the trimming of vascular plants may be an effective technique for examining the reaction of Sphagnum to desiccation.
461	'i-q-rin'	These estimates have been criticised (Feeley and Silman 2008) and an improved assessment of the extinction risks of Amazonian plant species suggests that by 2050 land use change may result in 12-24% reduction of forest habitat and that this may result in 5-9% species being committed to extinction (Feeley and Silman 2009).

462	'i-q-rin'	By analogy with paleoclimatic reconstructions, it is possible that such a MOC perturbation would have an impact on future climate, e.g.
463	'i-q-rin'	In particular, the further reduction in existing limited water resources combined with an increase in temperature may result in higher impacts on agricultural crops in the Mediterranean area than in other regions.
464	'i-q-rin'	However, a change in diet composition related to a change in availability of cod and salmonids may be expected.
465	'i-q-rin'	In the future, with the ablation of glaciers, the central positions of intense ablation may be disconnected and there may be a dead zone at the terminus.
466	'i-q-rin'	8, and it is possible that the rough upper surface of the actual tunnel leads to an irregular measurement image.
467	'ai'	Therefore, the central positions of intense ablation may be disconnected and there will be a dead ice zone at the terminus owing to the insulation of the ice body by debris (Fig. 10).
468	'i-q-rin'	Since amphibian growth and development is highly dependent on hormonal regulation (see, e.g., Denver 2009), it is possible that similar mechanisms operate in amphibians, with female hormonal
469	'i-q-rin'	The relatively high N:P ratios may lead to consistently high isotopic fractionation of N during uptake by phytoplankton, and may result in relatively slight covariation with productivity.
470	'i-q-rin'	Although the community in this study did not frame their arguments for park improvement explicitly around the heat mitigation benefits demonstrated in our research, this service may become more important in the future as the trend of increasing temperatures continues.
471	'i-q-rin','i-q-hnv'	We hypothesise that <i>P. arundinacea</i> is effective in C sequestration although the high percentage of aerenchyma may result in high levels of CH ₄ transfer from anoxic soil layers to the atmosphere bypassing oxic layers and thereby reducing soil CH ₄ oxidation potentials.
472	'i-q-rin','i-q-hnv'	Are the life cycle CO ₂ e emissions for tree production greater than CO ₂ stored by a tree planted in an urban environment? If so, large-scale tree planting initiatives may result in net CO ₂ emissions, not storage, even during the life of the tree.
473	'i-q-rin','i-q-rab'	Consequently, <i>R. alba</i> might be less dependent on the applied nutrients for its growth and therefore did not show a response to the added N and P. The N:P ratio in the leaves, however, differed among the nutrient treatments, suggesting that <i>R. alba</i> did use the nutrients to some extent.
474	'i-q-rin'	Secondly, the negative effect may be that temperature rise can increase the consumption of water and bring on a water deficit in some biomes (Liang and Maruyama, 1995).
475	'i-q'	The periplasmic nitrate reductase involved in the first reduction step during aerobic denitrification is not inhibited by O ₂ , unlike the membrane-bound nitrate reductase of anaerobic denitrification (Carter et al. 1995), and so it is possible that aerobic denitrification was occurring at 20% WFPS.
476	'i-q'	However, with our methodology, any coupled heterotrophic nitrification–denitrification would have been attributed to heterotrophic nitrification, and thus it is possible that this occurred in the 50% WFPS treatment.
477	'i-q'	limited soil N availability is believed to cause photosynthetic acclimation and subsequent inhibition of growth during the period from panicle initiation to maturity (Kim et al. 2001, 2003a; Weerakoon et al. 2005; Yang et al. 2008).
478	'i-q'	It is not known what these root attributes may be.
479	'i-q-rab'	In tropical lianas it may be that an active process associated with the abundant xylem parenchyma is especially well-developed, and thus protects against hydraulic failure.
480	'i-q-hnv'	There seems little doubt that they are indeed favoured by dry conditions (within limits), and given the widely accepted prediction that drought will increase in much of the tropics, they may be one example of a negative feedback that protects the tropical forest against the immediate impact of a reduction in water supply.
481	'i-q-rab','i-q'	While it might be true that the causes of some of society’s most vexing problems may be society itself, one reaches a point where that argument does not provide solace.
482	'i-q'	However, it remains unclear how typhoon activities are changed by large weather systems such as El Niño.
483	'i-q-hnv'	In these ecosystems, potential effects of global change might be of enormous importance, as the carbon sink capacity may be reduced or even changed into a carbon source (Bridgman et al. 1995).
484	'i-q-rab'	For while it may be plausible to argue that consumers face credit rationing, or have shorter or longer time horizons based on their
485	'i-q'	The Parties to the Framework Convention on Climate Change appear to be moving towards some protocol involving greenhouse gas emissions reductions, although the exact outcome of the negotiations that will take place in Kyoto in December 1997 remains unclear as of this writing.
486	'i-q-hnv'	Guidance by GHG programs also may be needed on how projects should undertake the stratification of complex
487	'i-q'	Aerobic denitrification may also be occurring, although the mechanism underlying this process remains unclear.
488	'i-q-hnv'	However there may be other reasons that still remain unclear.
489	'i-q'	However, the relative importance of exudation and root death on the amount of carbon sequestered in the soil under field conditions remains unclear (Rees et al. 2005).
490	'i'	Human-induced climate change will almost certainly result in increases of MAT, MAP and MAPER.
491	'i'	Although the mechanism involved in pH control of denitrification rate remains unclear, numerous laboratory and field measurements have demonstrated that the ratio of N ₂ O:N ₂ increases with the decrease of soil pH (Stevens et al. 1998; Simek and Cooper 2002; Dannenmann et al. 2008).

492	'i-q-hnv'	Therefore, in our study, the N ₂ O production from either nitrification or denitrification process may be potentially overestimated in acidic soil at pH 4.3. In summary, soils used for intensive vegetable cultivation exhibited N ₂ O emission rates significantly higher than those found in soil used for more traditional grain crops when tested over 96 h. Surprisingly, denitrification is an important contributor to N ₂ O emission under aerobic conditions, and the N ₂ O emission rates via heterotrophic nitrification were higher than those via autotrophic nitrification in vegetable soil.
493	'i'	Almost certainly the latter statement reflects estimates of extinctions which derive from the species–area relationship and loss of tropical forests (e.g. Myers 1988; Wilson 1988) and these are discussed later.
494	'i-q-hnv'	Further concern for the loss of species due to forest loss came from an analysis of the impact of expansion of road networks and other infrastructure in the Amazon that suggested as little as 40% might be unconverted forest by 2020 (Laurance et al. 2001) (but note even the areas termed heavily degraded or deforested are still mostly covered with forests—see also Chazdon et al. 2009 for discussion of regrowth forests).
495	'i'	More important, by using such a method, policy advisers and scientists will most probably stay unaware of the fact that the pluriformity of basic views plays the key role in the controversy or problem they want to solve.
496	'i'	Concern for the environmental implications of energy use has led Government and analysts to wonder what policies might best minimise atmospheric and air pollution without excessively burdening economic activity.
497	'i','i-q-hnv'	It is clear that more field experiments are needed, but several studies suggest that growth compensation may be more widespread among birds than we currently know (e.g. Bize et al. 2003; Brzek and Konarzewski 2004; Hegyi and Toth 2007; Honarmand et al. 2010).

G.2 Corpus d'étude - partie 2 (200 phrases)

ID	Catégories	Phrase
501	'i-q-hnv'	If spatial information on daily topographic global radiation is not available, this approach may be a promising alternative for predicting spatial and temporal patterns of soil temperature in terrain regions.
502	'i-q-hnv'	Hence, it is hypothesized that global warming will give a more significant positive feedback on the global carbon cycle in a clear-cut area than in an uncut area at a local scale and in tundra than in temperate forests in a global scale since LAI is lower in tundra ($\approx 2 \text{ m}^2 \text{ m}^{-2}$) than in temperate forest ($5\text{--}12 \text{ m}^2 \text{ m}^{-2}$) (Whittaker and Likens, 1975).
503	'i-q-hnv'	Recent experiments of Leadbetter and Bresnak (1996) also suggest oxygen tolerance of methanogens.
504	'i-q-hnv'	It is my belief that 50 years from now, the majority of the world's habitats and species will either be destroyed or on their way to recovery from a degraded state.
505	'i-q-hnv'	Including activities related to land use change and forestry in the international climate change agreement may stimulate new incentives for soil-conservation practices domestically.
506	'i-q-hnv'	If farmers are compensated for activities, rather than for measured changes in carbon stocks, it may not be necessary to establish extensive control sites.
507	'i-q-hnv'	No-till practices may also reduce the rate of methane consumption in soils (Cole et al., 1993).
508	'i-q-hnv'	Experience with other emissions sources suggests that — in the absence of sizable national investments in monitoring — national inventory accuracy does not necessarily increase significantly over a decade's time.
509	'i-q-hnv'	Although this has to be confirmed in future work, for now we assume that the productivity trends in the Okhotsk Sea and our two study regions are parallel.
510	'i-q-hnv'	Removing this difference may well be an intractable problem because it is probably impossible to realistically reproduce the wind fields of a real forest in an enclosed environment.
511	'i-q-hnv'	Verification with direct field measurements is needed.
512	'i-q-hnv'	Moreover, it may prove difficult to accurately track and credit any additional carbon benefits that might be gained off-site through harvested wood or cardboard and paper products (e.g., Blum et al. 1998; Mikales and Skog 1997; Marland et al. 1997).
513	'i-q-hnv'	Investors in this category might include power utilities that find it costly to quickly modify existing power-generating technologies.
514	'i-q-hnv'	The carbon benefits from forest conservation may be altered by the method of calculating carbon credits chosen by the Parties to the Climate Convention.
515	'i-q-hnv'	The assumptions by some authors (e.g. Stavins and Newell 1998) that the carbon discount rate is similar to the financial discount rate and that global warming impacts are linear over time are highly questionable.
516	'i-q-hnv'	We present a method for identifying and selecting projects that may improve the attractiveness of forest-based CDM for local, national, and international stakeholders.
517	'i-q-hnv'	Thus, the sustainability equation, and its energy or pollutant emission equivalent, suggest a multiplier effect between population, economic welfare, and emissions or resource intensity.
518	'i-q-hnv'	The abundance of coal is likely to place an upper limit on all fossil fuel prices at the synthetic fuel cost for the foreseeable future.
519	'i-q-hnv'	A more equitable sharing of world income and resources is likely to be a prerequisite for long-term sustainability.
520	'i-q-hnv'	Carbon dioxide released into the atmosphere from the burning of fossil fuels is thought to persist for around a 100 years.
521	'i-q-rab'	Contradictory findings suggest that no-till may involve lower nitrous oxide emissions than conventional tillage (Li et al., 1996), or higher nitrous oxide emissions (Cole et al., 1993).
522	'i-q-rab'	The ratios of Corg/Cinorg were usually significantly greater than one in both regions, suggesting that preferentially greater organic carbon from cytoplasm than skeletal inorganic carbon was exported from the surface layers.
523	'i-q-rab'	It is likely that the night-time difference is the result of poor relative calibration between two different types of Ta sensors, while the greater discrepancy during the day may be due to solar heating of the metal screen in which the Ta/RH probe is housed.
524	'i-q-rab'	Nonetheless, Fig. 6d shows that (σ_w/u^*) , the normalized standard deviation of w , is between 4 and 6, suggesting very stable conditions even at 3m height.
525	'i-q-rab'	Perhaps this is because the EC system is in dense vegetation at this height, and the air prefers to flow through gaps elsewhere in the canopy.
526	'i-q-rab'	Presumably, this is because the amount of moisture provided has been optimized to maintain a small but finite drainage from the soil column.
527	'i-q-rab'	Thus, an intensified southwest palaeomonsoon apparently transported large quantities of moisture to the eastern Sahara, producing the necessary conditions for the persisting existence of freshwater in the West Nubian Palaeolake Basin.
528	'i-q-rab'	Their extent is probably not adequately reflected in the model pattern, as once again outcropping yardangs imply a greater extent.
529	'i-q-rab'	This halite must be the result of recent efflorescence, as the oxygen isotope and element ratios Sr/Ca and Mg/Ca do not support these values.
530	'i-q-rab'	High element ratios that do not correlate with high $\delta^{18}\text{O}$ content suggest that high levels of salinity were the result of leaching of preexisting salt crusts — e.g., N94/P243 and A859 — or the dilution of a relict brine (Gasse et al., 1987).
531	'i-q-rab'	This stability is corroborated by the parallelism of the $\delta^{18}\text{O}$ curves with the $\delta^{13}\text{C}$ curves, which is only disturbed in the upper part of sections N94/P243 and A859 when major changes in the $\delta^{13}\text{C}$ values occurred towards more negative values, probably indicating a shift towards a larger influence of terrestrial C4 plants and shallowing lake levels.

532	'i-q-rab'	The average sedimentation rate for the sequences shows values between 0.8 and 1.1 mm/year, indicating that the carbonate date 3805 ± 65 years BP is probably much too young as the sediment accumulation rate would increase to > 1.9 mm/year.
533	'i-q-rab'	The lake's stability can only be interpreted as a result of the stable climate without persisting droughts.
534	'i-q-rab'	Regarding the importance of surface water to the lake's water budget, the extremely low content of clastics in the lake deposits must be caused by filtering of runoff through peri-lacustrine vegetation.
535	'i-q-rab'	The low clastic content may also reflect low rates of erosion due to the extensive vegetation cover (Ritchie et al., 1985; Neumann, 1989; Pachur and Hoelzmann, 1991; McKenzie, 1993).
536	'i-q-rab'	Alternatively, substantial variation among sites at a common latitude may also mask a subtle, but potentially important, latitudinal gradient.
537	'i-q-rab'	Regardless of C-input, the profiles showed that dissolved CH ₄ in the floodwater (0 cm depth) was negligible, presumably because CH ₄ diffusing into the floodwater from the soil was rapidly lost by volatilization and oxidation.
538	'i-q-rab'	But the measured concentrations were less than the standards, presumably because of errors either in the standard concentration calculated with Eq. (2) or in the degassing of the sampled solution calculated with Eq. (1).
539	'i-q-rab'	This might be due to exclusion of permafrost soils in the disposition of the Overexploitation Syndrome, relatively low biomass and the missing influence of mining activities (Katanga Syndrome) in the analysis.
540	'i-q-rab'	These differences may have their origin in analytical methodology and sampling site location.
541	'i-q-rin'	Nonetheless, a stable or even (in many countries) declining population within the next century may allow the survival of considerable biodiversity.
542	'i-q-rin'	However, there is every reason to believe that aggressive education campaigns will continue to win the hearts and minds of increasing numbers in future generations.
543	'i-q-rin'	Therefore, if present trends continue and intensify, carbon sequestration in agricultural soils would be most relevant during the first few commitment periods of the Kyoto Protocol.
544	'i-q-rin'	Human-induced climate change will almost certainly result in increases of MAT, MAP and MAPER.
545	'i-q-rin'	Under such conditions, the geographic distribution of current ecosystems will likely shift westward to maintain an adequate water balance.
546	'i-q-rin'	This volume of erosion could have serious impacts on the capacity of the reservoir, cause damage to water transport and filtration infrastructure, and will eventually necessitate the dredging of the reservoir and/or the addition of supplementary reservoirs.
547	'i-q-rin'	The NAPAG working group believes that a reduction of more than 20% in UK energy demand is achievable over the next 20 years with readily available, existing technologies.
548	'i-q-rin'	It is anticipated that the new Renewables Obligation, exemption under the Climate Change Levy, increased renewables R&D funding, and energy market reforms, will enable these new energy technologies to compete more effectively.
549	'i-q-rin'	On engineering grounds the IEE believes that only some 7% of electricity demand could be met from renewables by 2010.
550	'i-q-rin'	Models predict that forests of this region, especially Scots pine forests, will continue sequestering C in response to predicted climate change and increased atmospheric CO ₂ (Kellomäki and Kolström, 1994; Karjalainen and Kellomäki, 1995; Karjalainen, 1996).
551	'i-q-rin'	The positive association between total C content and warmer, moister climates at southern latitudes are consistent with simulation results suggesting that C sequestration rates in northern latitudes are likely to increase in response to predicted climate changes.
552	'i-m-int'	The run terminates when the slow soil organic matter pool reaches steady-state which may require up to 3000 simulation years.
553	'i-m-int'	From the results of in vitro methane emission from Japanese paddy soils in a laboratory, Koyama (1963) first estimated the annual global contribution of methane to be approximately 190 Tgy^{-1} .
554	'i-m-int'	However, there is considerable uncertainty in the estimated values of methane emission, varying widely from 20 Tgy^{-1} to 100 Tgy^{-1} .
555	'i-m-int'	The current value is approximately $0.5\% \text{ y}^{-1}$ (Steele et al., 1992).
556	'i-m-int'	Recent projections of agricultural soil carbon accumulation in the United States center on an estimate of approximately $17\text{--}39 \text{ Mt carbon/yr}$ (Lal et al., 1998), and an estimate at the high end of this range in Canada of $\sim 40 \text{ Mt carbon/year}$ (Dumanski et al., 1998; Lindwall, 1999).
557	'i-m-int'	These costs are high given that the potential value of these carbon offsets may be only \$1.5-\$3.0 billion assuming a value of \$50-\$100/tonne carbon applying to 30 Mt C sequestered.
558	'i-m-int'	Generally, elevations are between 500 and 1200 m.
559	'i-m-int'	Approximately 20% of the west portion has an MAP less than or equal to 300 mm, 70% of the region has an MAP less than or equal to 600 mm, and 30% has an MAP greater than or equal to 600 mm (Fig. 3(c)).
560	'i-m-int'	The S.D. of MAPET ranges from 10 to 80 mm, whereas MAPET ranges from approximately 550 to 700 mm year ⁻¹ .
561	'i-m-int'	Radiation absorption (approximately 50%) by the B2C glass and space frame is high in comparison with the 10–20% absorption reported for glass greenhouses (Mistriotis et al., 1997) because of the need for (6 cm diam) pipe struts to support the (12mm thick) glass roof and walls.
562	'i-m-int'	The mean 24-h emission rate \pm its standard error for the 14 days of measurement was $0.22 \pm 0.02 \text{ mg N}_2\text{O-N m}^{-2} \text{ d}^{-1}$.
563	'i-m-int'	Preliminary estimates suggest that inclusion of fabrication energy for plant and offices and the operative travel energy would increase the embodied impacts of aggregate by some 5-9%, other factors being equal.

564	'i-m-int'	The global warming potential of the N ₂ O is about a few hundred times greater on a weight basis than carbon dioxide, and the lifetime of N ₂ O in the atmosphere is about 150 yr (Bleakley et al., 1982).
565	'i-m-int'	The ratios of The ratios of 44N ₂ O, 45N ₂ O and 46N ₂ O were $81.05 \pm 0.12 : 17.93 \pm 0.12 : 1.02 \pm 0.01$ (when oxygen was 0% (anoxic)); $81.08 \pm 0.23 : 17.94 \pm 0.22 : 0.98 \pm 0.05$ (when oxygen was 5%); $81.12 \pm 0.19 : 17.85 \pm 0.20 : 1.03 \pm 0.06$ (when oxygen was 10%); $80.96 \pm 0.08 : 18.06 \pm 0.12 : 0.98 \pm 0.07$ (when oxygen was 15%).
566	'i-m-int'	The average at.% of ¹⁵ N were 50.31 ± 0.10 , 49.91 ± 0.14 , 49.86 ± 0.07 , when moisture contents were 40%, 50% and 60%, respectively.
567	'i-m-int'	Almost all of the N ₂ O was released by biological denitrification, regardless of the moisture content, at least between 40-60% moisture content.
568	'i-m-int'	N ₂ O/(N ₂ O + N ₂) ratios showed the largest value, 13.3%, when moisture content was 50%, and the ratios decreased almost to 0% as moisture content became 40% or 60%.
569	'i-m-int'	The dissolved CH ₄ concentration measured with the sampler was expressed on a $\mu\text{g g}^{-1}$ (dry soil) basis by multiplying the concentration in $\mu\text{g ml}^{-1}$ by the soil gravimetric moisture content at 10 cm depth which was approximately $1.4 \text{ ml g}^{-1} \pm 0.2$.
570	'i-m-int'	The normal annual precipitation is about 362 mm, nearly 70% of which falls from June through August.
571	'i-m-int'	Grazing density for studied sandy grassland was about five to eight sheep or sheep equivalents before a treatment of enclosure (Zhao, 1992).
572	'i-m-val'	This endangered population of 250 individuals has a 50% chance of extinction over the next 100 years
573	'i-m-int'	However, both IPCC and NGGIC attach large uncertainty to the magnitude of this source: a factor of about 2.
574	'i-m-val'	There was no significant difference between the production rates, both agreeing to within 2% (Leuning et al., 1999).
575	'i-q-rin'	Secondly, the negative effect may be that temperature rise can increase the consumption of water and bring on a water deficit in some biomes (Liang and Maruyama, 1995).
576	'i-m-int'	Amongst the various micrometeorological approaches, the bottom-up mass-balance approach was the most precise with a standard error of just over 30% of the daily mean.
577	'i-q-hnv', 'i-m'	The resulting figure of 1100 km ² (Table 1; Fig. 4) is thought to provide the best estimate of the average early to mid-Holocene lake surface.
578	'ai'	Small-scale fires which mimic the patchiness of the formerly uncleared landscape at a smaller scale within reserves may be a management option for plant populations in isolated fragments.
579	'i'	Human-induced climate change will almost certainly result in increases of MAT, MAP and MAPER.
580	'i-q-rab'	It might indicate that the net carbon budgets of forest ecosystems in the northeastern China were more sensitive to an increase of air temperature than a change of precipitation.
581	'i-q'	Secondly, the negative effect may be that temperature rise can increase the consumption of water and bring on a water deficit in some biomes (Liang and Maruyama, 1995).
582	'i-q-rab', 'i-q-hnv'	This may indicate an advantage of direct measurement of ¹⁵ N-N ₂ O rather than reliance on inhibitors that may also affect denitrification.
583	'i-q-rab', 'i-m-int'	On the contrary, the denitrification of nitrate might be responsible for the second peak, which occurred approximately 40 days after rice transplantation.
584	'i-q-hnv', 'i-m-int'	Within a global perspective, treeline position seems to coincide with a mean temperature in the growing season ranging between 6 and 7°C (Kolmer 2003), suggesting that growth processes might be strongly limited below this threshold.
585	'i-q-hnv', 'i-q', 'i'	Thus, results suggest that convective gas flow is not the primary transport mechanism although mature <i>P. arundinacea</i> may be capable of building slight pressure differentials resulting in slight CH ₄ transport during flowering and increased O ₂ transport to the rhizosphere during the second growth peak and the late growth season.
586	'i-q-rin'	Given the challenges that governments face in dealing effectively with the pace and complexity of scientific innovation at present, much less the longterm impacts, global issues such as climate change suggest that the redesign of governance may be the most critical endeavor for the long-term survival of human civilization.
587	'i-q-rab', 'i-q-hnv'	Several studies have suggested that this seasonal phase locking of the SST anomalies in the Southern Hemisphere may be attributed to the shallowing of the mixed layer during summertime in such way that the latent heat flux and shortwave radiation anomalies influence the SST anomalies more efficiently during this season (Suzuki et al. 2004; Huang and Shukla 2006, 2007; Morioka et al. 2009).
588	'i-q-rin', 'i-q-rab'	Consequently, <i>R. alba</i> might be less dependent on the applied nutrients for its growth and therefore did not show a response to the added N and P. The N:P ratio in the leaves, however, differed among the nutrient treatments, suggesting that <i>R. alba</i> did use the nutrients to some extent.
589	'i-q-rab', 'i-q-hnv'	These different conclusions may be due to differences in models, location and scale, and require further investigation.
590	'i-q-rin', 'i-q-hnv'	Are the life cycle CO ₂ e emissions for tree production greater than CO ₂ stored by a tree planted in an urban environment? If so, large-scale tree planting initiatives may result in net CO ₂ emissions, not storage, even during the life of the tree.
591	'i-q-rab', 'i'	Part of these differences might relate to different rates of adaptation to occupy these relatively new habitats, but the possibility that they might be caused by species-specific differences in perceived predation risk and vulnerability to disturbance means that we should be cautious in applying general predictions.
592	'i-q-hnv', 'i'	It must be stressed that there may be some uncertainties in these determinations.
593	'i-q-hnv'	During the daytime, net radiative heat transfer occurs downward, while at night, or on cloudy days, the net exchange might be upward through the emission of long-wave thermal radiation.
594	'i'	However, due to the uncertainties inherent in model parameters and driving variables (climate, soils...), these should not be taken for solid year predictions.

595	'i'	A contradictory result was obtained with a Gobindabhog cultivated rice field (a strain of rice with aroma), where methane yield under anoxic was greater than that under atmospheric conditions.
596	'i-m'	Flooded rice paddy fields have been considered to be a most prominent source of abiogenic methane emission, though considerable uncertainty exists regarding the true estimates of methane emission.
597	'i'	However, considerable disagreement exists between data obtained from different sources.
598	'i-q-hnv'	Moreover, water inundated rice-paddy fields (though thought to be entirely anoxic) are likely to have low oxygen tension and methane emission under such conditions or atmospheric conditions may suggest possible aerotolerance (or microaerophily) of methanogens, known to be obligate anaerobes.
599	'i-q-hnv'	Microaerophily in methanogens may be due to the possession of some enzymic self-defence against oxygen toxicity (Jones et al., 1983; Kengen et al., 1991).
600	'i-q-hnv'	Moreover, in the scale of evolution, adaptive changes for aerotolerance may be likely for obligate anaerobes like methanogens.
601	'i'	The uncertainties are due to:
602	'i-m'	Thus, the methane emission values are not only uncertain but over estimated.
603	'i'	It is apparent that considerable uncertainty exists as to the contribution of methane from rice paddy soils to the global methane budget.
604	'i-q-hnv', 'i-m-int'	Anthropogenic sources are supposed to contribute 340 Tgy-1 to annual methane budget of which flooded rice fields contribute about 50 Tgy-1 i.e. 15% of the anthropogenic contribution.
605	'i-m-int'	In order to stabilize methane concentration at the present level, a 15-20% cut in methane emission through rice cultivation management has been suggested.
606	'i-q-rin'	I believe that most of the biodiversity that will be lost to humans will be lost in the next fifty years.
607	'i-q-rin'	But I also believe that population stabilization, land abandonment, and biodiversity awareness provide a window of opportunity (Waggoner et al., 1996) for shaping a world in which future losses will not only become less likely in the latter half of the next century, but will begin to be reversed.
608	'i-q-hnv', 'i-m-int'	Annual accumulation in agricultural soils could be equivalent to about 10% of Annex I carbon dioxide emissions, and therefore options for limiting sink credits from soils should be considered.
609	'i'	The Kyoto Protocol to the UN Framework Convention on Climate Change left unanswered the extent to which parties will be permitted to count net changes in carbon stocks from some land use changes against their future greenhouse gas emissions commitments.
610	'i'	The level of uncertainty/transparency/verification is pertinent to this decision, and in practical terms, the costs of achieving a reasonable level of accuracy should be highly relevant.
611	'ai'	The Protocol also leaves open the question of whether this source should be included in its entirety, or as a subset identified by type of activity and/or region or according to the timescale of activities.
612	'i'	Recent regional and national estimates of agricultural soil carbon flux have been based on extrapolations from experimental plots, and the uncertainties associated with these national estimates have been unknown and usually, unspecified.
613	'i'	Existing investigations (2) in some regions are not believed to be sufficiently comprehensive in spatial area, crop type, and typography to provide sufficient data for a reasonably accurate model-based national estimates (IPCC, 1996).
614	'i', 'i-q-hnv'	Lead investigators in the USA (Paustian, 1999) and Canada (McKonkey, 1999) agree that a convincing understanding of the uncertainties involved in estimating carbon flux that considers the many variables of soil type, crop type, management, and climate is not yet available, but they hold that experiments under way may be able to demonstrate an improved modeling capability.
615	'i-q-hnv', 'i-m-int'	Although costs are not up-front, agents spend an estimated half day of labor to sample a typical farm (Vaughan, 1999), and total costs including laboratory and labor charges may involve an estimated \$20-\$30 per sample taken.
616	'i-q-hnv'	Many agricultural practices can enhance carbon accumulation in soils, but conservation tillage and cover crops may be the most important categories that also offer other environmental benefits.
617	'i-q-hnv'	Land area under cover crops in the USA is believed to be increasing but the USDA does not include cover crops in their soil conservation incentive program, and data quantifying their area are not available.
618	'i-q'	The implications of conservation tillage on methane and nitrous oxide emissions requires additional research.
619	'i-q-hnv', 'i-m-int'	These emissions may range from 7.9-11.3 tC/ha/yr for crops on converted peatlands (Armentano and Menges, 1986) and, on an annual basis, greatly exceed the per-area flux related to soil-conservation practices.
620	'i-q-hnv'	If it is characterized by a rapidly increasing trend-line, the increment is likely to be minor and may not justify the expense of monitoring and including these Article 3.4 categories.
621	'i-m'	For example, in regions where the moldboard plow is still in widespread use, such as in the former Soviet Union, the potential for increased carbon sequestration in agricultural soils is much greater, perhaps by an order of magnitude.
622	'i'	It is not realistic that discounting according to uncertainty would be done precisely, because in actuality the uncertainties associated with agricultural soil sequestration will not be known precisely.
623	'i-q-hnv'	However, sufficient evidence may not be available to ensure that these requirements would be met, given that no country has yet demonstrated the ability to complete accurate national assessments for agricultural soils.
624	'i-q-hnv'	Given that parties could offset sizable emissions in other sources, through offsets on agricultural soils, if parties agree to include agricultural soils in Kyoto Protocol inventories, it may be necessary to limit credits for the first commitment period (2008)2012).
625	'i-q-hnv', 'i-m'	To our knowledge, however, none except the one in Gulf of Alaska and the present study lasted more than a year.
626	'i'	This is to elucidate seasonal and interannual flux variabilities and to characterize related biogeochemical cycles of biological elements such as carbon and silicon.

627	'i-m-int'	This was performed using an aliquot of 1/64 original sample and an electronic balance with a precision of ± 10 μg .
628	'i-m-int'	Precision and accuracy of the Analyzer are ± 0.2 and $\pm 0.3\%$ for carbon and ± 0.1 and $\pm 0.2\%$ for nitrogen, respectively.
629	'i-q-hnv', 'i-m-int'	For example, Manganini et al. (1994) and Honjo et al. (1995) suggest that soluble organic carbon is not a major problem for sediment traps deployed at depths greater than 2000 m, showing release of less than $10 \pm 4\%$ of carbon in the formalin-preserved cups.
630	'i-m-int'	Moreover, Honjo et al. (1995) showed that only $5 \pm 2\%$ of biogenic SiO_2 become soluble in formalin preserved samples, small enough so as not to be a major problem.
631	'i-q-hnv', 'i-m-int'	Based on the observed flux values, it is fair to conclude that the Bering Sea is roughly twice as productive as the pelagic region just outside.
632	'i-q-hnv', 'i-m-int'	With this assumption we speculate that total mass fluxes in the Okhotsk Sea in a typical year would be much higher than the $129 \text{ mg m}^{-2} \text{ d}^{-1}$ measured in 1990-1991.
633	'i-q'	The reason why the correlation is relatively weak cannot be readily answered until detailed studies of both siliceous and calcareous plankton groups have been further advanced.
634	'i-q-hnv'	Detailed flux studies on diatoms, radiolarians, calcareous nannoplankton, and planktonic foraminifers in the sediment trap material would no doubt advance understanding of the ecosystem, production, and material cycles.
635	'i-q-rab'	Although timing of the total mass flux maxima varies significantly from one year to another, there are systematic changes in timing and magnitude of the fluxes that seem to be governed by environmental parameters such as hydrography.
636	'i-q-hnv'	For example, simulating fire suppression and a reduction in wildfire activity may result in a larger proportion of the landscape in older seral stages.
637	'i-m'	There is little quantitative information about such second-order effects.
638	'i-q-hnv'	While this assumption may be justified in some ecological systems (e.g. Gustafson, 1996), in the dry interior of British Columbia and in many other ecosystems, broad scale stand-replacing wildfires and bark beetles are examples of agents with significant impacts on landscape characteristics.
639	'i-q-hnv', 'i-m-int'	Such maps can demonstrate whether certain parts of the landscape are more likely to be affected by natural disturbances.
640	'i-m-int'	For example, certain insect outbreaks are more likely to occur in those parts of a landscape that contain susceptible vegetation.
641	'i-q-hnv', 'i-m-int'	Parts of a landscape that are highly fragmented by roads, wetlands or lakes may have lower probability of burning than other parts with the same vegetation types.
642	'i-m-int'	Thirty per cent area of the region has an MAT less than or equal to 0°C , 70% has an MAT less than or equal to 3° , and approximately 30% has an MAT greater than 3°C (Fig. 3(a)).
643	'i-m-int'	Mean annual precipitation (MAP) is 100–300 mm in the western semi-arid, 300–600 mm in the middle semi-humid, and up to 1000 mm in the eastern humid portions (Fig. 3(c)).
644	'i-m-int'	About 60% of the area has MAP between 100 mm and 500 mm, whereas a similar proportion of the region has MAPET equal to or greater than 500 mm (Fig. 4(a)).
645	'i-m-int'	Observations show that, over the last 100 years, atmospheric carbon dioxide (CO_2) concentration has increased by about 70 ppm (Schimel et al., 1996) and global average near-surface air temperature has increased by $0.7\text{--}0.9^\circ\text{C}$ (Nicholls et al., 1996).
646	'i'	However, our current understanding of the functioning of these forest ecosystems does not allow accurate prediction of their carbon sink and sources strength because it is not known how these ecosystems will respond over time to changes in atmospheric CO_2 and global warming.
647	'i-m-int'	The normal temperature range of the air passing through the air handlers is $21\text{--}29^\circ\text{C}$.
648	'i-m-int'	Dry periods rarely last more than a week.
649	'i-m-int'	A comparison of T_a measured by one thermocouple and both T_a/RH probes installed at 15.2m height at the central location for 10 days (27 May 1998–5 June 1998) showed that, although the probes agreed well with each other, the thermocouple measurements were systematically lower by $1\pm 1^\circ\text{C}$ at night and by $4\pm 2^\circ\text{C}$ during daylight hours.
650	'i-m-int'	However, the reader is advised that values of T_a shown later may be prone to absolute errors on the order of a few degree Celsius.
651	'i-m-int'	In the B2C rain forest, monthly mean values of T_a , VP, and VPD are typically $27\text{--}31^\circ\text{C}$, 2.8–3.5 kPa, and 0.6–1.5 kPa, respectively, while monthly mean values of these variables in the Amazon rain forest are typically $22\text{--}26^\circ\text{C}$, 2.0–2.7 kPa, and 0.4–1.4 kPa, respectively (see also Culf et al., 1996).
652	'i-m-int'	There is, on the average, typically a $7\text{--}10^\circ\text{C}$ diurnal cycle in T_a , which also is similar throughout the year and which does not change greatly between sites.
653	'i-m-int'	There is a marked daily cycle in VPD in the Amazon, with near saturation at night and, on the average, VPD reaching peak values of 1–2 kPa during the day.
654	'i-m', 'i-m-int', 'i-q-hnv'	The value of friction velocity, u^* , is similarly unreliable at most measurement levels but may have some relevance at the base of the canopy where there are sustained winds due to the air handling systems.
655	'i-q-hnv'	The higher and sustained turbulence at this level is almost certainly associated with the movement of air between the air handlers that are located below the canopy in the B2C rain forest biome.
656	'i'	It is intriguing that there is less evidence of a systematic daily cycle in vertical wind speed in the middle of the canopy (at 7.5 m).
657	'i-q-hnv'	Arguably, a better approach to enhance similarity with the real world would be to remove the energy near the roof of the B2C where it is captured by cooling the air.
658	'i-q-hnv'	It seems that seeking greater comparability between enclosed and real-world conditions necessarily requires that air be removed from the enclosed environment near the roof because this is where the containment has the most effect.
659	'i-m-int', 'i-q-hnv', 'i-m-val'	About 30% is believed to be produced in grazed pastures from nitrogen provided by fertilisers, fixed by legumes and voided by animals, but this estimate has an uncertainty factor of 2.

660	'i-m'	Results were at the top end of the range of inventory predictions, but uncertainty levels are high.
661	'i-m'	These estimates are based almost wholly on extrapolations of enclosure experiments to the field scale and uncertainty levels are high.
662	'i-m', 'i-q-hnv'	Although uncertainty levels remain high due to large temporal and spatial variability, the micrometeorological measurements suggested that N ₂ O emissions might be considerably larger than those predicted by NGGIC algorithms which use emission factors of 0.4% for urine and 1.25% for dung, but appear to be predicted more closely by IPCC algorithms which use 2% for both.
663	'i-m-int'	Census data for 1995 indicated that there were approximately 0.3 cattle and 1.9 sheep ha ⁻¹ in the local government district surrounding the experimental site (Leuning et al., 1997).
664	'i-m'	The standard errors of the daily means are quite large.
665	'i-m-int'	Typical enrichments in N ₂ O on the downwind boundaries of the test plot were in the range 2±5 ppbv, whereas the FTIR measurement precision was 0.2 ppbv (Leuning et al., 1997).
666	'i-q-hnv', 'i-q-rab'	It seems unlikely that the large standard errors were due to limitations of the FTIR measurements, but perhaps they reflect the scattered nature of the N ₂ O sources and their different times of creation.
667	'i-q-hnv', 'i'	The apparent uptake of N ₂ O on 1 November is difficult to explain.
668	'i-m-int'	The average emission of N ₂ O for all the days of measurement was 54±17 mg N ₂ O-N m ⁻² d ⁻¹ .
669	'i-m-int'	Peak rates of emission were near 100 ng N ₂ O-N m ⁻² s ⁻¹ and the mean rate over 24 h was 2.2 mg N ₂ O-N m ⁻² d ⁻¹ .
670	'i-m-int'	For all 14 days in the study, the mean daytime emission rate (06:00-18:00) calculated in this way was 9.8±6.6 ng N ₂ O-N m ⁻² s ⁻¹ and the mean nighttime rate (18:00-06:00) was 3.8±3.1 ng N ₂ O-N m ⁻² s ⁻¹ .
671	'i-m-int'	These rates correspond to a mean daily emission of 0.59±0.32 mg N ₂ O-N m ⁻² d ⁻¹ .
672	'i-m-int'	The corresponding gradient flux for the same period was 59±81 ng N ₂ O-N m ⁻² s ⁻¹ .
673	'i-m-int'	The DCBL estimate of the mean N ₂ O flux for those periods was 3.0±6.1 ng N ₂ O-N m ⁻² s ⁻¹ and the ICBL estimate was 7.4±6.0 ng N ₂ O-N m ⁻² s ⁻¹ .
674	'i-m', 'i'	However, the standard errors on the micrometeorological flux estimates are too large to speak with any certainty.
675	'i', 'i-m'	The high precision required in atmospheric N ₂ O concentration measurements coupled with quite heterogeneous source distributions, uncertainty about the footprint of the largescale micrometeorological flux measurements because of varying wind directions, and the sporadic nature of the emission itself (being heavily dependent on rainfall events in particular) make it difficult to arrive at definitive flux values from a limited number of observations.
676	'i', 'i-q-hnv', 'i-m-int'	Unfortunately, much uncertainty about emission factors remains and in the present study, micrometeorological approaches would not have been able to detect even modest changes of, say, 20% or 30%, which might be very important for inventory and regulatory purposes as we point out below.
677	'i-m-int'	They report measured fluxes equivalent to 7.6±0.6 and 1.3±0.6 ng N ₂ O-N m ⁻² s ⁻¹ in different investigations.
678	'i-m-int'	By way of comparison, our measurements of daytime and nighttime fluxes were 9.8±6.6 and 3.8±3.1 ng N ₂ O-N m ⁻² s ⁻¹ , respectively.
679	'i-m'	Again, a dedicated high-resolution N ₂ O gas analyser operating continuously at the reference level would improve the level of precision.
680	'i'	Finally, we conclude that while our knowledge of N ₂ O emission factors for grazed pastures is still uncertain, the work described in this paper has been useful in pointing up what is needed for their determination on a field scale.
681	'i-q-hnv'	Prospective investors in the CDM and host countries might place widely different values on the CDM's twin objectives, with investors favoring projects that provide the most cost-effective emissions reductions, and host countries favoring projects that contribute most strongly to national and local development priorities.
682	'i-q-hnv'	We assume in our analysis that sound technical guidelines for measuring and monitoring emissions reductions are defined for the full range of prospective LUCF measures, and that all are approved for crediting.
683	'i-m'	At this time, accurate estimates of future financial flows into the CDM are impossible.
684	'i-q-hnv'	Hence, the cost-effectiveness incentive for CDM investments may be most compelling in the initial years of a carbon reduction agreement.
685	'ai'	Because this requirement is unique to the CDM, it may effectively serve as a tax that places CDM projects at a competitive disadvantage relative to other mitigation measures (Humphreys et al. 1998).
686	'ai'	Alternatively, they could decide to allow Annex I nations participating in the CDM to fulfill the adaptation assistance requirement at the national level, perhaps through a large annual lump sum payment to a multilateral fund, rather than at the investor/project level.
687	'i'	Third, there are considerable uncertainties over how prices will reflect risk.
688	'i', 'i-q-hnv'	Indeed, perceived risk, coupled with weak institutional capacity to identify secure projects, could initially limit the participation of many developing countries in the CDM, including many experiencing rapid deforestation and loss of biodiversity (Frumhoff et al. 1998).
689	'i-q-hnv', 'i-m-int'	Boscolo et al. (1997), for example, estimate that damage to unharvested trees at a project site in Sabah, Malaysia, can be reduced from 40 to 15 percent at a cost of \$150/ha and \$5/ton C.
690	'i-q-hnv'	This constraint is likely to make plantations relatively unattractive options to investors that need immediate carbon credit to offset large volumes of emissions in the short run.
691	'i-q-hnv'	The carbon benefits from forest conservation may be altered by the method of calculating carbon credits chosen by the Parties to the Climate Convention.

692	'i'	Currently, scientists do not yet know the specific relationship between changes in atmospheric concentrations of greenhouse gases, climate change, and consequent harm to the biosphere (Trexler and Associates 1998).
693	'i-q-hnv'	Bilateral or multilateral development organizations, and environmental non-governmental organizations, could certainly contribute to the costs of such projects to help make them more attractive to investors.
694	'ai'	As with all carbon-offset projects, any leakage would need to be monitored and quantified in order to accurately credit the project's emissions reductions.
695	'i-q-hnv'	Reid (1998) identifies a large portfolio of similar cases where forest conservation in critical watersheds could generate local and regional socioeconomic benefits.
696	'i-q-hnv', 'i-m-val', 'i-m'	He estimates that these critical watersheds comprise 13 percent of the world's land area at present, and 25 percent by the middle of the next century.
697	'i-m'	Although forest conservation and restoration are often necessary components of erosion, flood, and water quality management, they are not sufficient in isolation.
698	'i-q-hnv'	Indeed, each setting for a forest-based project would present unique characteristics and constraints that may diminish Reid's sanguine estimate of potential watersheds where socioeconomic benefits could be derived from forest conservation and restoration.
699	'ai'	It is difficult to estimate the magnitude of funds that will be directed to the CDM, and what portion of those funds will ultimately finance the conservation and restoration of natural forests.
700	'i-q-rin'	These side effects give rise to potential environmental hazards on a local, regional and global scale.

G.3 Corpus d'évaluation (300 phrases)

ID	Catégories	Phrase
501	'i-q-hnv'	If spatial information on daily topographic global radiation is not available, this approach may be a promising alternative for predicting spatial and temporal patterns of soil temperature in terrain regions.
502	'i-q-hnv'	Hence, it is hypothesized that global warming will give a more significant positive feedback on the global carbon cycle in a clear-cut area than in an uncut area at a local scale and in tundra than in temperate forests in a global scale since LAI is lower in tundra ($\approx 2 \text{ m}^2 \text{ m}^{-2}$) than in temperate forest ($5\text{--}12 \text{ m}^2 \text{ m}^{-2}$) (Whittaker and Likens, 1975).
503	'i-q-hnv'	Recent experiments of Leadbetter and Bresnak (1996) also suggest oxygen tolerance of methanogens.
504	'i-q-hnv'	It is my belief that 50 years from now, the majority of the world's habitats and species will either be destroyed or on their way to recovery from a degraded state.
505	'i-q-hnv'	Including activities related to land use change and forestry in the international climate change agreement may stimulate new incentives for soil-conservation practices domestically.
506	'i-q-hnv'	If farmers are compensated for activities, rather than for measured changes in carbon stocks, it may not be necessary to establish extensive control sites.
507	'i-q-hnv'	No-till practices may also reduce the rate of methane consumption in soils (Cole et al., 1993).
508	'i-q-hnv'	Experience with other emissions sources suggests that — in the absence of sizable national investments in monitoring — national inventory accuracy does not necessarily increase significantly over a decade's time.
509	'i-q-hnv'	Although this has to be confirmed in future work, for now we assume that the productivity trends in the Okhotsk Sea and our two study regions are parallel.
510	'i-q-hnv'	Removing this difference may well be an intractable problem because it is probably impossible to realistically reproduce the wind fields of a real forest in an enclosed environment.
511	'i-q-hnv'	Verification with direct field measurements is needed.
512	'i-q-hnv'	Moreover, it may prove difficult to accurately track and credit any additional carbon benefits that might be gained off-site through harvested wood or cardboard and paper products (e.g., Blum et al. 1998; Mikales and Skog 1997; Marland et al. 1997).
513	'i-q-hnv'	Investors in this category might include power utilities that find it costly to quickly modify existing power-generating technologies.
514	'i-q-hnv'	The carbon benefits from forest conservation may be altered by the method of calculating carbon credits chosen by the Parties to the Climate Convention.
515	'i-q-hnv'	The assumptions by some authors (e.g. Stavins and Newell 1998) that the carbon discount rate is similar to the financial discount rate and that global warming impacts are linear over time are highly questionable.
516	'i-q-hnv'	We present a method for identifying and selecting projects that may improve the attractiveness of forest-based CDM for local, national, and international stakeholders.
517	'i-q-hnv'	Thus, the sustainability equation, and its energy or pollutant emission equivalent, suggest a multiplier effect between population, economic welfare, and emissions or resource intensity.
518	'i-q-hnv'	The abundance of coal is likely to place an upper limit on all fossil fuel prices at the synthetic fuel cost for the foreseeable future.
519	'i-q-hnv'	A more equitable sharing of world income and resources is likely to be a prerequisite for long-term sustainability.
520	'i-q-hnv'	Carbon dioxide released into the atmosphere from the burning of fossil fuels is thought to persist for around a 100 years.
521	'i-q-rab'	Contradictory findings suggest that no-till may involve lower nitrous oxide emissions than conventional tillage (Li et al., 1996), or higher nitrous oxide emissions (Cole et al., 1993).
522	'i-q-rab'	The ratios of Corg/Cinorg were usually significantly greater than one in both regions, suggesting that preferentially greater organic carbon from cytoplasm than skeletal inorganic carbon was exported from the surface layers.
523	'i-q-rab'	It is likely that the night-time difference is the result of poor relative calibration between two different types of Ta sensors, while the greater discrepancy during the day may be due to solar heating of the metal screen in which the Ta/RH probe is housed.
524	'i-q-rab'	Nonetheless, Fig. 6d shows that (σ_w/u^*) , the normalized standard deviation of w , is between 4 and 6, suggesting very stable conditions even at 3m height.
525	'i-q-rab'	Perhaps this is because the EC system is in dense vegetation at this height, and the air prefers to flow through gaps elsewhere in the canopy.
526	'i-q-rab'	Presumably, this is because the amount of moisture provided has been optimized to maintain a small but finite drainage from the soil column.
527	'i-q-rab'	Thus, an intensified southwest palaeomonsoon apparently transported large quantities of moisture to the eastern Sahara, producing the necessary conditions for the persisting existence of freshwater in the West Nubian Palaeolake Basin.
528	'i-q-rab'	Their extent is probably not adequately reflected in the model pattern, as once again outcropping yardangs imply a greater extent.
529	'i-q-rab'	This halite must be the result of recent efflorescence, as the oxygen isotope and element ratios Sr/Ca and Mg/Ca do not support these values.
530	'i-q-rab'	High element ratios that do not correlate with high $\delta^{18}\text{O}$ content suggest that high levels of salinity were the result of leaching of preexisting salt crusts — e.g., N94/P243 and A859 — or the dilution of a relict brine (Gasse et al., 1987).
531	'i-q-rab'	This stability is corroborated by the parallelism of the $\delta^{18}\text{O}$ curves with the $\delta^{13}\text{C}$ curves, which is only disturbed in the upper part of sections N94/P243 and A859 when major changes in the $\delta^{13}\text{C}$ values occurred towards more negative values, probably indicating a shift towards a larger influence of terrestrial C4 plants and shallowing lake levels.

532	'i-q-rab'	The average sedimentation rate for the sequences shows values between 0.8 and 1.1 mm/year, indicating that the carbonate date 3805 ± 65 years BP is probably much too young as the sediment accumulation rate would increase to > 1.9 mm/year.
533	'i-q-rab'	The lake's stability can only be interpreted as a result of the stable climate without persisting droughts.
534	'i-q-rab'	Regarding the importance of surface water to the lake's water budget, the extremely low content of clastics in the lake deposits must be caused by filtering of runoff through peri-lacustrine vegetation.
535	'i-q-rab'	The low clastic content may also reflect low rates of erosion due to the extensive vegetation cover (Ritchie et al., 1985; Neumann, 1989; Pachur and Hoelzmann, 1991; McKenzie, 1993).
536	'i-q-rab'	Alternatively, substantial variation among sites at a common latitude may also mask a subtle, but potentially important, latitudinal gradient.
537	'i-q-rab'	Regardless of C-input, the profiles showed that dissolved CH ₄ in the floodwater (0 cm depth) was negligible, presumably because CH ₄ diffusing into the floodwater from the soil was rapidly lost by volatilization and oxidation.
538	'i-q-rab'	But the measured concentrations were less than the standards, presumably because of errors either in the standard concentration calculated with Eq. (2) or in the degassing of the sampled solution calculated with Eq. (1).
539	'i-q-rab'	This might be due to exclusion of permafrost soils in the disposition of the Overexploitation Syndrome, relatively low biomass and the missing influence of mining activities (Katanga Syndrome) in the analysis.
540	'i-q-rab'	These differences may have their origin in analytical methodology and sampling site location.
541	'i-q-rin'	Nonetheless, a stable or even (in many countries) declining population within the next century may allow the survival of considerable biodiversity.
542	'i-q-rin'	However, there is every reason to believe that aggressive education campaigns will continue to win the hearts and minds of increasing numbers in future generations.
543	'i-q-rin'	Therefore, if present trends continue and intensify, carbon sequestration in agricultural soils would be most relevant during the first few commitment periods of the Kyoto Protocol.
544	'i-q-rin'	Human-induced climate change will almost certainly result in increases of MAT, MAP and MAPER.
545	'i-q-rin'	Under such conditions, the geographic distribution of current ecosystems will likely shift westward to maintain an adequate water balance.
546	'i-q-rin'	This volume of erosion could have serious impacts on the capacity of the reservoir, cause damage to water transport and filtration infrastructure, and will eventually necessitate the dredging of the reservoir and/or the addition of supplementary reservoirs.
547	'i-q-rin'	The NAPAG working group believes that a reduction of more than 20% in UK energy demand is achievable over the next 20 years with readily available, existing technologies.
548	'i-q-rin'	It is anticipated that the new Renewables Obligation, exemption under the Climate Change Levy, increased renewables R&D funding, and energy market reforms, will enable these new energy technologies to compete more effectively.
549	'i-q-rin'	On engineering grounds the IEE believes that only some 7% of electricity demand could be met from renewables by 2010.
550	'i-q-rin'	Models predict that forests of this region, especially Scots pine forests, will continue sequestering C in response to predicted climate change and increased atmospheric CO ₂ (Kellomäki and Kolström, 1994; Karjalainen and Kellomäki, 1995; Karjalainen, 1996).
551	'i-q-rin'	The positive association between total C content and warmer, moister climates at southern latitudes are consistent with simulation results suggesting that C sequestration rates in northern latitudes are likely to increase in response to predicted climate changes.
552	'i-m-int'	The run terminates when the slow soil organic matter pool reaches steady-state which may require up to 3000 simulation years.
553	'i-m-int'	From the results of in vitro methane emission from Japanese paddy soils in a laboratory, Koyama (1963) first estimated the annual global contribution of methane to be approximately 190 Tgy^{-1} .
554	'i-m-int'	However, there is considerable uncertainty in the estimated values of methane emission, varying widely from 20 Tgy^{-1} to 100 Tgy^{-1} .
555	'i-m-int'	The current value is approximately $0.5\% \text{ y}^{-1}$ (Steele et al., 1992).
556	'i-m-int'	Recent projections of agricultural soil carbon accumulation in the United States center on an estimate of approximately $17\text{--}39 \text{ Mt carbon/yr}$ (Lal et al., 1998), and an estimate at the high end of this range in Canada of $\sim 40 \text{ Mt carbon/year}$ (Dumanski et al., 1998; Lindwall, 1999).
557	'i-m-int'	These costs are high given that the potential value of these carbon offsets may be only \$1.5-\$3.0 billion assuming a value of \$50-\$100/tonne carbon applying to 30 Mt C sequestered.
558	'i-m-int'	Generally, elevations are between 500 and 1200 m.
559	'i-m-int'	Approximately 20% of the west portion has an MAP less than or equal to 300 mm, 70% of the region has an MAP less than or equal to 600 mm, and 30% has an MAP greater than or equal to 600 mm (Fig. 3(c)).
560	'i-m-int'	The S.D. of MAPET ranges from 10 to 80 mm, whereas MAPET ranges from approximately 550 to 700 mm year ⁻¹ .
561	'i-m-int'	Radiation absorption (approximately 50%) by the B2C glass and space frame is high in comparison with the 10–20% absorption reported for glass greenhouses (Mistriotis et al., 1997) because of the need for (6 cm diam) pipe struts to support the (12mm thick) glass roof and walls.
562	'i-m-int'	The mean 24-h emission rate \pm its standard error for the 14 days of measurement was $0.22 \pm 0.02 \text{ mg N}_2\text{O-N m}^{-2} \text{ d}^{-1}$.
563	'i-m-int'	Preliminary estimates suggest that inclusion of fabrication energy for plant and offices and the operative travel energy would increase the embodied impacts of aggregate by some 5-9%, other factors being equal.

564	'i-m-int'	The global warming potential of the N ₂ O is about a few hundred times greater on a weight basis than carbon dioxide, and the lifetime of N ₂ O in the atmosphere is about 150 yr (Bleakley et al., 1982).
565	'i-m-int'	The ratios of The ratios of 44N ₂ O, 45N ₂ O and 46N ₂ O were $81.05 \pm 0.12 : 17.93 \pm 0.12 : 1.02 \pm 0.01$ (when oxygen was 0% (anoxic)); $81.08 \pm 0.23 : 17.94 \pm 0.22 : 0.98 \pm 0.05$ (when oxygen was 5%); $81.12 \pm 0.19 : 17.85 \pm 0.20 : 1.03 \pm 0.06$ (when oxygen was 10%); $80.96 \pm 0.08 : 18.06 \pm 0.12 : 0.98 \pm 0.07$ (when oxygen was 15%).
566	'i-m-int'	The average at.% of ¹⁵ N were 50.31 ± 0.10 , 49.91 ± 0.14 , 49.86 ± 0.07 , when moisture contents were 40%, 50% and 60%, respectively.
567	'i-m-int'	Almost all of the N ₂ O was released by biological denitrification, regardless of the moisture content, at least between 40-60% moisture content.
568	'i-m-int'	N ₂ O/(N ₂ O + N ₂) ratios showed the largest value, 13.3%, when moisture content was 50%, and the ratios decreased almost to 0% as moisture content became 40% or 60%.
569	'i-m-int'	The dissolved CH ₄ concentration measured with the sampler was expressed on a $\mu\text{g g}^{-1}$ (dry soil) basis by multiplying the concentration in $\mu\text{g ml}^{-1}$ by the soil gravimetric moisture content at 10 cm depth which was approximately $1.4 \text{ ml g}^{-1} \pm 0.2$.
570	'i-m-int'	The normal annual precipitation is about 362 mm, nearly 70% of which falls from June through August.
571	'i-m-int'	Grazing density for studied sandy grassland was about five to eight sheep or sheep equivalents before a treatment of enclosure (Zhao, 1992).
572	'i-m-val'	This endangered population of 250 individuals has a 50% chance of extinction over the next 100 years
573	'i-m-int'	However, both IPCC and NGGIC attach large uncertainty to the magnitude of this source: a factor of about 2.
574	'i-m-val'	There was no significant difference between the production rates, both agreeing to within 2% (Leuning et al., 1999).
575	'i-q-rin'	Secondly, the negative effect may be that temperature rise can increase the consumption of water and bring on a water deficit in some biomes (Liang and Maruyama, 1995).
576	'i-m-int'	Amongst the various micrometeorological approaches, the bottom-up mass-balance approach was the most precise with a standard error of just over 30% of the daily mean.
577	'i-q-hnv', 'i-m'	The resulting figure of 1100 km ² (Table 1; Fig. 4) is thought to provide the best estimate of the average early to mid-Holocene lake surface.
578	'ai'	Small-scale fires which mimic the patchiness of the formerly uncleared landscape at a smaller scale within reserves may be a management option for plant populations in isolated fragments.
579	'i'	Human-induced climate change will almost certainly result in increases of MAT, MAP and MAPER.
580	'i-q-rab'	It might indicate that the net carbon budgets of forest ecosystems in the northeastern China were more sensitive to an increase of air temperature than a change of precipitation.
581	'i-q'	Secondly, the negative effect may be that temperature rise can increase the consumption of water and bring on a water deficit in some biomes (Liang and Maruyama, 1995).
582	'i-q-rab', 'i-q-hnv'	This may indicate an advantage of direct measurement of ¹⁵ N-N ₂ O rather than reliance on inhibitors that may also affect denitrification.
583	'i-q-rab', 'i-m-int'	On the contrary, the denitrification of nitrate might be responsible for the second peak, which occurred approximately 40 days after rice transplantation.
584	'i-q-hnv', 'i-m-int'	Within a global perspective, treeline position seems to coincide with a mean temperature in the growing season ranging between 6 and 7°C (Kol'mer 2003), suggesting that growth processes might be strongly limited below this threshold.
585	'i-q-hnv', 'i-q', 'i'	Thus, results suggest that convective gas flow is not the primary transport mechanism although mature <i>P. arundinacea</i> may be capable of building slight pressure differentials resulting in slight CH ₄ transport during flowering and increased O ₂ transport to the rhizosphere during the second growth peak and the late growth season.
586	'i-q-rin'	Given the challenges that governments face in dealing effectively with the pace and complexity of scientific innovation at present, much less the longterm impacts, global issues such as climate change suggest that the redesign of governance may be the most critical endeavor for the long-term survival of human civilization.
587	'i-q-rab', 'i-q-hnv'	Several studies have suggested that this seasonal phase locking of the SST anomalies in the Southern Hemisphere may be attributed to the shallowing of the mixed layer during summertime in such way that the latent heat flux and shortwave radiation anomalies influence the SST anomalies more efficiently during this season (Suzuki et al. 2004; Huang and Shukla 2006, 2007; Morioka et al. 2009).
588	'i-q-rin', 'i-q-rab'	Consequently, <i>R. alba</i> might be less dependent on the applied nutrients for its growth and therefore did not show a response to the added N and P. The N:P ratio in the leaves, however, differed among the nutrient treatments, suggesting that <i>R. alba</i> did use the nutrients to some extent.
589	'i-q-rab', 'i-q-hnv'	These different conclusions may be due to differences in models, location and scale, and require further investigation.
590	'i-q-rin', 'i-q-hnv'	Are the life cycle CO ₂ e emissions for tree production greater than CO ₂ stored by a tree planted in an urban environment? If so, large-scale tree planting initiatives may result in net CO ₂ emissions, not storage, even during the life of the tree.
591	'i-q-rab', 'i'	Part of these differences might relate to different rates of adaptation to occupy these relatively new habitats, but the possibility that they might be caused by species-specific differences in perceived predation risk and vulnerability to disturbance means that we should be cautious in applying general predictions.
592	'i-q-hnv', 'i'	It must be stressed that there may be some uncertainties in these determinations.
593	'i-q-hnv'	During the daytime, net radiative heat transfer occurs downward, while at night, or on cloudy days, the net exchange might be upward through the emission of long-wave thermal radiation.
594	'i'	However, due to the uncertainties inherent in model parameters and driving variables (climate, soils...), these should not be taken for solid year predictions.

595	'i'	A contradictory result was obtained with a Gobindabhog cultivated rice field (a strain of rice with aroma), where methane yield under anoxic was greater than that under atmospheric conditions.
596	'i-m'	Flooded rice paddy fields have been considered to be a most prominent source of abiogenic methane emission, though considerable uncertainty exists regarding the true estimates of methane emission.
597	'i'	However, considerable disagreement exists between data obtained from different sources.
598	'i-q-hnv'	Moreover, water inundated rice-paddy fields (though thought to be entirely anoxic) are likely to have low oxygen tension and methane emission under such conditions or atmospheric conditions may suggest possible aerotolerance (or microaerophily) of methanogens, known to be obligate anaerobes.
599	'i-q-hnv'	Microaerophily in methanogens may be due to the possession of some enzymic self-defence against oxygen toxicity (Jones et al., 1983; Kengen et al., 1991).
600	'i-q-hnv'	Moreover, in the scale of evolution, adaptive changes for aerotolerance may be likely for obligate anaerobes like methanogens.
601	'i'	The uncertainties are due to:
602	'i-m'	Thus, the methane emission values are not only uncertain but over estimated.
603	'i'	It is apparent that considerable uncertainty exists as to the contribution of methane from rice paddy soils to the global methane budget.
604	'i-q-hnv', 'i-m-int'	Anthropogenic sources are supposed to contribute 340 Tgy-1 to annual methane budget of which flooded rice fields contribute about 50 Tgy-1 i.e. 15% of the anthropogenic contribution.
605	'i-m-int'	In order to stabilize methane concentration at the present level, a 15-20% cut in methane emission through rice cultivation management has been suggested.
606	'i-q-rin'	I believe that most of the biodiversity that will be lost to humans will be lost in the next fifty years.
607	'i-q-rin'	But I also believe that population stabilization, land abandonment, and biodiversity awareness provide a window of opportunity (Waggoner et al., 1996) for shaping a world in which future losses will not only become less likely in the latter half of the next century, but will begin to be reversed.
608	'i-q-hnv', 'i-m-int'	Annual accumulation in agricultural soils could be equivalent to about 10% of Annex I carbon dioxide emissions, and therefore options for limiting sink credits from soils should be considered.
609	'i'	The Kyoto Protocol to the UN Framework Convention on Climate Change left unanswered the extent to which parties will be permitted to count net changes in carbon stocks from some land use changes against their future greenhouse gas emissions commitments.
610	'i'	The level of uncertainty/transparency/verification is pertinent to this decision, and in practical terms, the costs of achieving a reasonable level of accuracy should be highly relevant.
611	'ai'	The Protocol also leaves open the question of whether this source should be included in its entirety, or as a subset identified by type of activity and/or region or according to the timescale of activities.
612	'i'	Recent regional and national estimates of agricultural soil carbon flux have been based on extrapolations from experimental plots, and the uncertainties associated with these national estimates have been unknown and usually, unspecified.
613	'i'	Existing investigations (2) in some regions are not believed to be sufficiently comprehensive in spatial area, crop type, and typography to provide sufficient data for a reasonably accurate model-based national estimates (IPCC, 1996).
614	'i', 'i-q-hnv'	Lead investigators in the USA (Paustian, 1999) and Canada (McKonkey, 1999) agree that a convincing understanding of the uncertainties involved in estimating carbon flux that considers the many variables of soil type, crop type, management, and climate is not yet available, but they hold that experiments under way may be able to demonstrate an improved modeling capability.
615	'i-q-hnv', 'i-m-int'	Although costs are not up-front, agents spend an estimated half day of labor to sample a typical farm (Vaughan, 1999), and total costs including laboratory and labor charges may involve an estimated \$20-\$30 per sample taken.
616	'i-q-hnv'	Many agricultural practices can enhance carbon accumulation in soils, but conservation tillage and cover crops may be the most important categories that also offer other environmental benefits.
617	'i-q-hnv'	Land area under cover crops in the USA is believed to be increasing but the USDA does not include cover crops in their soil conservation incentive program, and data quantifying their area are not available.
618	'i-q'	The implications of conservation tillage on methane and nitrous oxide emissions requires additional research.
619	'i-q-hnv', 'i-m-int'	These emissions may range from 7.9-11.3 tC/ha/yr for crops on converted peatlands (Armentano and Menges, 1986) and, on an annual basis, greatly exceed the per-area flux related to soil-conservation practices.
620	'i-q-hnv'	If it is characterized by a rapidly increasing trend-line, the increment is likely to be minor and may not justify the expense of monitoring and including these Article 3.4 categories.
621	'i-m'	For example, in regions where the moldboard plow is still in widespread use, such as in the former Soviet Union, the potential for increased carbon sequestration in agricultural soils is much greater, perhaps by an order of magnitude.
622	'i'	It is not realistic that discounting according to uncertainty would be done precisely, because in actuality the uncertainties associated with agricultural soil sequestration will not be known precisely.
623	'i-q-hnv'	However, sufficient evidence may not be available to ensure that these requirements would be met, given that no country has yet demonstrated the ability to complete accurate national assessments for agricultural soils.
624	'i-q-hnv'	Given that parties could offset sizable emissions in other sources, through offsets on agricultural soils, if parties agree to include agricultural soils in Kyoto Protocol inventories, it may be necessary to limit credits for the first commitment period (2008)2012).
625	'i-q-hnv', 'i-m'	To our knowledge, however, none except the one in Gulf of Alaska and the present study lasted more than a year.
626	'i'	This is to elucidate seasonal and interannual flux variabilities and to characterize related biogeochemical cycles of biological elements such as carbon and silicon.

627	'i-m-int'	This was performed using an aliquot of 1/64 original sample and an electronic balance with a precision of ± 10 μg .
628	'i-m-int'	Precision and accuracy of the Analyzer are ± 0.2 and $\pm 0.3\%$ for carbon and ± 0.1 and $\pm 0.2\%$ for nitrogen, respectively.
629	'i-q-hnv', 'i-m-int'	For example, Manganini et al. (1994) and Honjo et al. (1995) suggest that soluble organic carbon is not a major problem for sediment traps deployed at depths greater than 2000 m, showing release of less than $10 \pm 4\%$ of carbon in the formalin-preserved cups.
630	'i-m-int'	Moreover, Honjo et al. (1995) showed that only $5 \pm 2\%$ of biogenic SiO_2 become soluble in formalin preserved samples, small enough so as not to be a major problem.
631	'i-q-hnv', 'i-m-int'	Based on the observed flux values, it is fair to conclude that the Bering Sea is roughly twice as productive as the pelagic region just outside.
632	'i-q-hnv', 'i-m-int'	With this assumption we speculate that total mass fluxes in the Okhotsk Sea in a typical year would be much higher than the $129 \text{ mg m}^{-2} \text{ d}^{-1}$ measured in 1990-1991.
633	'i-q'	The reason why the correlation is relatively weak cannot be readily answered until detailed studies of both siliceous and calcareous plankton groups have been further advanced.
634	'i-q-hnv'	Detailed flux studies on diatoms, radiolarians, calcareous nannoplankton, and planktonic foraminifers in the sediment trap material would no doubt advance understanding of the ecosystem, production, and material cycles.
635	'i-q-rab'	Although timing of the total mass flux maxima varies significantly from one year to another, there are systematic changes in timing and magnitude of the fluxes that seem to be governed by environmental parameters such as hydrography.
636	'i-q-hnv'	For example, simulating fire suppression and a reduction in wildfire activity may result in a larger proportion of the landscape in older seral stages.
637	'i-m'	There is little quantitative information about such second-order effects.
638	'i-q-hnv'	While this assumption may be justified in some ecological systems (e.g. Gustafson, 1996), in the dry interior of British Columbia and in many other ecosystems, broad scale stand-replacing wildfires and bark beetles are examples of agents with significant impacts on landscape characteristics.
639	'i-q-hnv', 'i-m-int'	Such maps can demonstrate whether certain parts of the landscape are more likely to be affected by natural disturbances.
640	'i-m-int'	For example, certain insect outbreaks are more likely to occur in those parts of a landscape that contain susceptible vegetation.
641	'i-q-hnv', 'i-m-int'	Parts of a landscape that are highly fragmented by roads, wetlands or lakes may have lower probability of burning than other parts with the same vegetation types.
642	'i-m-int'	Thirty per cent area of the region has an MAT less than or equal to 0°C , 70% has an MAT less than or equal to 3° , and approximately 30% has an MAT greater than 3°C (Fig. 3(a)).
643	'i-m-int'	Mean annual precipitation (MAP) is 100–300 mm in the western semi-arid, 300–600 mm in the middle semi-humid, and up to 1000 mm in the eastern humid portions (Fig. 3(c)).
644	'i-m-int'	About 60% of the area has MAP between 100 mm and 500 mm, whereas a similar proportion of the region has MAPET equal to or greater than 500 mm (Fig. 4(a)).
645	'i-m-int'	Observations show that, over the last 100 years, atmospheric carbon dioxide (CO_2) concentration has increased by about 70 ppm (Schimel et al., 1996) and global average near-surface air temperature has increased by $0.7\text{--}0.9^\circ\text{C}$ (Nicholls et al., 1996).
646	'i'	However, our current understanding of the functioning of these forest ecosystems does not allow accurate prediction of their carbon sink and sources strength because it is not known how these ecosystems will respond over time to changes in atmospheric CO_2 and global warming.
647	'i-m-int'	The normal temperature range of the air passing through the air handlers is $21\text{--}29^\circ\text{C}$.
648	'i-m-int'	Dry periods rarely last more than a week.
649	'i-m-int'	A comparison of T_a measured by one thermocouple and both T_a/RH probes installed at 15.2m height at the central location for 10 days (27 May 1998–5 June 1998) showed that, although the probes agreed well with each other, the thermocouple measurements were systematically lower by $1\pm 1^\circ\text{C}$ at night and by $4\pm 2^\circ\text{C}$ during daylight hours.
650	'i-m-int'	However, the reader is advised that values of T_a shown later may be prone to absolute errors on the order of a few degree Celsius.
651	'i-m-int'	In the B2C rain forest, monthly mean values of T_a , VP, and VPD are typically $27\text{--}31^\circ\text{C}$, 2.8–3.5 kPa, and 0.6–1.5 kPa, respectively, while monthly mean values of these variables in the Amazon rain forest are typically $22\text{--}26^\circ\text{C}$, 2.0–2.7 kPa, and 0.4–1.4 kPa, respectively (see also Culf et al., 1996).
652	'i-m-int'	There is, on the average, typically a $7\text{--}10^\circ\text{C}$ diurnal cycle in T_a , which also is similar throughout the year and which does not change greatly between sites.
653	'i-m-int'	There is a marked daily cycle in VPD in the Amazon, with near saturation at night and, on the average, VPD reaching peak values of 1–2 kPa during the day.
654	'i-m', 'i-m-int', 'i-q-hnv'	The value of friction velocity, u^* , is similarly unreliable at most measurement levels but may have some relevance at the base of the canopy where there are sustained winds due to the air handling systems.
655	'i-q-hnv'	The higher and sustained turbulence at this level is almost certainly associated with the movement of air between the air handlers that are located below the canopy in the B2C rain forest biome.
656	'i'	It is intriguing that there is less evidence of a systematic daily cycle in vertical wind speed in the middle of the canopy (at 7.5 m).
657	'i-q-hnv'	Arguably, a better approach to enhance similarity with the real world would be to remove the energy near the roof of the B2C where it is captured by cooling the air.
658	'i-q-hnv'	It seems that seeking greater comparability between enclosed and real-world conditions necessarily requires that air be removed from the enclosed environment near the roof because this is where the containment has the most effect.
659	'i-m-int', 'i-q-hnv', 'i-m-val'	About 30% is believed to be produced in grazed pastures from nitrogen provided by fertilisers, fixed by legumes and voided by animals, but this estimate has an uncertainty factor of 2.

660	'i-m'	Results were at the top end of the range of inventory predictions, but uncertainty levels are high.
661	'i-m'	These estimates are based almost wholly on extrapolations of enclosure experiments to the field scale and uncertainty levels are high.
662	'i-m', 'i-q-hnv'	Although uncertainty levels remain high due to large temporal and spatial variability, the micrometeorological measurements suggested that N ₂ O emissions might be considerably larger than those predicted by NGGIC algorithms which use emission factors of 0.4% for urine and 1.25% for dung, but appear to be predicted more closely by IPCC algorithms which use 2% for both.
663	'i-m-int'	Census data for 1995 indicated that there were approximately 0.3 cattle and 1.9 sheep ha ⁻¹ in the local government district surrounding the experimental site (Leuning et al., 1997).
664	'i-m'	The standard errors of the daily means are quite large.
665	'i-m-int'	Typical enrichments in N ₂ O on the downwind boundaries of the test plot were in the range 2±5 ppbv, whereas the FTIR measurement precision was 0.2 ppbv (Leuning et al., 1997).
666	'i-q-hnv', 'i-q-rab'	It seems unlikely that the large standard errors were due to limitations of the FTIR measurements, but perhaps they reflect the scattered nature of the N ₂ O sources and their different times of creation.
667	'i-q-hnv', 'i'	The apparent uptake of N ₂ O on 1 November is difficult to explain.
668	'i-m-int'	The average emission of N ₂ O for all the days of measurement was 54±17 mg N ₂ O-N m ⁻² d ⁻¹ .
669	'i-m-int'	Peak rates of emission were near 100 ng N ₂ O-N m ⁻² s ⁻¹ and the mean rate over 24 h was 2.2 mg N ₂ O-N m ⁻² d ⁻¹ .
670	'i-m-int'	For all 14 days in the study, the mean daytime emission rate (06:00-18:00) calculated in this way was 9.8±6.6 ng N ₂ O-N m ⁻² s ⁻¹ and the mean nighttime rate (18:00-06:00) was 3.8±3.1 ng N ₂ O-N m ⁻² s ⁻¹ .
671	'i-m-int'	These rates correspond to a mean daily emission of 0.59±0.32 mg N ₂ O-N m ⁻² d ⁻¹ .
672	'i-m-int'	The corresponding gradient flux for the same period was 59±81 ng N ₂ O-N m ⁻² s ⁻¹ .
673	'i-m-int'	The DCBL estimate of the mean N ₂ O flux for those periods was 3.0±6.1 ng N ₂ O-N m ⁻² s ⁻¹ and the ICBL estimate was 7.4±6.0 ng N ₂ O-N m ⁻² s ⁻¹ .
674	'i-m', 'i'	However, the standard errors on the micrometeorological flux estimates are too large to speak with any certainty.
675	'i', 'i-m'	The high precision required in atmospheric N ₂ O concentration measurements coupled with quite heterogeneous source distributions, uncertainty about the footprint of the largescale micrometeorological flux measurements because of varying wind directions, and the sporadic nature of the emission itself (being heavily dependent on rainfall events in particular) make it difficult to arrive at definitive flux values from a limited number of observations.
676	'i', 'i-q-hnv', 'i-m-int'	Unfortunately, much uncertainty about emission factors remains and in the present study, micrometeorological approaches would not have been able to detect even modest changes of, say, 20% or 30%, which might be very important for inventory and regulatory purposes as we point out below.
677	'i-m-int'	They report measured fluxes equivalent to 7.6±0.6 and 1.3±0.6 ng N ₂ O-N m ⁻² s ⁻¹ in different investigations.
678	'i-m-int'	By way of comparison, our measurements of daytime and nighttime fluxes were 9.8±6.6 and 3.8±3.1 ng N ₂ O-N m ⁻² s ⁻¹ , respectively.
679	'i-m'	Again, a dedicated high-resolution N ₂ O gas analyser operating continuously at the reference level would improve the level of precision.
680	'i'	Finally, we conclude that while our knowledge of N ₂ O emission factors for grazed pastures is still uncertain, the work described in this paper has been useful in pointing up what is needed for their determination on a field scale.
681	'i-q-hnv'	Prospective investors in the CDM and host countries might place widely different values on the CDM's twin objectives, with investors favoring projects that provide the most cost-effective emissions reductions, and host countries favoring projects that contribute most strongly to national and local development priorities.
682	'i-q-hnv'	We assume in our analysis that sound technical guidelines for measuring and monitoring emissions reductions are defined for the full range of prospective LUCF measures, and that all are approved for crediting.
683	'i-m'	At this time, accurate estimates of future financial flows into the CDM are impossible.
684	'i-q-hnv'	Hence, the cost-effectiveness incentive for CDM investments may be most compelling in the initial years of a carbon reduction agreement.
685	'ai'	Because this requirement is unique to the CDM, it may effectively serve as a tax that places CDM projects at a competitive disadvantage relative to other mitigation measures (Humphreys et al. 1998).
686	'ai'	Alternatively, they could decide to allow Annex I nations participating in the CDM to fulfill the adaptation assistance requirement at the national level, perhaps through a large annual lump sum payment to a multilateral fund, rather than at the investor/project level.
687	'i'	Third, there are considerable uncertainties over how prices will reflect risk.
688	'i', 'i-q-hnv'	Indeed, perceived risk, coupled with weak institutional capacity to identify secure projects, could initially limit the participation of many developing countries in the CDM, including many experiencing rapid deforestation and loss of biodiversity (Frumhoff et al. 1998).
689	'i-q-hnv', 'i-m-int'	Boscolo et al. (1997), for example, estimate that damage to unharvested trees at a project site in Sabah, Malaysia, can be reduced from 40 to 15 percent at a cost of \$150/ha and \$5/ton C.
690	'i-q-hnv'	This constraint is likely to make plantations relatively unattractive options to investors that need immediate carbon credit to offset large volumes of emissions in the short run.
691	'i-q-hnv'	The carbon benefits from forest conservation may be altered by the method of calculating carbon credits chosen by the Parties to the Climate Convention.

692	'i'	Currently, scientists do not yet know the specific relationship between changes in atmospheric concentrations of greenhouse gases, climate change, and consequent harm to the biosphere (Trexler and Associates 1998).
693	'i-q-hnv'	Bilateral or multilateral development organizations, and environmental non-governmental organizations, could certainly contribute to the costs of such projects to help make them more attractive to investors.
694	'ai'	As with all carbon-offset projects, any leakage would need to be monitored and quantified in order to accurately credit the project's emissions reductions.
695	'i-q-hnv'	Reid (1998) identifies a large portfolio of similar cases where forest conservation in critical watersheds could generate local and regional socioeconomic benefits.
696	'i-q-hnv', 'i-m-val', 'i-m'	He estimates that these critical watersheds comprise 13 percent of the world's land area at present, and 25 percent by the middle of the next century.
697	'i-m'	Although forest conservation and restoration are often necessary components of erosion, flood, and water quality management, they are not sufficient in isolation.
698	'i-q-hnv'	Indeed, each setting for a forest-based project would present unique characteristics and constraints that may diminish Reid's sanguine estimate of potential watersheds where socioeconomic benefits could be derived from forest conservation and restoration.
699	'ai'	It is difficult to estimate the magnitude of funds that will be directed to the CDM, and what portion of those funds will ultimately finance the conservation and restoration of natural forests.
700	'i-q-rin'	These side effects give rise to potential environmental hazards on a local, regional and global scale.