



HAL
open science

Cinquante ans d'un discours institutionnel socio-écologique qui se cherche encore

Bénédicte Pincemin, Ioan Negrutiu

► To cite this version:

Bénédicte Pincemin, Ioan Negrutiu. Cinquante ans d'un discours institutionnel socio-écologique qui se cherche encore: Exercice textométrique mettant en lumière le récit santé-ressources. 2024. halshs-04550985

HAL Id: halshs-04550985

<https://shs.hal.science/halshs-04550985>

Preprint submitted on 18 Apr 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License

Cinquante ans d'un discours institutionnel socio-écologique qui se cherche encore

Exercice textométrique mettant en lumière le récit santé-ressources

Bénédicte PINCEMIN (CNRS, IHRIM UMR5317, ENS de Lyon)

Ioan NEGRUTIU (ENS de Lyon, Institut Michel Serres, Lab. IXXI et RDP UMR5667)

Résumé

Nous avons voulu comprendre l'évolution des discours institutionnels en matière socio-environnementale pour la période 1972-2021, période qui a suivi les Trente Glorieuses et qui a été parcourue par des crises imbriquées. Les 9 textes choisis se distinguent par une qualité et une envergure intellectuelles et politiques, constituant ainsi des références sociétales. Ces textes parlent à la fois de principes, valeurs, règles, droits et devoirs, dans une perception socio-écosystémique des agissements humains. Ces agissements concernent les cycles extraction-production-consommation-déchets, mais aussi des modes de développement, des styles de vie, et des modèles de société ayant produit des conséquences de plus en plus marquées sur la santé des personnes et des milieux. Mais la lecture de ces documents ne permet pas d'identifier clairement les ressemblances et contrastes entre les textes et les particularités discursives de la période analysée. Les textes les plus récents, cependant, pointent plus nettement le besoin d'un récit mobilisateur permettant de transformer les préconisations discursives successives en actes.

Nous avons donc étudié ces discours à l'aide de techniques d'analyse de données textuelles (sur un total d'un demi-million de mots) susceptibles d'objectiver le contenu discursif et de fournir des clés de lecture plus précises de ces documents diplomatiques. L'analyse nous a permis d'examiner cinq thématiques majeures, à savoir un ensemble reliant au fil du temps les questions ressources/rareté, pollution/déchets, développement social et économique, justice, et santé. Nous avons également mis en évidence des affinités textuelles et contextuelles (chapitres, paragraphes, syntagmes) entre, par exemple, santé et pollution. Sur ces bases, nous formulons comme hypothèse de travail la nécessité d'encastrier l'économie dans le social et l'écologie en passant par le récit de la santé commune. Il s'agit d'une santé inclusive et réciproque des milieux, des sociétés et des personnes. Car dans une société visant la justice sociale et la responsabilité écologique tel que différemment évoqué par les textes, la santé serait à adopter comme valeur universelle et objectif social et politique partagés par tous.

Mots-clés :

approche socio-écologique, analyse des données textuelles, santé commune, pollution-déchets, justice, Stockholm Declaration, Rio Summit, UNEP / PNUE reports, déclaration Prix Nobel, logiciel TXM.



Ce document est publié sous licence Creative Commons Attribution 4.0

<https://creativecommons.org/licenses/by/4.0/>

17 avril 2024

Table des matières

1. Introduction	3
2. Méthode et données	5
2.1. L'approche textométrique	5
2.2. Choix des textes	6
2.3. Préparation du corpus pour l'analyse.....	7
3. Parcours d'analyse et résultats	9
3.1. Aperçu général de la configuration du corpus	9
3.1.1. Volumétrie.....	9
3.1.2. Principaux contrastes internes	9
3.1.3. Vocabulaire caractéristique de chaque texte	12
3.2. Mots du corpus et points d'entrée pour l'analyse	14
3.2.1. Vocabulaire d'ensemble du corpus et place des notions étudiées.....	14
3.2.2. Du lexique à la thématique : élaboration de requêtes étendues	16
3.2.3. Lorsqu'un même mot cache plusieurs usages : illustration avec la distinction singulier/pluriel de <i>resource</i>	18
3.3. Contexte situationnel et affinités contextuelles	19
3.3.1. La répartition des mentions : à quelle période ? dans quels textes ? dans quels chapitres ? Spécialisations textuelles en diachronie.....	19
3.3.2. Attirances à l'échelle du paragraphe	23
3.3.3. Autres évolutions lexicales à noter en toile de fond.....	24
3.4. De la syntaxe à la sémantique : observation systématique des constructions dans lesquelles s'ancrent les notions à l'étude	26
3.4.1. Les constructions d'emploi du nom <i>health</i>	27
3.4.2. Les constructions d'emploi du nom <i>pollution</i>	31
4. Discussion et conclusions.	34
4.1. Bilan et perspectives socio-écologiques	34
4.2. Bilan et ouvertures méthodologiques.....	37
Remerciements.....	38
Références	38
Bibliographie scientifique.....	38
Textes du corpus.....	40
Annexes.....	41
Annexe 1. Tableau de synthèse du corpus.....	41
Annexe 2. Vocabulaire caractéristique des neuf textes du corpus (Spécificités)	47

1. Introduction

Ce document de travail vise à illustrer des techniques d'analyse de texte susceptibles d'objectiver le contenu discursif et de fournir des clés de lecture de textes diplomatiques, ici autour des questions socio-environnementales et pour la période 1972-2021.

Le contexte dans lequel s'inscrit notre questionnement de recherche est le suivant. Après les Trente Glorieuses qui ont "épuisé" l'héritage politique post-deuxième guerre mondiale (Collart Dutilleul, 2011 et 2017), la réalité des 50 dernières années est que 75% de la population mondiale vit encore dans la pauvreté, avec 10% dans l'extrême pauvreté. Si cette dernière a été réduite en partie, les inégalités sociales ont explosé (NASEM, mars 2021). De même, ressort le contraste entre le niveau des fonds totaux alloués à la protection de la biodiversité (approximativement 130 millions de dollars par an) et les subventions dommageables à l'environnement (de l'ordre de 2,5 milliards de dollars par an) (Narain et al., 2022). Ainsi, si la richesse globale a augmenté, cela s'est fait au détriment de la nature et de ses ressources vitales.

Et pourtant, le rapport pour le Club de Rome (1972 ; Meadows et al., 2005) actait une remise en question politique, économique, sociale, environnementale et scientifique du modèle économique dominant, en partant de la question de la finitude des ressources physiques. Il enclenche, ou il est au moins contemporain, d'un cheminement scientifique et diplomatique qui a vu l'arrivée sur les agendas politiques de nouveaux concepts :

- inséparabilité du social, de l'environnemental et du développement durable (voir Binder et al., 2013, pour une utile synthèse),
- empreinte écologique et limites planétaires (Steffen et al., 2015 ; Whitmee et al., 2015) pour quantifier les pressions humaines sur la biosphère,
- limites sociétales (Brand et al., 2021) pour préciser les cadres dans lesquels s'inscrivent des inégalités sociales de plus en plus fortes,
- réponses institutionnelles à partir du sommet de Rio (1992) qui se sont focalisées sur les Objectifs du Millénaire pour le Développement (OMD), suivi des Objectifs de Développement Durable (ODD), mais dont les effets politiques se sont révélés faibles (Biermann et al., 2022).

Nous retenons ici le fait que depuis la Déclaration de Stockholm en 1972, le développement des sociétés humaines est perçu comme étroitement dépendant de la nature. Ainsi, les systèmes économiques et de gouvernance exercent des effets directs et indirects sur les écosystèmes (leurs fonctions, les services et les ressources qu'ils procurent, leur productivité). En retour, les externalités qui en résultent rétroagissent (effets-retour) sur les structures sociales et de gouvernance (Ostrom, 2009 ; Bourgeron et al., 2018). Les systèmes alimentaires et la pollution globale sont des exemples parlants de ces impacts et de leurs conséquences sur la santé humaine et les inégalités sociales (Negrutiu, 2022). Nous formulons l'hypothèse suivante : le système liant ressources, pollution globale, et santé, constitue la « fabrique » d'une déconstruction sociale et écologique porteuse des crises systémiques actuelles. Pour aller plus loin, nous avons calqué sur la période 1972-2021 les crises énergétiques, politiques et financières de la période et neuf textes diplomatiques choisis comme corpus d'analyse afin de disséquer les éléments de langage et l'évolution des discours institutionnels déployés sur ces cinquante dernières années (**Figure 1 ; Tableau 1**).

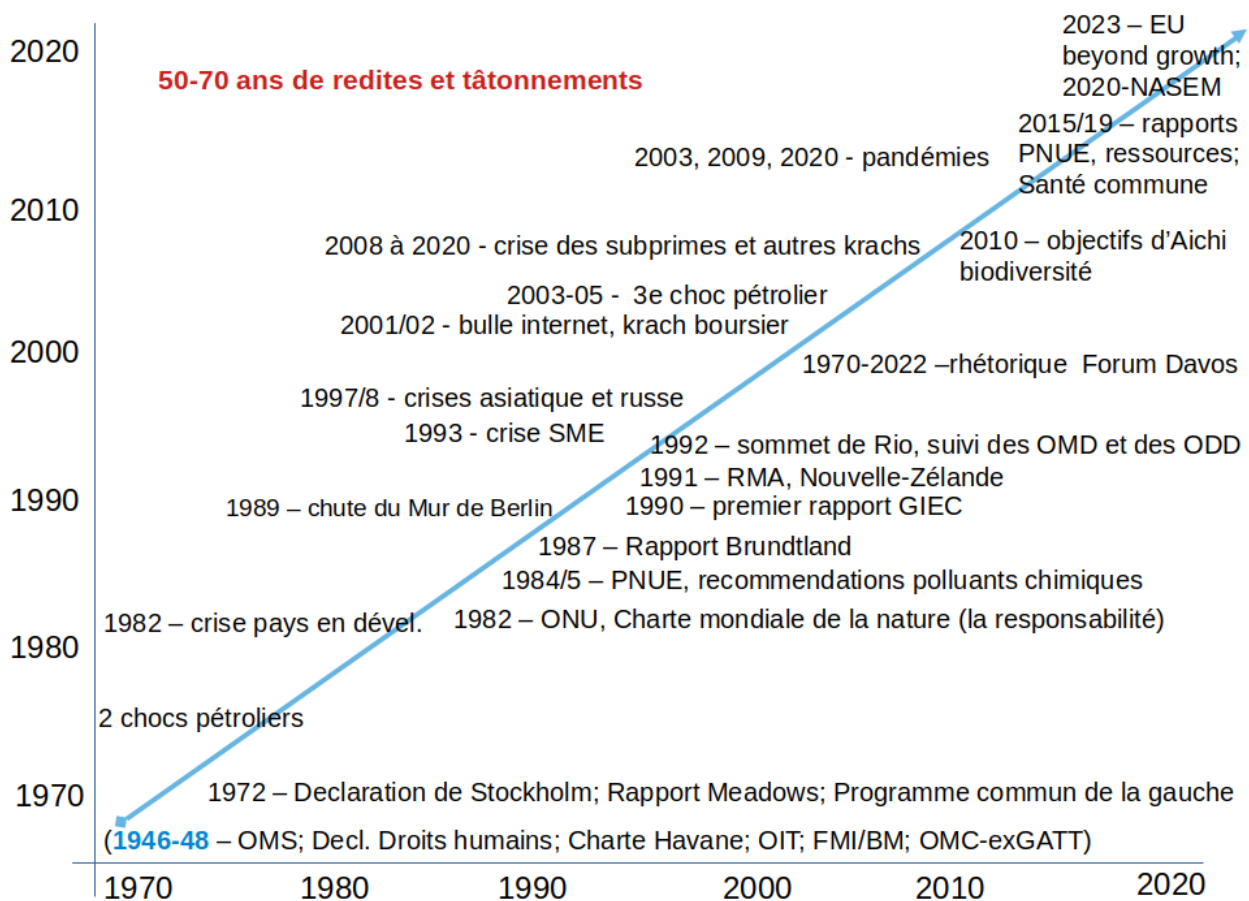


Figure 1. Mise en perspective historique du corpus analysé. À partir des Trente Glorieuses (1945-1972), la réalité des 50 dernières années est esquissée par le temps des crises (axe des ordonnées) qui ont été essentiellement financières et énergétiques, et le temps des agendas et des rhétoriques institutionnelles et diplomatiques (axe des abscisses) qui déclinent sous diverses formes et avec une intensité variable la responsabilité partagée et l'urgente nécessité de mesures appropriées, aux niveaux national et international, individuel et collectif, privé et public, pour protéger la nature et ses ressources. Les choix des exemples n'est pas exhaustif. **Notes et abréviations.** OMS, Organisation Mondiale de la Santé ; OIT, Organisation Internationale du Travail ; FMI, Fond Monétaire International ; BM, Banque Mondiale ; OMC ex-GATT, Organisation Mondiale du Commerce ; GIEC, Groupement International des Experts pour le Climat ; SME, Système Monétaire Européen, et prise de conscience plus générale des limites des mouvements (trop libres) de capitaux ; RMA, Resource Management Act ; OMD, Objectifs du Millénaire pour le Développement ; ODD, Objectifs de Développement Durable ; PNUE, Programme des Nations Unies pour l'Environnement ; NASEM, National Academies of Sciences, Engineering, and Medicine des Etats Unis – appel et rapport des prix Nobel portant sur la science de la soutenabilité.

Les textes d'Olof Palme (1972), la Déclaration de Stockholm (1972) et le rapport Brundtland (1987) vont du récit centré sur des principes et valeurs (Palme) à leur déclinaison par des règles centrées sur ressources/rareté, pollution, développement social et économique (dont la santé), le tout dans une pensée socio-écologique (responsabilité, équité, solidarité, les communs, planification).

Avec Rio (1992), à la démarche ressources-pollution s'est ajoutée celle, plus floue, du développement durable (introduite dès le rapport Brundtland en 1987) et ses variations liées aux Objectifs du Millénaire et leur suite, les Objectifs de développement durable. Rio-1992 est à l'image de la décennie 1980-1990, qui est celle des libérations commerciales et financières imposées par les pays riches à l'ensemble du monde.

Les textes UNEP (40 ans plus tard) se positionnent sur des priorités stratégiques, dont changement climatique, pollution chimique et déchets, et ressources et efficacité des ressources, en décalage avec l'esprit du moment, "grow now, clean up later". Ils insistent sur "la demande de consommation

incessante basée sur des modèles d'industrialisation et de développement non durables"¹, ainsi que sur "la distribution inégale des bénéfices de l'utilisation des ressources, et les impacts de plus en plus globaux et graves de celle-ci sur le bien-être humain et la santé des écosystèmes".²

Enfin, arrivent ces dernières années les questionnements et positionnements des Prix Nobel réunis en conclave par les Académies américaines (NASEM, mars 2021). En voici quelques exemples :

*Nous n'avons jamais eu à faire face à des problèmes de l'ampleur de ceux auxquels est aujourd'hui confrontée la société mondialement interconnectée. Personne ne sait avec certitude ce qui va fonctionner, si bien qu'il est important de construire un système qui puisse évoluer et s'adapter rapidement.*³

*Qui a le pouvoir de décider ce qu'est une "bonne transformation" ?*⁴

*Un changement systémique peut-il naître d'initiatives locales ?*⁵

*Les Objectifs de Développement Durable s'avèrent davantage des perspectives que des supports opérationnels.*⁶

*La nécessité d'une vision convaincante pour donner une direction, la nécessité de former des réseaux et d'affaiblir les forces structurelles dominantes.*⁷

*Sans action de transformation au cours de la présente décennie (2021-2030), l'humanité prend des risques colossaux pour notre avenir commun.*⁸

La lecture de cet ensemble nous a donc incités à tenter de scruter l'évolution sémantique des textes à l'aide d'une analyse textométrique, pour mieux saisir le sens que les structures institutionnelles mêmes ont donné à une pensée sociale et écologique déjà présente durant ces cinq décennies passées, avec une matrice d'analyse reliant ressources, pollution, santé, justice et finance.

Dans la section qui suit (2), nous présentons plus concrètement les données textuelles et l'outillage sur lesquels est basée l'étude. La section 3, centrale, expose les observations tirées du corpus, en commençant par des vues d'ensemble, puis en se focalisant progressivement sur certains thèmes, en particulier autour de la santé et de la pollution. Notre dernière section (4) replace dans la réflexion socio-écologique les enseignements issus du parcours analytique, et relit l'expérience d'un point de vue méthodologique. Des matériaux supplémentaires sont par ailleurs réunis en annexe.

2. Méthode et données

2.1. L'approche textométrique

L'analyse textométrique (Lebart et Salem, 1994 ; Née, 2017 ; Poudat et Landragin, 2017 ; Lebart et al., 2019) propose une approche à la fois quantitative et qualitative des corpus numériques. Le corpus dans son ensemble est exploré à travers les mots qu'il contient et leurs contextes d'emploi (mots voisins, apparition dans certains textes ou dans certaines parties). Les aspects *quantitatifs*

¹ "the story of relentless demand and of unsustainable patterns of industrialization and development" (UNEP, 2019, p.5)

² "the story of the unequal distribution of the benefits of resource use and its increasingly global and severe impacts on human well-being and ecosystem health." (UNEP 2019, p. 5)

³ We have never had to deal with problems of the scale facing today's globally interconnected society. No one knows for sure what will work, so it is important to build a system that can evolve and adapt rapidly.

⁴ Who has agency to decide what is a "good transformation"?

⁵ Can systemic change arise from local initiatives?

⁶ SDGs serve more as aspiration than practical level guidance.

⁷ The need for a compelling vision to provide direction, forming networks, and weakening dominant structural forces.

⁸ Without transformational action this decade (2021-2030), humanity is taking colossal risks with our common future.

exploitent des décomptes fréquentiels et des mesures statistiques correspondant à des questionnements de lectures (prépondérance, affinités, particularités, etc.). Le volet *qualitatif* confère une place centrale au “retour au texte”, c’est-à-dire à un travail interprétatif toujours ancré sur l’observation des données telles qu’elles se manifestent dans le document source⁹. Toute une gamme de traitements quantitatifs et qualitatifs seront illustrés plus avant en section 3.

La méthode est robuste et endogène, au sens où sa valeur ajoutée ne provient pas de l’application au corpus de connaissances extérieures (annotation sémantique, référentiels ontologiques...) mais d’une observation méthodique et systématique du comportement contextuel des mots et expressions linguistiques à l’intérieur du corpus construit pour rendre compte d’un intertexte de lecture (Pincemin, 2022). On comprend donc que la construction du corpus, qui pose le cadre d’analyse et définit ce qu’on pourrait appeler l’univers de référence, participe fondamentalement à l’analyse : tous les résultats de traitements sont relatifs à ce cadre posé au départ.

Plus généralement, la démarche textométrique n’est pas une automatisation presse-bouton mais implique activement le chercheur dans toutes ses étapes (Pincemin, 2020). Elle consiste à construire un parcours d’analyse, à commencer donc par le choix des textes qui définit le contexte d’ensemble déterminant pour l’analyse, puis en interrogeant et explorant le corpus selon une problématique de recherche, qui permet d’affiner le choix des points d’entrée, de noter les points d’intérêt à approfondir, et d’interpréter de façon critique et éclairée les sorties logicielles. L’outil logiciel ne remplace pas la connaissance et la lecture des textes, mais fournit un instrument pour lire autrement et pousser plus avant l’exploration de son expression verbale.

La réalisation de la présente étude ne pouvait donc pas simplement consister à appliquer les outils textométriques à un texte fourni : il fallait mobiliser une problématique scientifique liée aux documents concernés permettant de définir le corpus (introduction et section 2), de guider son exploration (section 3), et de donner sens aux résultats de traitement (section 4). C’est pourquoi ce travail, à travers ses deux auteurs, allie une expertise de la méthode textométrique et une expertise de recherche dans le domaine du développement durable : cette double expertise est essentielle pour la valeur scientifique de la démonstration.

2.2. Choix des textes

Notre intérêt se centre sur la prise de conscience et la place dans les discours des relations réunissant ressources, pollution, santé, et justice, qui caractérisent le mieux les modes de développement, les styles de vie, et in fine des modèles de sociétés devenus problématiques.

Dans un premier temps, nous avons retenu neuf documents internationaux nous permettant d’observer ces concepts dans un discours général sur l’environnement, sur une échelle de temps d’une cinquantaine d’années, de 1972 à 2022. La synthèse de 50 années de programmes institutionnels se centre sur l’évolution dans la réalisation des objectifs environnementaux et socio-économiques - ou plus exactement dans l’élaboration du discours à leur propos. Cela couvre la démarche initiée par la Déclaration de Stockholm de 1972, le rapport Brundtland qui a suivi en 1987, la Conférence de Rio de 1992, la réflexion de l’UNEP sur la période 2015-2019, et les lauréats du Nobel envisageant une science de la soutenabilité (*sustainability science*) en 2020-2021 (**Tableau 1**).

⁹ En ce sens, la textométrie se démarque d’un pur *distant reading* (en imposant une combinaison de *distant reading* et de *close reading*) ou de beaucoup d’approches de *text mining* tendues vers la production de visualisations synthétiques davantage conçues pour remplacer la lecture des données sources que pour revenir les examiner de plus près.

Tableau 1. Les neuf documents composant le corpus.

Code dans le corpus	Date	Nom courant	Émetteur	Nb. pages
1972palme	1972	Olof Palme Statement	Swedish Delegation to the UN Conference on the Human Environment	12
1972stockholm	1972	Stockholm declaration	United Nations	10
1987brundtland	1987	Brundtland report	United Nations	300
1992rio	1992	Rio Conference, Earth Summit	United Nations	492
2015unep	2015	Policy Coherence SDGs	United Nations - UNEP International Resource Panel	56
2017unep	2017	Assessing Global Resource Use	United Nations - UNEP International Resource Panel	104
2019unep	2019	Global Resources Outlook	United Nations - UNEP International Resource Panel	162
202103nasem	2021	Sustainability Science	National Academies of Sciences, Engineering, and Medicine	14
202104nasem	2021	Our Planet, Our Future	National Academies of Sciences, Engineering, and Medicine	7

2.3. Préparation du corpus pour l'analyse

Pour mener l'analyse textométrique, nous avons choisi le logiciel *open-source* TXM¹⁰ (Heiden et al., 2010) développé en principal dans notre laboratoire, car il offre une large gamme de traitements textométriques (dont des calculs statistiques issus de R¹¹), un puissant langage d'interrogation (intégration du moteur CQP¹²), et une capacité à prendre en compte une représentation riche des textes notamment selon les recommandations internationales de la TEI¹³ (avec la possibilité de ne pas traiter en bloc, mais de façon structurée, nos documents dont certains font plusieurs centaines de pages). Nous avons donc besoin de préparer une édition numérique de nos neuf textes, et nous avons choisi de produire une version de type XML TEI, car c'est le format qui nous ouvre ensuite le plus de possibilités pour ajuster la représentation des textes en fonction des besoins pour les analyses.¹⁴

Nous avons téléchargé des versions numériques publiques des documents originaux en version anglophone depuis le site de leur émetteur officiel, pour travailler à partir de sources de première main.

Nous avons eu affaire à trois types de formats :

¹⁰ La version utilisée est la version 0.8.3.

¹¹ R est une bibliothèque logicielle de référence open-source pour le calcul statistique : <https://www.r-project.org/>

¹² Le Corpus Query Processor est le composant central de l'Open Corpus Workbench (CWB) initié à l'IMS de l'université de Stuttgart : <https://cwb.sourceforge.io/>

¹³ La Text Encoding Initiative (TEI) est un consortium de la communauté scientifique internationale qui développe et améliore continuellement un standard pour la représentation des textes numériques : <https://tei-c.org/>

¹⁴ Mais un simple format TXT, ou un format DOCX, seraient aussi directement importables dans TXM, si l'on avait opté pour un corpus simplifié, sans les repérages fins à l'intérieur des textes.

- des PDF-image (numérisation de pages de documents tapuscrits) pour les documents les plus anciens (à l'exception de 1987brundtland), pour lesquels l'obtention du contenu textuel passe par une reconnaissance optique de caractères¹⁵ ;
- des PDF-texte (PDF créé par export depuis le texte édité dans un traitement de texte) pour la majorité des documents diffusés à partir des années 2000, pour lesquels l'obtention du contenu textuel suppose aussi des traitements pour restituer les mots et l'ordre du texte (par delà des mises en page souvent élaborées : colonnes, encadrés, etc.)¹⁶ ; et
- un document en ligne (nasem202104), dont le contenu textuel est directement accessible par enregistrement ou export depuis le navigateur.

À l'aide des utilitaires de préparation de corpus de TXM et dans l'environnement d'édition du logiciel Oxygen XML Editor, nous avons converti les contenus textuels extraits des neuf documents en fichiers XML utilisant le vocabulaire TEI. Pour se repérer dans les documents, nous avons fait en sorte de conserver la pagination des PDF originaux et un repérage mnémotique des chapitres (numérotation + mot-clé). Nous avons éliminé les éléments répétitifs de mise en page (comme les titres courants) de façon à ce que l'analyse reste centrée sur le contenu ; dans le même ordre d'idée, nous avons repéré certaines parties à l'intérieur des textes pour pouvoir les écarter si besoin des traitements : ce qui ne concerne pas notre sujet (informations commerciales, éditoriales, etc.), les références bibliographiques, les marqueurs d'un plan-type répétitif (par exemple des intertitres récurrents dans 1992rio). Nous avons également procédé à des corrections typographiques et orthographiques semi-automatiques, pour éliminer des erreurs importantes ou massives issues du processus d'extraction textuelle.

Les fichiers XML préparés pour les neuf textes, accompagnés d'un tableau de métadonnées décrivant les textes (année, titre, émetteur, référence bibliographique, url source, genre textuel, nombre de pages, etc.) ont été importés dans TXM pour produire le corpus POLLUTION-HEALTH-221110 comptant un demi-million de mots¹⁷. Lors de l'import, une analyse morphosyntaxique a été réalisée à la volée avec le logiciel TreeTagger (Schmid, 1994), ajoutant à chaque mot une information de lemme (à savoir sa forme non fléchie normalisée, type entrée de dictionnaire) et de catégorie grammaticale ; ces informations linguistiques sont ensuite mobilisables pour les interrogations du corpus.¹⁸

En principe, le corpus ainsi préparé est disponible¹⁹ pour les lecteurs qui voudraient reproduire nos résultats ou prolonger l'analyse.

¹⁵ Nous avons eu recours au service AbbyCloud d'Huma-Num (CNRS) ou au logiciel Acrobat Reader d'Adobe selon les documents (choix au cas par cas de l'extraction optimale).

¹⁶ Nous avons utilisé la bibliothèque open-source PDFBox, intégrée dans l'utilitaire prototype PDF2TXT de TXM.

¹⁷ Les mots des passages ou parties codées comme en dehors de notre sujet (copyright,...), ceux des références bibliographiques, et ceux des éléments structuraux répétitifs de structuration ou mise en page, ne sont pas comptabilisés et n'interviendront pas dans les analyses (ils sont seulement restitués pour la consultation des documents, pour interpréter les mots dans leur contexte).

¹⁸ L'analyse morphosyntaxique a été ici réalisée pour des textes en anglais, mais le logiciel TreeTagger peut fonctionner avec des modèles d'autres langues, dont le français, voir par exemple les fichiers paramètres disponibles sur la page du logiciel : <https://cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

¹⁹ Dans la limite des droits qui pourront nous être accordés, nous envisageons un dépôt du corpus dans un entrepôt pour la diffusion des données de recherche, tel que celui offert par la plateforme Ortolang, a priori sous l'appellation Resources-Pollution-Health.

3. Parcours d'analyse et résultats

3.1. Aperçu général de la configuration du corpus

3.1.1. Volumétrie

Un premier constat à noter d'entrée de jeu sont les **tailles très inégales** des différents documents (**Figure 2**) : les trois documents les plus courts comptent autour de 2500 mots, et quatre "poids lourds" de 50 000 à 180 000 mots se partagent 93 % du volume du corpus (les deux gros rapports 1987brundtland et 1992rio représentent chacun un tiers du volume total, et les deux rapports UNEP de 2017 et 2019 forment un petit troisième tiers). Les approches quantitatives du corpus seront donc par construction essentiellement déterminées par ces quatre documents. En revanche, les approches qualitatives pourront toujours prendre en considération les documents moins volumineux.

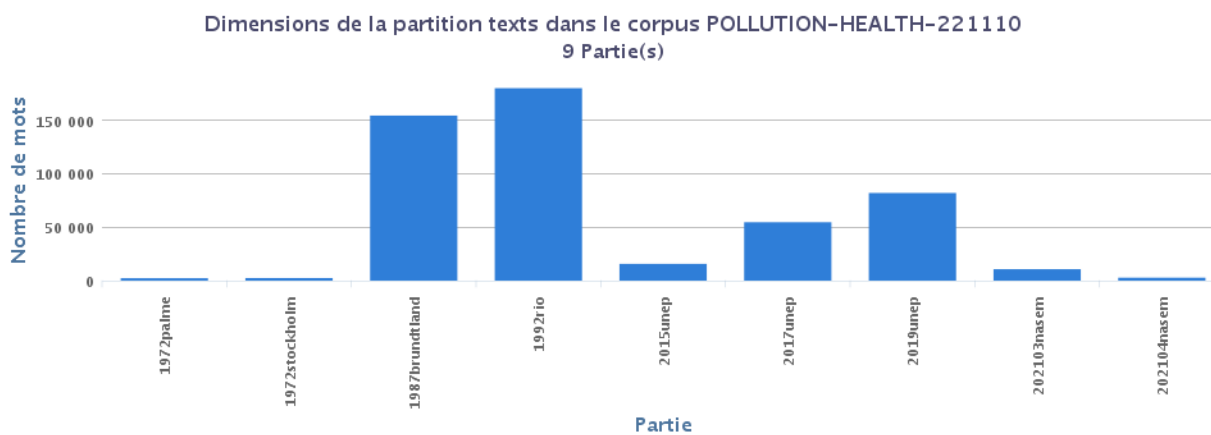


Figure 2. Diagramme des tailles des neuf documents du corpus POLLUTION-HEALTH-221110.

3.1.2. Principaux contrastes internes

Une analyse factorielle des correspondances nous apporte une visualisation globale de l'organisation des textes entre eux en fonction de leur contenu lexical (**Figure 3**), plus précisément ici²⁰ en fonction de leur profil d'emploi des 200 noms communs les plus fréquents du corpus.

²⁰ Il s'agit bien d'une analyse factorielle possible de ce corpus, et non de son (*unique*) analyse factorielle. En effet, d'une part le corpus pourrait être divisé autrement qu'en textes (en chapitres, en années, en décennies, etc.) et d'autre part la caractérisation pourrait se faire autrement qu'avec 200 noms (on pourrait utiliser tous les mots du corpus, ou tous ceux de fréquence supérieure à 100, ou des regroupements comme les cinq thématiques définies plus loin, etc. etc.). Le choix des textes et des 200 lemmes nominaux les plus fréquents correspond à un point d'entrée usuel, car il produit une visualisation pas trop chargée et souvent assez parlante en lien avec les thématiques du corpus. La lemmatisation permet de rassembler dans le même mot les formes au singulier et au pluriel de chaque nom, comme dans un dictionnaire. C'est en poursuivant l'analyse que d'autres façons de diviser le corpus ou de représenter les textes peuvent suggérer le calcul d'analyses factorielles des correspondances complémentaires.

Plan factoriel de l'analyse des correspondances de POLLUTION-HEALTH-221110/texts/< [enpos="NNS?" & word!="\p{P}"]>@enlemma ≤506 347 /197/@enlemma /197

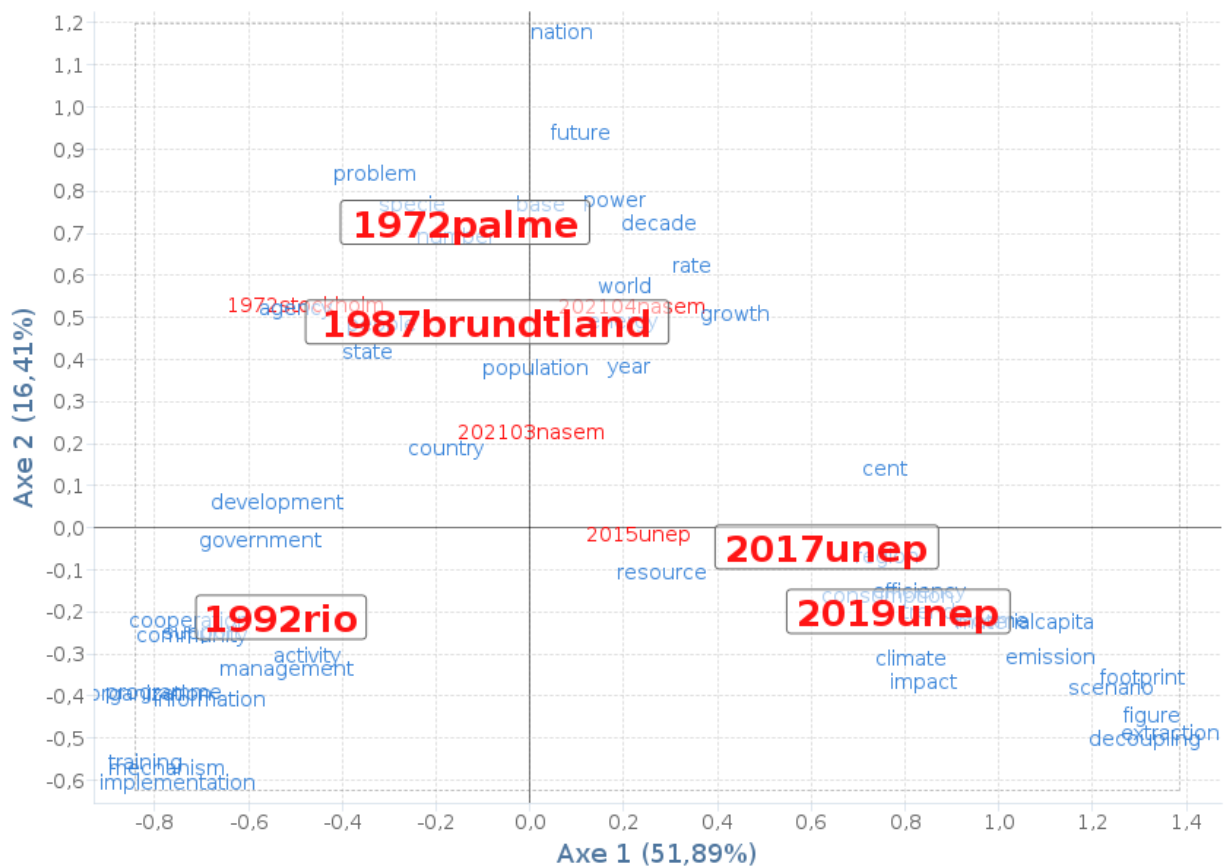


Figure 3. Cartographie des ressemblances et contrastes entre les textes selon leur volume d'emploi des noms les plus fréquents (analyse factorielle des correspondances des 9 textes caractérisés par les 197 noms communs²¹ les plus fréquents du corpus, lemmatisés). Pour les textes (en rouge), on a mis en gros caractères ceux qui sont les mieux représentés (les moins déformés) dans la projection plane (cosinus carré d'au moins 0,57, ainsi que 1972palme avec la valeur suivante, de 0,17). Pour les noms (en bleu), pour la lisibilité du graphique, n'ont été affichés que ceux qui apportent une contribution d'au moins 1% sur l'un des deux axes, c'est-à-dire qui pèsent le plus dans le contraste principal du corpus.

[Encart explicatif Analyse factorielle des Correspondances] Chaque texte est simplement représenté par sa fréquence d'emploi de chacun des 200 noms. Cela est la base d'une modélisation géométrique, car une distance peut être calculée entre ces représentations (vectorielles) des textes : des textes ayant tendance à employer beaucoup les mêmes mots, et peu d'autres mots semblables, vont être proches dans cet espace. Les textes sont donc positionnés en fonction de leurs choix lexicaux et dominantes de vocabulaire. L'espace géométrique complet compte ici 8 dimensions²². Le calcul de l'analyse factorielle consiste alors à trouver le repère qui oriente et hiérarchise ces 8 dimensions de telle sorte que (i) les dimensions sont orthogonales (elles n'ont aucune information en commun, elles sont totalement complémentaires), et (ii) quel que soit l'entier n , les n premières dimensions

²¹ Nous avons retiré trois mots incorrectement catégorisés comme noms par l'analyseur morphosyntaxique automatique. Le nombre de noms traités est donc finalement de 197.

²² Le nombre de dimensions de l'espace total est défini par le minimum du nombre de colonnes (ici les 9 textes) et du nombre de lignes (ici les 197 noms) du tableau de données, minimum ensuite diminué d'une unité, soit $9-1 = 8$.

représentent le maximum d'information²³ sur les contrastes présents dans les données. L'origine (le centre) du référentiel correspond à la moyenne du corpus, physiquement le centre de gravité des textes (pondérés par leur taille) et des mots (pondérés par leur fréquence). La position centrale signale donc un texte ou un mot "dans la moyenne", qui ne marque pas d'originalité dans ses usages (du point de vue quantitatif). Une position excentrée signale une originalité au moins en partie partagée avec les autres textes ou mots excentrés dans la même direction. Dans la visualisation plane que nous observons en **Figure 3**, l'espace est réduit à deux dimensions, les deux premières afin de maximiser l'information représentée (les indications de pourcentage d'inertie sont données au niveau de chaque axe, ici donc avec 2 dimensions on peut capter $51,89 + 16,41 = 68,3$ % de l'information). Néanmoins, les autres dimensions sont "écrasées", projetées sur le plan, et donc certains points apparemment proches (c'est-à-dire proches pour les deux premières dimensions) peuvent être écartés par une ou plusieurs des dimensions suivantes. Il faut donc interpréter les proximités visuelles avec prudence et contrôler les apparentes associations avec d'autres indicateurs (comme la qualité de représentation dans le plan (cosinus carré au plan), les coordonnées sur les dimensions suivantes, etc.) ou avec un calcul de spécificités (qui permet d'établir clairement des relations statistiques entre mots et textes). Mais en première instance, la représentation factorielle plane est d'abord utilisée pour révéler la structure globale et ce qui apparaît comme le plus fort contraste au sein du corpus (première dimension, conventionnellement représentée horizontalement). Les axes du graphique ne correspondent donc pas à des grandeurs simples, mais chacun est une construction et synthèse mathématique abstraite à interpréter à partir des mots et des textes qui contribuent le plus à le construire (indicateur de contribution à l'inertie de l'axe, en %) ou/et qui lui correspondent le mieux (indicateur du cosinus à l'axe).

Dans cet espace lexical (**Figure 3**), il apparaît que **les textes les plus anciens (en particulier 1972palme et 1987brundtland) se positionnent de façon préceuse** : non seulement ce ne sont pas eux qui marquent le plus fort contraste lexical avec les documents les plus récents (le pôle des documents UNEP structurant le premier axe en s'opposant à 1992rio), mais ils marquent même une affinité avec les toutes dernières productions émises par les Académies. Les deux textes de 1972, ceux de la NASEM et le rapport Brundtland forment ainsi un groupe dont se distinguent d'une part les trois textes UNEP, et d'autre part 1992rio (qui indique un éloignement institutionnel des "idéaux" des années 1970), formant ainsi trois grands pôles dans notre corpus (que confirmera une classification, cf. **Figure 4**). L'analyse des correspondances apporte un éclairage lexical à ce qui motive cette structuration. Ainsi, l'imposant volume de **Rio (1992) est marqué par sa structure formelle très régulière** abordant systématiquement les modes de mise en œuvre des recommandations thème par thème (usage récurrent de *implementation, training, management*, etc.). Et de leur côté, **les rapports UNEP ressortent par une introduction massive des graphiques**, drainant le vocabulaire des données visualisées (*figure, per capita, income*, etc.). Cette analyse factorielle sur les noms les plus fréquents (**Figure 3**) attire ainsi notre attention sur le fait que le contenu lexical de notre corpus est loin d'être purement thématique, il est aussi façonné par les **formes d'expressions différentes** adoptées selon les années et le cadre d'interlocution (qui s'adresse à qui, avec quelle visée).

La lecture de l'analyse factorielle peut être complétée par une classification (**Figure 4**), dans la mesure où la visualisation 2D de l'analyse factorielle est une projection qui déforme les distances entre les éléments représentés (même si c'est le moins possible, par construction). La classification montre les proximités en rendant compte de façon plus complète des différences qui peuvent séparer certains textes.

²³ Il s'agit d'une mesure d'inertie, qui combine la masse (fréquence pour un mot, taille pour un texte) et l'originalité (écart à la moyenne, correspondant à la distance au centre du graphique). Autrement dit, l'information croît à mesure de la masse de données textuelles concernées ou/et de leur originalité.

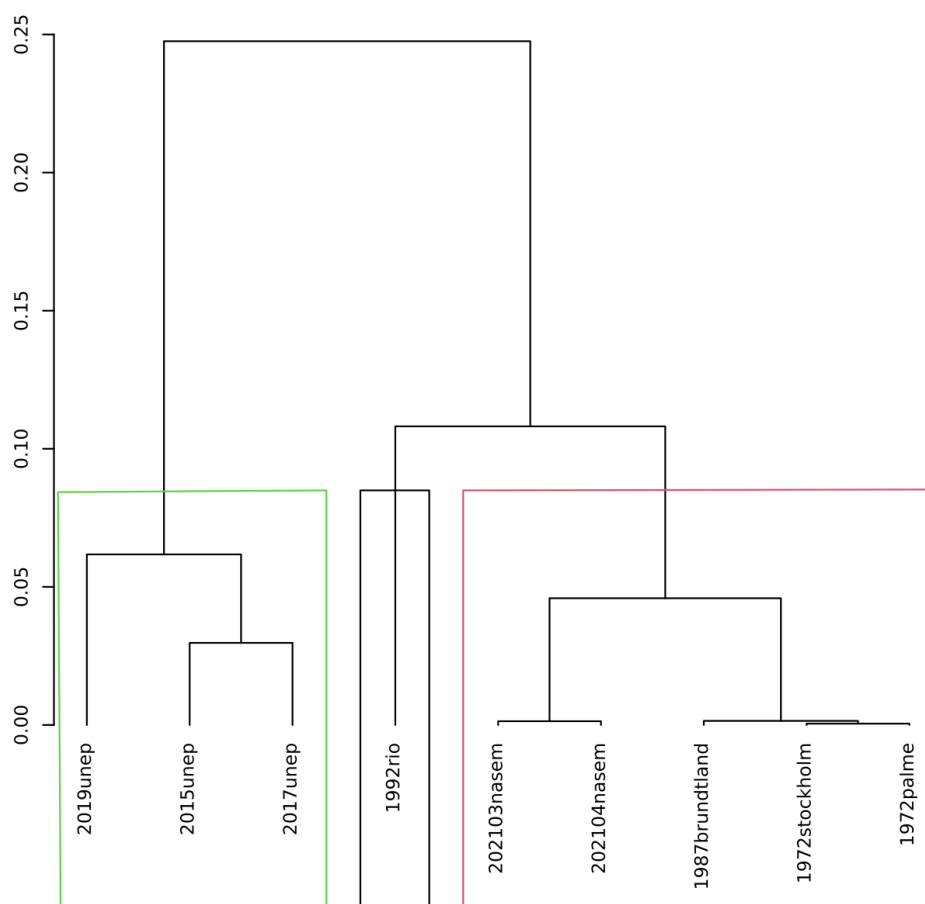


Figure 4. Arbre des ressemblances et contrastes entre les textes selon leur volume d'emploi des noms les plus fréquents (classification ascendante hiérarchique des 9 textes caractérisés par les 197 noms communs les plus fréquents du corpus, lemmatisés).

[Encart explicatif Classification ascendante hiérarchique] Les neuf textes ont des coordonnées dans l'espace factoriel. Cet espace est ici considéré dans ses 5 premières dimensions sur 8, soit plus de 92% de l'information, pour rendre la classification plus claire et efficace en neutralisant les variations mineures, singulières et isolées, pouvant s'apparenter à des perturbations. Les coordonnées permettent de calculer des distances entre textes (distance euclidienne usuelle). La classification opère des regroupements en fonction à la fois de la distance et de la taille (masse) des textes, en faisant successivement les regroupements les plus homogènes (méthode dite de Ward). Le dendrogramme (**Figure 4**) se lit ainsi de bas en haut : les textes qui forment les regroupements les plus homogènes sont les deux textes de 1972 et les deux textes de la NASEM (cela est en partie favorisé par leur petite taille et par le fait d'avoir écarté les dernières dimensions de l'espace factoriel, qui rendaient compte des variations les plus marginales), et 1987brundtland s'apparente aux deux textes de 1972. Les regroupements suivants portent sur les textes UNEP (le dernier étant un peu moins semblable aux deux premiers) et les textes NASEM qui rejoignent le paquet 1972 et 1987brundtland. À ce stade il ressort de notre corpus 3 grands groupes : un groupe UNEP qui s'oppose aux deux autres, le groupe des 3 textes les plus anciens et des deux NASEM récents, et 1992rio qui forme le troisième groupe à lui seul.

3.1.3. Vocabulaire caractéristique de chaque texte

Le calcul des spécificités permet de détecter les mots anormalement fréquents dans un texte au regard de leur fréquence moyenne dans le corpus.

[Encart explicatif Spécificités] Plus précis qu'une comparaison de pourcentages ou qu'une règle de trois, le calcul des Spécificités intègre une évaluation statistique²⁴ qui permet de ne considérer que les variations difficilement attribuables au hasard, celles qui sortent des fluctuations contingentes. Un indice de 4 pour le mot M dans le texte T par exemple signifie que, si les mots avaient été répartis au hasard, le mot M avait une probabilité de moins d'une chance sur dix mille (le nombre de zéros correspond à l'indice, ici 4 dans notre exemple) d'apparaître avec une fréquence aussi élevée (égale à celle qu'on observe ou supérieure) ; autrement dit, plus l'indice est grand, moins la force de la fréquence est attribuable au hasard, et donc plus elle semble avoir fait l'objet d'un choix marqué dans ce texte. La mesure peut s'effectuer aussi bien pour repérer des emplois anormalement fréquents (indice positif) que des emplois anormalement peu fréquents (indice négatif). L'indice nul correspond à la fréquence moyenne : le mot est employé dans le texte avec une fréquence correspondant à la moyenne du corpus entier. On convient usuellement de considérer comme non significatives les spécificités comprises entre -3 et 3 (voire -2 et 2, lorsque la taille du texte – ou la fréquence totale du mot – est petite, ce qui réduit l'échelle sur laquelle les statistiques peuvent s'exprimer).

Nous considérons ici (**Tableau 2**) les mots tels que sélectionnés par le calcul pour la Déclaration de Stockholm, lorsqu'on la compare à l'ensemble de notre corpus. Le détail des résultats du calcul pour chacun des autres textes est donné en **Annexe 2**.

Tableau 2. Mots (lemmes) les plus spécifiques de 1972stockholm (12 max. par catégorie et indice de spécificité supérieur ou égal à 2).

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
principle environment man life jurisdiction earth posterity state mankind co-operation domination harm	human common environmental man-made essential social adverse irreversible future present adequate <i>irreplaceable</i> ²⁵	co-ordinate safeguard improve bear protect transform reach <i>propel</i> <i>entrust</i> <i>remedy</i> <i>imperilled</i> <i>condemn</i>	must shall	his their him	all wisely continuously heedlessly gravely wrongly

1972stockholm se présente comme un texte au ton plus solennel et formel (*man, shall, principle*), posant sans détour (*must*) les notions fondamentales de la préoccupation socio-écologique (*environment* et *environmental, earth, co-operation* et *domination, posterity, future* et *irreversible, safeguard, protect, improve* et *transform*).

²⁴ Probabilité selon la loi hypergéométrique, basée sur le décompte de toutes les distributions possibles des mots dans les différentes parties du corpus (ici les textes).

²⁵ En italiques, les mots dont la seule et unique occurrence est dans le texte considéré. Il y a généralement un lot de mots ainsi ex-aequo : pour compléter la liste à 12, nous avons choisi ceux qui nous paraissent les plus parlants.

3.2. Mots du corpus et points d'entrée pour l'analyse

3.2.1. Vocabulaire d'ensemble du corpus et place des notions étudiées

Le corpus peut d'abord être observé dans le détail de son vocabulaire, à travers des listes complètes des mots utilisés, avec indication de leur fréquence (leur nombre d'attestations). Ces listes peuvent être dressées selon certains critères de sélection (par exemple, ne considérer que les noms communs) et adopter différents points de vue sur les mots (par exemple, sur les verbes, garder le détail des flexions - en temps, personne, nombre... - ou bien grouper toutes ces variations derrière une seule entrée lexicale conventionnellement à l'infinitif, dite "lemme"). Ces listes peuvent être parcourues alphabétiquement, tel un dictionnaire du corpus, rapprochant les mots par leur partie initiale, ou bien selon un point de vue quantitatif, généralement par fréquence décroissante.

Le **Tableau 3** recense ainsi les 100 noms communs les plus fréquents du corpus.

Tableau 3. Les 100 noms communs les plus fréquents du corpus.²⁶ *Le décompte est effectué sur les lemmes, c'est-à-dire en neutralisant les variations flexionnelles singulier/pluriel et les variations typographiques, toutes les variantes sont comptabilisées ensemble.*

Nom	Fréq.	Nom	Fréq.	Nom	Fréq.	Nom	Fréq.
resource	2919	growth	653	figure	443	effect	324
country	2276	information	621	source	440	datum	321
development	2268	consumption	608	process	439	way	321
use	1471	population	606	group	424	ecosystem	317
programme	1340	world	606	region	419	sustainability	317
policy	1197	need	577	climate	418	effort	315
impact	1184	pollution	577	action	414	role	315
material	1121	community	576	income	410	capita	314
system	1081	people	568	implementation	408	goal	314
government	1054	strategy	544	problem	402	institution	306
level	994	forest	536	industry	397	time	306
water	973	food	529	assessment	387	objective	301
management	956	research	520	state	387	issue	297
cent	941	year	518	service	379	supply	285
area	938	cost	496	extraction	371	planning	284
environment	908	efficiency	489	trade	368	education	281
activity	846	cooperation	487	conservation	367	woman	281
production	820	capacity	468	approach	365	rate	278
waste	814	emission	465	infrastructure	364	account	275
land	803	support	462	product	361	fuel	275
technology	800	sector	460	trend	361	agency	269
energy	798	term	459	risk	347	access	265
organization	732	city	451	nation	342	chemical	264
health	673	measure	451	training	336	productivity	261
change	672	economy	444	example	331	framework	260

Le nom **resource** (dans le détail : 1378 singuliers et 1541 pluriels) arrive en tête. Nous pouvons vérifier à l'aide des concordances triées sur les contextes gauche ou droit, que cette fréquence particulièrement élevée n'est pas attribuable à un *artefact* -par exemple, un titre courant répété à chaque page que nous aurions manqué de neutraliser au moment de la préparation du corpus. On constate que ce nom joue en partie un rôle sémantique d'hyperonyme et un rôle morphosyntaxique de support, au sens où il se décline de multiples façons à l'aide d'un complément nominal ou adjectival qui le spécialise sémantiquement, par exemple pour citer les plus fréquents : *natural resource(s)* (510 occ.), *water resources* (119 occ.), *human resource(s)* (92 occ.), *land resource*

²⁶ Requête CQL : [enpos = "NNS?" & enlemma!="%"]

(65 occ.), *genetic resources* (51 occ.), *financial resources* (48 occ.), etc. Cette souplisse d'emplois tantôt généraux, tantôt plus précis, en fait un "passe-partout" omniprésent.

On peut noter que *health* (673 occurrences) et *pollution* (577 occ.) figurent également quantitativement en bonne place dans le corpus, en compagnie de *waste* (814), *land* (803), *technology* (800), *energy* (798), *population* (606), *forest* (536) et *food* (529).

En prenant du recul et en passant du mot isolé au champ lexical, le bilan du tableau peut faire l'objet d'agrégations par affinité sur la première moitié des 100 mots. Ainsi, en additionnant les fréquences :

(1) *resources+energy+use+efficiency* = (5677 occurrences)

et *government+management+development+policy* = (5475 occurrences)

indiquent un poids déterminant et assez équilibré des aspects liant les ressources physiques et leur politique de gestion.

(2) *pollution+waste+emissions* = (1856 occurrences)

health+environment = (1581 occurrences)

permettent d'observer un certain équilibre causes - effets des pollutions sur la santé et l'environnement. La problématique *waste* dépasse celle de la *pollution* sur la durée de l'analyse²⁷.

(3) La question démographique (*population*), mais aussi celle concernant les capacités de mesure (*measure+assessment=451+387* occurrences) recouvrent des positions significatives dans l'ensemble du corpus, mais la question se pose de leur articulation aux défis de référence que constituent ressources, pollution, santé, droits.

Tableau 4. Les 50 adjectifs qualificatifs²⁸ et les 50 verbes²⁹ les plus fréquents du corpus (lemmes).

Adjectif	Fréq.	Adjectif	Fréq.	Verbe	Fréq.	Verbe	Fréq.
environmental	1606	agricultural	318	be	9581	must	397
international	1386	sound	297	have	2407	see	386
sustainable	1386	public	283	should	1849	would	379
national	1208	technical	282	develop	1722	grow	373
global	1114	total	282	include	1387	do	348
economic	1069	specific	275	can	982	encourage	299
other	1015	particular	272	will	936	base	290
such	960	important	271	increase	825	give	287
appropriate	888	available	266	provide	772	exist	277
natural	730	effective	265	promote	724	relate	261
human	712	different	255	use	626	depend	246
regional	668	marine	243	improve	623	lead	236
local	607	necessary	236	need	553	build	231
many	556	great	232	reduce	548	follow	227
new	555	good	229	take	533	identify	226
relevant	472	non-governmental	229	make	520	consider	221
high	471	rural	225	require	520	assess	216
social	465	integrated	223	strengthen	471	enhance	208
more	446	hazardous	212	support	449	meet	208
urban	381	small	212	achieve	437	help	207
industrial	360	essential	211	ensure	422	manage	203
large	351	domestic	210	implement	416	address	200
scientific	345	future	207	could	414	create	196
financial	327	institutional	205	establish	399	involve	195
major	325	most	198	may	399	contribute	190

²⁷ Le nombre d'occurrences de *waste* fait un bond de près de 400 occurrences sur les seuls deux chapitres 20 et 21 de l'Agenda 21 de 1992rio, précisément consacrés aux déchets ; les mentions des déchets sont aussi sensiblement plus présentes que celles de la pollution dans le rapport 2019unep.

²⁸ Requête CQL : [enpos = "JJ.?" & enlemma!="@card@"]

²⁹ Requête CQL : [enpos = "[VM].*"]

De la même façon que pour les noms communs, on peut procéder au relevé des adjectifs qualificatifs ou des verbes quantitativement dominants (**Tableau 4**).

Considérons à présent les différentes réalisations en corpus des radicaux *-health-* et *-pollut-* (**Tableau 5**). La famille de *health* totalise 786 occurrences, dont 726 (92%) pour le seul nom simple *health* (toujours au singulier). On observe davantage de formes dérivées autour de la pollution (32 formes au lieu des 19 pour *health*), mais domine toujours une forme nominale de base et au singulier (*pollution*), qui représente la majeure partie des occurrences (575 sur 712, soit plus de 80%).

Tableau 5. Toutes les formes de mots³⁰ du corpus contenant les radicaux *-health-*³¹ ou *-pollut-*³²

<i>Famille de Health</i>	<i>Fréq</i>	<i>Famille de Pollution</i>	<i>Fréq</i>	<i>Famille de Pollution</i>	<i>Fréq</i>
health	664	pollution	530	anti-pollution	1
Health	61	Pollution	45	High-pollution	1
healthy	33	pollutants	43	less-polluting	1
unhealthy	4	polluting	24	pollute	1
health-care	3	pollutant	9	Polluter	1
health-related	3	polluter	7	polluter-pays	1
Healthy	3	air-pollution	6	pollutes	1
healthier	2	polluted	6	Polluting	1
ill-health	2	pollution-related	5	pOLLutiOn	1
pollution-health	2	low-polluting	4	POLLUTION	1
HEALTH	1	pollution-free	4	pollution-abatement	1
health-	1	Pollutants	3	pollution-control	1
health_topics	1	Pollution-intensive	3	pollution-intensive	1
health-based	1	non-polluting	2	Pollution-Intensive	1
health-car	1	Pollutant	2	pollutive	1
Healthier	1	polluters	2		
health-protecting	1	pollution-health	2		
non-health	1				
resource-infrastructure-health	1				

3.2.2. Du lexique à la thématique : élaboration de requêtes étendues

D'un point de vue sémantique, les thèmes de santé ou de pollution pourraient cependant être exprimés à travers d'autres mots, basés sur d'autres radicaux. Nous pouvons utiliser un calcul de cooccurrences pour nous suggérer³³ des candidats (**Tableau 6**).

³⁰ Il s'agit de mots au sens d'une délimitation automatique, qui n'est pas exempte d'erreurs, sur des textes numériques eux-mêmes non sans coquilles. Pour avoir une vision très claire (directe) des mots concernés, le recensement est établi au plus près des données, sur les formes graphiques des mots telles qu'elles se présentent dans les fichiers source (sans lemmatisation). Ainsi apparaissent ici par exemple toutes les variantes d'écriture avec ou sans majuscules.

³¹ Requête CQL : `".*health.*"%c`

³² Requête CQL : `".*pollut.*"%c`

³³ Le calcul fournit des résultats bruts, tels que montrés dans le **Tableau 6** (où nous avons juste omis les chiffres, symboles et ponctuations). Ils ne sont bien sûr pas directement utilisables tels quels et demandent à être interprétés pour faire la part entre les relations de voisinage qui peuvent nous intéresser pour une généralisation thématique et celles qui proviennent de cas particuliers dans nos textes ou d'autres types de liens. Par exemple, dans les cooccurents des mots de la famille de *pollution*, *Delhi* vient de l'étude de cas développée dans 2017unep ; *Ramaswami* est un auteur abondamment cité dans cette même section. D'une certaine façon, nos cooccurrences rendent compte de cette étude de cas.

Tableau 6. Les 50 premiers cooccurents (lemmes) des mots des familles de *health* et *pollution*, dans un contexte d'une distance de 50 mots avant et après.

Cooccurents de <i>-health-</i>					Cooccurents de <i>-pollut-</i>				
Mot	F tot.	Co-F	Ind.	Dist	Mot	F tot.	Co-F	Ind.	Dist
health	673	300	111	20,5	air ³⁴	302	245	189	4,1
impact	1243	334	57	15,2	pollution	577	283	128	22,6
human	735	221	46	9,9	control	348	153	61	9,4
care	95	67	42	1,9	city	451	173	59	16,9
PM	48	44	37	3,9	emission	465	150	40	14,7
disease	124	72	37	19,6	industry	397	109	23	19,7
particulate	44	39	32	8,8	PM2	37	28	20	18,7
exposure	55	37	22	13,5	infrastructure	364	97	19	20,1
matter	84	44	20	11,1	industrial	360	94	18	21,7
pollution	577	140	20	18,3	premature	34	25	18	11,7
air	302	91	19	16,0	damage	148	54	17	15,4
education	281	85	18	7,1	reduce	548	120	16	14,1
child	110	48	18	16,5	health	673	138	16	18,6
risk	356	96	17	13,4	multiple	107	43	16	14,0
communicable	27	22	16	23,3	waste	832	158	15	19,5
city	451	107	14	22,6	co-benefits	46	26	14	14,8
reproductive	17	16	14	4,4	quality	219	60	13	23,2
safe	138	49	14	19,7	resource-efficient	47	25	13	14,9
infrastructure	364	87	12	24,4	pollutant	57	27	12	17,6
environment	908	169	12	12,5	Delhi	54	26	12	15,7
mortality	37	22	12	15,9	death	51	25	12	13,7
diet	36	21	11	12,2	fuel	297	70	11	19,2
Health	53	25	10	16,9	exposure	55	25	11	15,2
sanitation	101	36	10	16,7	coal	85	30	9	23,0
chemical	300	72	10	23,9	transboundary	90	31	9	7,4
safety	107	37	10	12,6	Ramaswami	20	14	9	24,0
co-benefits	46	22	9	11,3	burning	30	17	9	26,2
traffic	46	22	9	20,3	energy	798	134	9	22,1
hazard	55	24	9	6,8	marine	243	57	9	16,3
worker	55	24	9	16,0	water	979	156	9	16,4
toxic	112	36	9	23,0	gas	215	52	9	18,7
nutrition	33	18	9	7,0	discharge	36	18	9	22,9
toxicity	33	18	9	10,8	associate	92	30	8	12,9
well-being	195	50	8	10,4	urban-industrial	17	12	8	15,7
woman	281	64	8	15,9	acidification	39	18	8	17,2
effect	326	71	8	13,9	bundle	15	11	8	22,7
Human	50	21	7	12,0	indoor	15	11	8	4,6
row	10	9	7	25,7	mitigation	54	21	8	11,0
Russia	10	9	7	28,7	reduction	199	46	7	12,7
climate	418	83	7	15,1	brick	16	11	7	8,7
indoor	15	11	7	22,7	coastal	163	40	7	24,0
pollutant	57	22	7	23,9	atmospheric	43	18	7	7,4
environmental	1606	240	7	17,5	waste-related	11	9	7	5,9
hazardous	212	50	7	19,9	end-of-pipe	17	11	7	11,5
drug	16	11	7	18,1	ship	17	11	7	13,6
control	348	71	7	21,2	source	443	79	7	16,9
maternal	11	9	7	15,2	flue	7	7	7	20,4
Ressource	11	9	7	27,9	hazard	55	20	7	18,6
respiratory	14	10	6	15,9	land-based	25	13	6	14,8
urban-industrial	17	11	6	18,1	feature	33	15	6	12,9

³⁴ Pour ce calcul, les lemmes *air* et *Air*, correspondant en fait tous deux au nom commun *air* et apparaissant tous deux dans les 25 premiers cooccurents de la famille de *pollution*, ont été regroupés.

Le calcul de cooccurrence fournit un indice statistique lié à la probabilité que le mot cooccurrent apparaisse avec une fréquence aussi élevée (c'est-à-dire supérieure ou égale à celle constatée) au voisinage du mot choisi comme pivot initial. Les voisinages sont ici définis comme un empan de 50 mots de part et d'autre du mot pivot : d'autres réglages seraient possibles mais une fenêtre d'une centaine de mots nous semble bien en mesure de capter des relations thématiques.

À partir de ces listes, d'intuitions complémentaires, et des observations des contextes pour vérifier l'interprétation, nous pouvons construire quelques thématiques plus couvrantes :

- On complète le radical *-pollut-* avec les dérivés de *chemical*, *emission* et *waste*³⁵, ce qui correspond à 86 formes de mot différentes et 2470 occurrences en tout dans le corpus ;
- Le radical *-health-* peut quant à lui être étendu avec les dérivés de *disease* et *mortality*³⁶, pour couvrir 25 formes de mot différentes et 949 occurrences en tout.

De la même façon, nous définissons des requêtes pour étoffer les autres thématiques qui nous intéressent :

- La problématique financière pourra être appréhendée aussi à travers les dérivés de *insurance*, *invest-*, et *subsid-*³⁷, soit 42 formes de mot différentes pour 841 occurrences.
- La place de la *justice* s'appuiera également sur les mentions de *right*, *law* et *responsibility* (singulier et pluriel)³⁸, totalisant 17 formes de mot différentes pour 547 occurrences.
- Pour *resource*, il ne nous a pas semblé utile d'étendre au-delà de la famille du mot³⁹, qui rassemble déjà 59 formes de mot différentes pour 3472 occurrences.

3.2.3. Lorsqu'un même mot cache plusieurs usages : illustration avec la distinction singulier/pluriel de *resource*

Les cooccurrents sont par ailleurs globalement très éclairants sur les usages du mot pivot en corpus, car ils forment une synthèse statistique et lexicale de ses contextes d'emploi. Nous pouvons par exemple nous demander si l'on a des usages analogues ou différenciés entre le singulier et le pluriel de *resource*. On trouve que *resource* et *resources* ont quelques cooccurrents en commun tels que *natural* et *use*, mais que certains autres cooccurrents sont plutôt en affinité avec le singulier (*efficiency*, *decoupling*, *panel*, *base*, *impacts*, *sustainability*, *extraction*, *consumption*...), et d'autres encore avec le pluriel (*water*, *management*, *biological*, *land*, *sustainable*, *living*, *conservation*, *marine*...). Sémantiquement, cela rejoint un comportement souvent observé en textométrie : le singulier du nom prend un sens plus abstrait, théorique et général ; ici il semble aussi lié à la description des processus autour des ressources ; quant au pluriel, il semble en affinité avec un sens plus concret, des contextes plus détaillés, et l'évocation d'entités plutôt que de processus. Dans le cas du mot *resource* donc, il peut être intéressant pour certaines analyses de considérer séparément le singulier et le pluriel.

³⁵ Requête CQL : [enlemma=".*(chemical|emission|pollut|wast).**%c]

³⁶ Requête CQL : [enlemma=".*(disease|health|mortality).**%c]

³⁷ Requête CQL : [enlemma=".*(financ|insurance|invest|subsid).**%c & word!=".*investig.**%c]

³⁸ Requête CQL : [enlemma="justice|law|rights?|responsibility"%c]

³⁹ Requête CQL : ".*resource.**%c ou [enlemma=".*resource.**%c] (même résultat)

3.3. Contexte situationnel et affinités contextuelles

3.3.1. La répartition des mentions : à quelle période ? dans quels textes ? dans quels chapitres ? Spécialisations textuelles en diachronie

Tableau 7. Fréquence des 5 thématiques dans les textes⁴⁰

	1972 palme	1972 stockholm	1987 brundt land	1992 rio	2015 unep	2017 unep	2019 unep	2021 03 nase	2021 04 nase	Total
Taille	2257	2435	154850	180549	15725	54797	82235	10726	2773	506347
1POLLUTION	9	4	542	908	32	588	380	3	4	2470
2HEALTH	1	3	213	441	35	104	132	6	12	947
3JUSTICE	2	13	235	207	7	24	39	20	0	547
4FINANCE	0	2	271	347	21	85	60	5	3	794
5RESOURCE	10	10	619	925	258	727	911	10	1	3471

Les décomptes permettent de hiérarchiser les thématiques à partir des fréquences totales (**Tableau 7**). Ainsi, les ressources apparaissent comme une préoccupation majeure dans le corpus, suivies par la pollution et la santé ; la dimension finance devance les aspects justice, ces deux thématiques étant moins représentées et semblant employées de façon plus variable dans le corpus, en étant même absentes de certains textes. Cependant, le recours à une mesure statistique est utile pour mieux apprécier les variations : la textométrie propose à cet effet le calcul des spécificités.

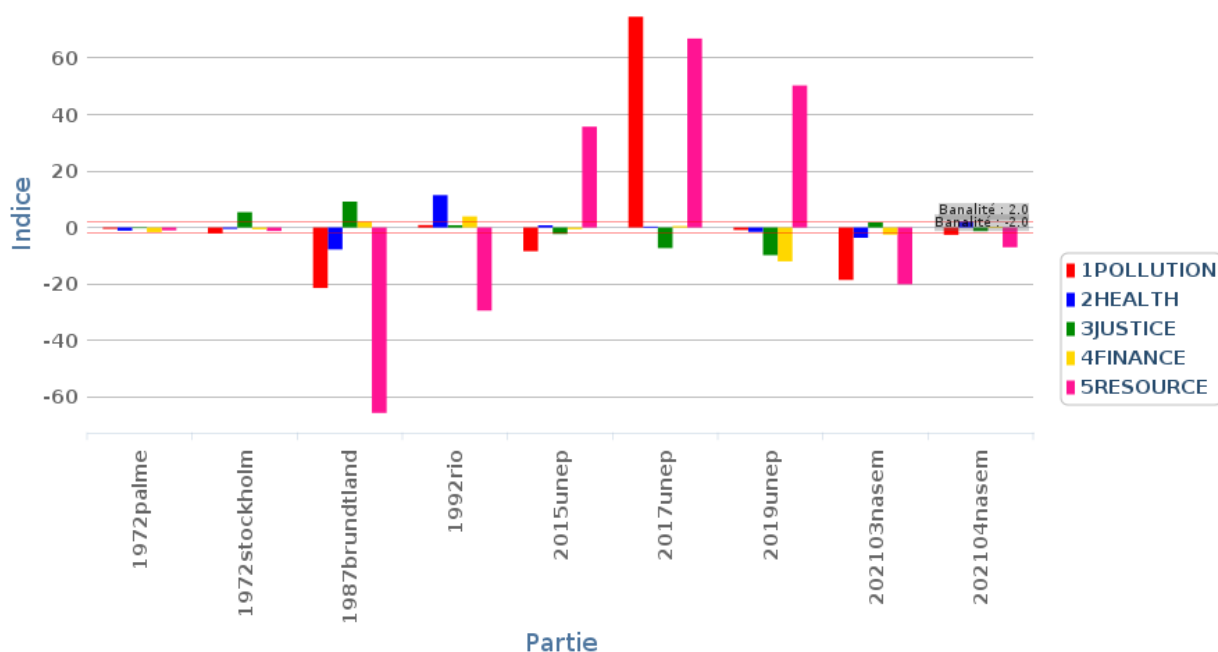


Figure 5. Répartition (Spécificités) des 5 thématiques POLLUTION (pollution, waste...), HEALTH (health, diseases...), JUSTICE (rights...), FINANCE (investment...) et RESOURCE (resources...) dans les textes du corpus

⁴⁰ Les légères différences de décompte avec les effectifs indiqués pour les thématiques viennent du fait qu'ici dans le tableau chaque mot n'est comptabilisé qu'une seule fois, et donc un mot comme "pollution-health" a été ici compté dans le thème 1POLLUTION et manque donc dans le thème 2HEALTH.

Un diagramme de spécificités calculées sur les 5 thématiques (**Figure 5**) donne à lire les informations suivantes :

- La **pollution** ressort comme sur-représentée dans **2017unep** ; ses mentions semblent en revanche sous-représentées dans 1987brundtland, mais on verra plus tard que c'est un effet un peu trompeur du caractère très large du rapport Brundtland et de son organisation en chapitres thématiques : le terme *pollution* est très densément présent dans le chapitre consacré au sujet.
- La **santé** semble particulièrement présente dans **1992rio**, mais aussi dans **202104nase** (la notion dépasse le seuil de spécificité malgré la courte taille du texte qui le pénalise statistiquement).
- Les préoccupations de **justice** sont plus verbalisées dans **Stockholm** et **Brundtland** (notamment chapitres finaux, 10 sur les *Commons* et 12 sur l'institutionnel) ; elles sont en contrepoint moins présentes dans les rapports UNEP.
- Pour les aspects **financiers**, le vocabulaire semble plus fréquent dans **1992rio**, mais cela pourrait être en partie lié à des phraséologies récurrentes (ce texte présente beaucoup de reprises et de répétitions internes, la mise au point du texte semble avoir utilisé de nombreuses briques rédactionnelles sous forme de phrases génériques et de paragraphes-types réutilisés et adaptés de section en section).
- Les **ressources** font clairement partie du vocabulaire central et très fréquent dans les **rapports UNEP** ; le sous-emploi dans le rapport Brundtland doit bien être compris comme relatif, il y a tout de même 619 occurrences dans ce rapport (mais elles apparaissent plus "diluées" vue la longueur).

Les spécificités donnent une appréciation globale, texte par texte. On peut compléter cet indicateur statistique global par une courbe de Progression, qui visualise le détail des apparitions des occurrences au fil du corpus.

[Encart explicatif Progression] Les courbes de Progression ont en abscisse le numéro d'ordre de chaque mot dans le corpus (dans l'ordre chronologique des textes et syntagmatique à l'intérieur des textes), et en ordonnée le nombre d'occurrences rencontrées depuis le premier mot du corpus. Ce sont des courbes cumulatives, elles ne peuvent donc qu'être ascendantes. Leur lecture doit s'attacher non pas à leur position relative (au-dessus ou en-dessous l'une de l'autre) mais à leur pente : une pente fortement croissante dénote une concentration d'apparition du thème ; à l'inverse, un plateau horizontal rend compte d'un passage sans mention du thème.

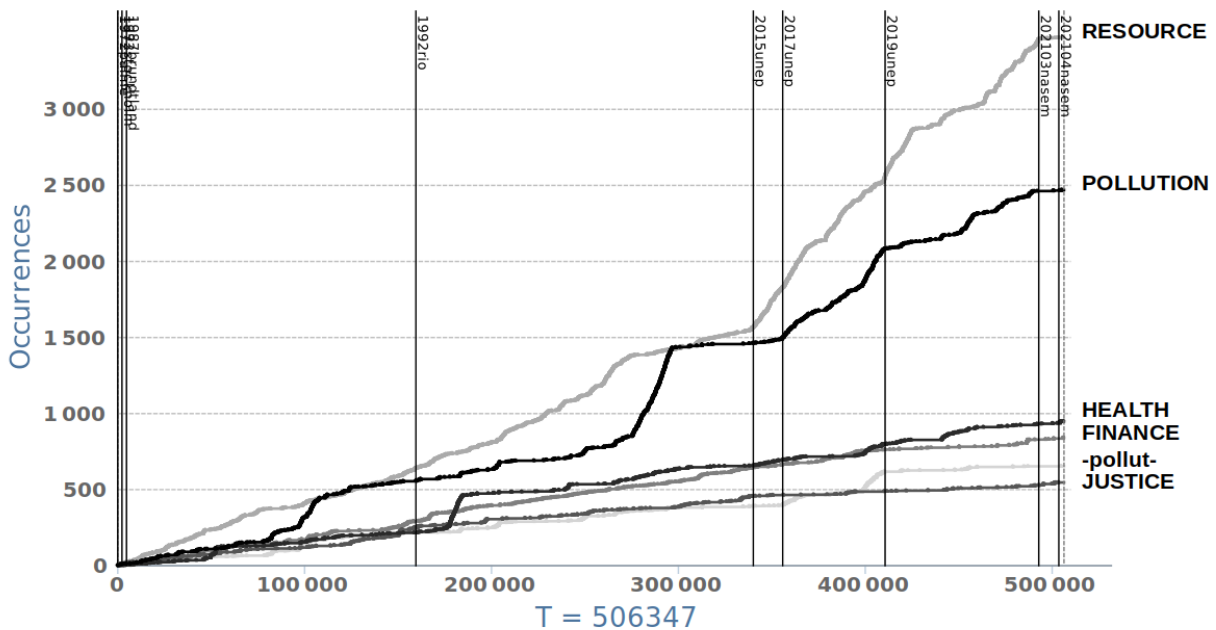


Figure 6. Évolution (Progression) des 5 thématiques POLLUTION (pollution, waste...), HEALTH (health, diseases...), JUSTICE (rights...), FINANCE (investment...) et RESOURCE (resources...). La 6e courbe (tracée avec le gris le plus clair) représente les mots de la famille de pollution.

La figure de Progression (**Figure 6**) confirme la forte présence de la thématique **RESOURCE**, relativement régulière, et plus marquée dans les rapports UNEP ; elle fait apparaître sa moindre présence dans la seconde moitié de 1992rio (à partir du chapitre 19 de l'Agenda 21, sur la gestion des produits toxiques) et dans les documents de la NASEM (que le calcul statistique mettait moins en évidence du fait de leur petite taille). La Progression montre en revanche un usage moins régulier de la thématique **POLLUTION**, laquelle est plus concentrée dans certains chapitres, en particulier pour les deux gros rapports 1987brundtland et 1992rio.

Cependant la lecture des courbes pour les quatre autres thématiques, de fréquence totale moindre et donc de moindre altitude, est plus difficile, d'où le "zoom" en figure suivante (**Figure 7**) pour mieux observer leurs variations de pente.

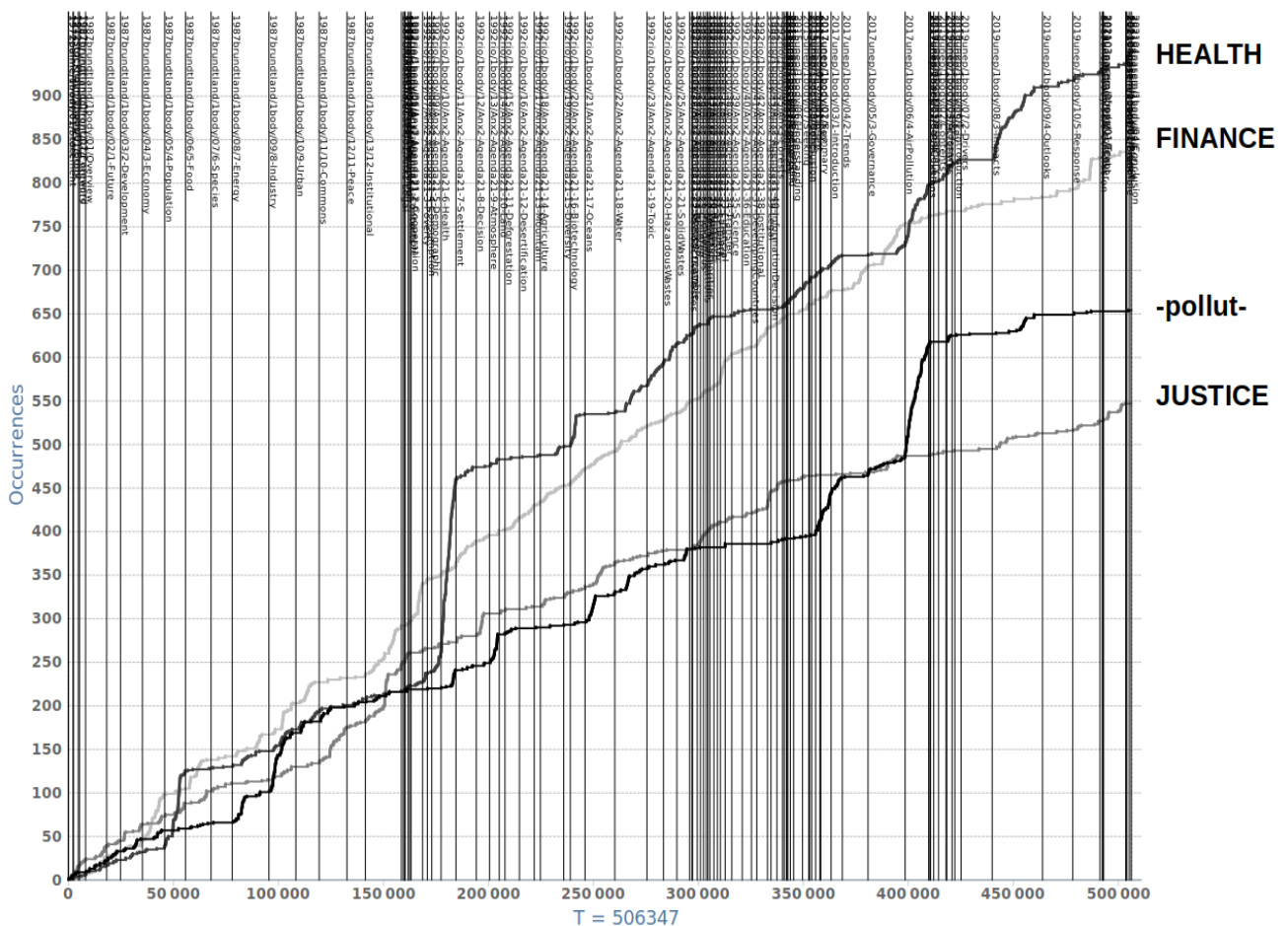


Figure 7. Évolution (Progression) : focus sur les chapitres des textes et sur les mots de la famille de pollution et les thématiques HEALTH (health, diseases...), JUSTICE (rights...), et FINANCE (investment...).

Cette nouvelle figure (**Figure 7**) permet de poursuivre l'examen des quatre thématiques moins fréquentes :

- Les thématiques de **pollution** et de **santé** ont tendance à avoir des zones de très forte concentration (**chapitres dédiés**) ; certains moments sont communs : les chapitres de Brundtland sur l'industrie (8) et la ville (9), dans l'agenda 21 de 1992rio la toute fin du chapitre santé (6), et dans 2019unep le chapitre 3 sur les impacts ; il ressort aussi que le rapport UNEP 2017 semble conjuguer très étroitement les deux notions à travers tout son propos.
- Les mentions relatives à la **finance** sont relativement régulières (un peu moins dans Brundtland, où se succèdent des zones de concentration et des zones d'absence) ; mais un tassement a tendance à apparaître à partir de 2019 (pente générale moins forte).
- Le vocabulaire lié à la **justice** prend une plus grande place dans les premiers textes ; il s'efface de façon assez notable des rapports UNEP (sachant que l'UNEP se charge davantage de questions matérielles environnementales) mais le **premier NASEM** (202103) lui donne à nouveau une place analogue à celle des premiers textes.

De ces différentes observations de répartition intertextuelle et intratextuelle, nous retenons que les cinq thématiques traversent tout le corpus, mais non sans certaines affinités. Le discours UNEP apparaît particulièrement centré sur les **ressources**, ce à quoi se superpose la place prise par la **pollution** dans le rapport 2017unep. La **santé** est massivement abordée dans 1992rio et 202104nasem, mais fait aussi l'objet de **chapitres dédiés** dans tous les autres gros rapports (1987brundtland, 2017unep, 2019unep) ; il est notable que plusieurs **chapitres associent étroitement santé et pollution**. Le vocabulaire de la **justice** apparaît plus présent dans les documents plus

anciens (en particulier 1972stockholm et 1987brundtland) et correspond moins à l'angle de vue des documents UNEP.

3.3.2. Attirances à l'échelle du paragraphe

Nous pouvons par ailleurs mettre à profit le calcul de cooccurrences pour évaluer plus précisément les attirances contextuelles réciproques des notions sur lesquelles nous centrons notre étude. Dans ce travail nous avons défini le contexte de cooccurrence comme un empan d'une distance de 50 mots à gauche et à droite, de l'ordre du paragraphe⁴¹.

[Encart explicatif cooccurrence] Le calcul des cooccurrences pour un mot P (dit pivot) consiste à calculer les spécificités de l'ensemble des contextes du mot P par rapport au corpus total. Autrement dit, un mot M qui a un indice de cooccurrence de 5 avec le mot P est un mot qui n'avait qu'une chance sur cent mille (le nombre de zéros correspond à la valeur de l'indice) d'apparaître avec une fréquence aussi élevée au voisinage du mot P. La manière de définir le voisinage (contexte) est un paramètre important du calcul. Il peut être défini par une distance en nombre de mots, à gauche ou/et à droite du pivot. Dans le logiciel TXM, il peut également être défini par toute unité structurelle (phrase, paragraphe, tour de parole, etc.) qui a été codée dans le corpus.

On notera tout d'abord, en revenant au **Tableau 6**, le **lien statistique très fort de health et pollution** entre eux (indice statistique entre 16 et 20⁴²), autrement dit les passages parlant de santé ont tendance à évoquer la pollution, et vice-versa. Les deux notions apparaissent contextuellement liées dans le corpus, non seulement donc au niveau des chapitres mais aussi à celui de passages plus étroits. Près de 140 contextes réunissent les mots (lemmes) *health* et *pollution*, de la Déclaration de Stockholm (1972) à celle de la NASEM (2021)⁴³, et à travers de multiples chapitres à l'intérieur des documents longs.

Si nous nous intéressons au lien du mot *resource* avec nos deux thématiques HEALTH et POLLUTION, nous obtenons l'information que *resource* (au singulier) apparaît de façon marquée (plus souvent qu'en moyenne) dans les contextes de la thématique POLLUTION (indice de 7). Si au lieu de *resource* on considère le mot **energy, le lien avec la thématique POLLUTION devient très marqué** (indice statistique de 15), car de fait plus du tiers des nombreuses occurrences d'*energy* (269 sur 753) sont dans un voisinage de 50 mots d'un mot de la thématique POLLUTION. En revanche pour la santé, l'attirance contextuelle avec *resource* reste très marginale (elle tient essentiellement à 12 contextes communs avec *resource-related* sur les 26 apparitions de ce mot, rencontres toutes concentrées dans 2019unep).

Aucune attirance n'est contextuellement décelée au niveau des paragraphes entre les mots du thème FINANCE et ceux de la pollution ou de la santé. Si ces thèmes peuvent se rencontrer (par exemple on peut dénombrer 173 apparitions d'un mot de la pollution et d'un autre de la finance à moins de 50 mots de distance⁴⁴), ces rencontres ne se produisent pas avec une fréquence supérieure à ce qu'on peut attendre vu les fréquences des deux thématiques et la taille du corpus.

⁴¹ Le logiciel permettrait de choisir le paragraphe comme unité de contexte, mais c'est notre corpus qui ne dispose pas de cette information (celle-ci était trop complexe à obtenir de façon formelle depuis les formats PDF de nos sources, avec les moyens à notre disposition).

⁴² Soit moins d'une chance sur 10 millions de milliard (1 suivi de 16 zéros, probabilité de 10E-16) que les rencontres des mots dans les mêmes contextes puissent être attribuées au hasard.

⁴³ À l'exception de deux documents courts, 1972palme et 202103nasem.

⁴⁴ Requête CQL :

```
([enlemma=".*(chemical|emission|pollut|wast).**%c"]{0,50}[enlemma=".*(financ|insurance|investm|subsid).**%c & word!=".*investig.**%c"]|([enlemma=".*(financ|insurance|investm|subsid).**%c & word!=".*investig.**%c"]{0,50}[enlemma=".*(chemical|emission|pollut|wast).**%c"]
```


Pour ce qui concerne la thématique JUSTICE, un seul lien apparaît, celui du mot *right* (au singulier) avec la santé (avec 21 contextes communs, soit concernant près du quart des 87 occurrences de *right*), mais dans des voisinages larges : on évoque un droit en relation avec la santé, mais jamais directement un “droit à la santé”.

Si donc les 5 thématiques RESSOURCE, POLLUTION, HEALTH, JUSTICE et FINANCE ont une forte présence à l'échelle globale du corpus, **seul le lien entre santé et pollution semble particulièrement exprimé dans le détail des contextes**, au sens où ces deux notions semblent s'attirer mutuellement et se concentrent dans les mêmes passages.

3.3.3. Autres évolutions lexicales à noter en toile de fond

En s'intéressant aux autres notions très présentes dans le corpus (les 100 noms communs les plus fréquents du **Tableau 3**), on peut repérer les **notions dont l'usage se fait moindre au fil du temps** (**Figure 8**) (on peut faire des regroupements d'ordre sémantique : **development, environment et conservation** ; **country, area et government** ; **industry et chemical**), et inversement, des **notions dont l'usage s'intensifie**⁴⁵ (**Figure 9**) (**sustainability et goal** dont les *Sustainable Development Goals, climate* et *change* -souvent combinés en *climate change*-, **emission, example**).

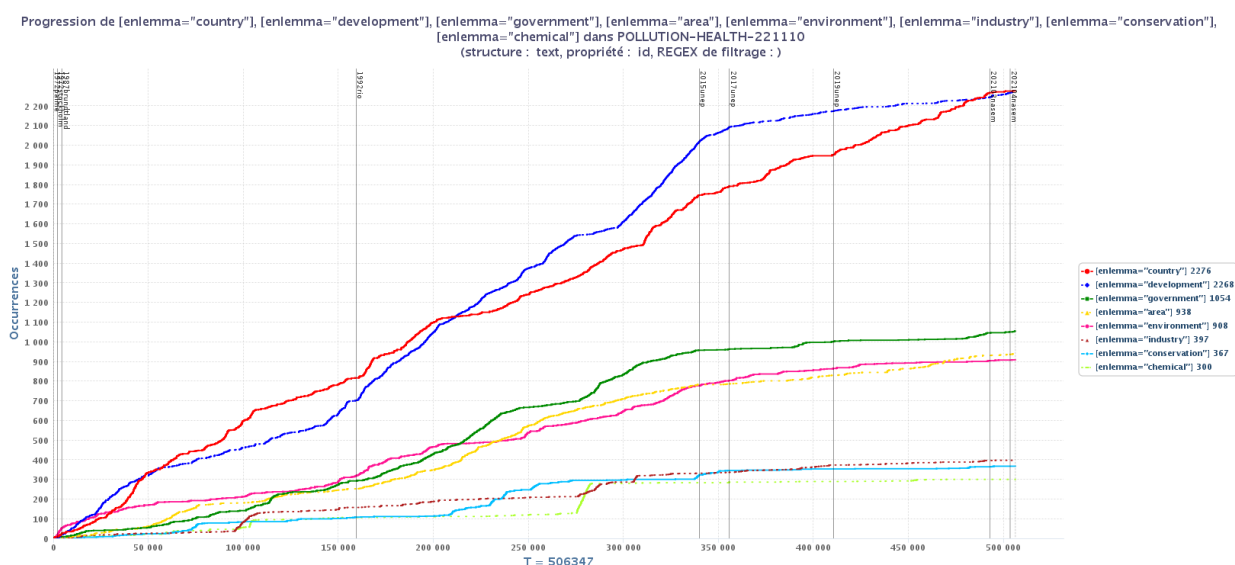


Figure 8. Vocabulaire en décroissance. Courbes cumulatives au fil du corpus (Progression) des 8 noms communs sur les 100 plus fréquents du corpus présentant un usage qui s'amenuise au fil du temps.

⁴⁵ La récolte n'est pas si abondante car on a veillé à ne pas considérer certains noms devenant très fréquents dans les rapports UNEP, mais peu représentés dans les textes de la NASEM (par exemple : *impact, material, consumption, efficiency, etc.*) : l'interprétation serait alors plutôt de les considérer comme spécifiques du discours UNEP, plutôt que comme spécifiques de l'air du temps.

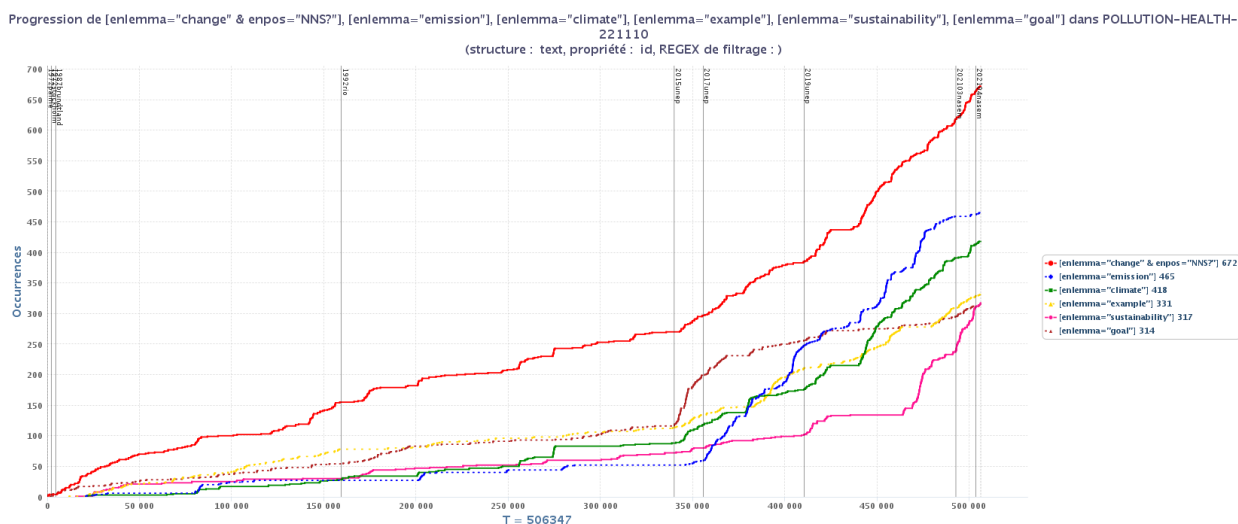


Figure 9. Vocabulaire montant. Courbes cumulatives au fil du corpus (Progression) des 6 noms communs sur les 100 plus fréquents du corpus présentant un usage qui s'intensifie au fil du temps.

À l'occasion de cette étude focalisée sur la répartition des 100 noms les plus fréquents, on remarque aussi que les rapports **UNEP semblent éviter un vocabulaire négativement** connoté en se distinguant par le sous-emploi tant de *problem* que de *risk*, *effort* et *issue* (**Figure 10**).

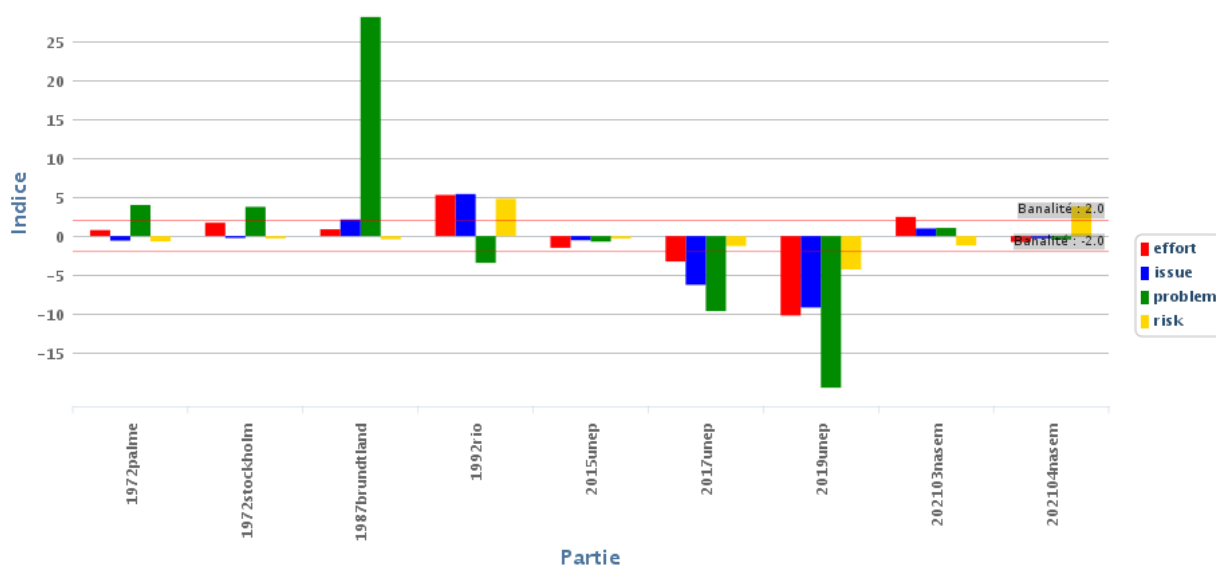


Figure 10. Répartition d'un vocabulaire négatif/pessimiste. Spécificités des noms effort, issue, problem et risk sur les neuf textes du corpus.

Enfin, nous avons par ailleurs noté la bonne présence de la question démographique avec la forte fréquence du nom *population*, tout en nous interrogeant sur la façon dont cette question se déploie et s'articule dans le corpus. Un graphe de Progression (**Figure 11**) comme un diagramme de spécificités (**Figure 12**) mettent en évidence que c'est un sujet toujours présent mais qui n'occupe le devant de la scène que dans le rapport Brundtland (1987). La croissance de la population est pensée en lien avec la croissance économique et la préservation du bien-être, misant sur la technologie (1992) qui se précise en efficacité énergétique et en recyclage. A contrario, une notion qui pourrait être directement liée mais plus exigeante comme la sobriété ne semble pas une seule fois évoquée.⁴⁶

⁴⁶ Aucun mot du corpus ne contient la suite de caractères "sob". Nous ne pouvons cependant pas formellement conclure à l'absence certaine du mot du fait des coquilles issues de l'extraction des textes à partir des PDF.

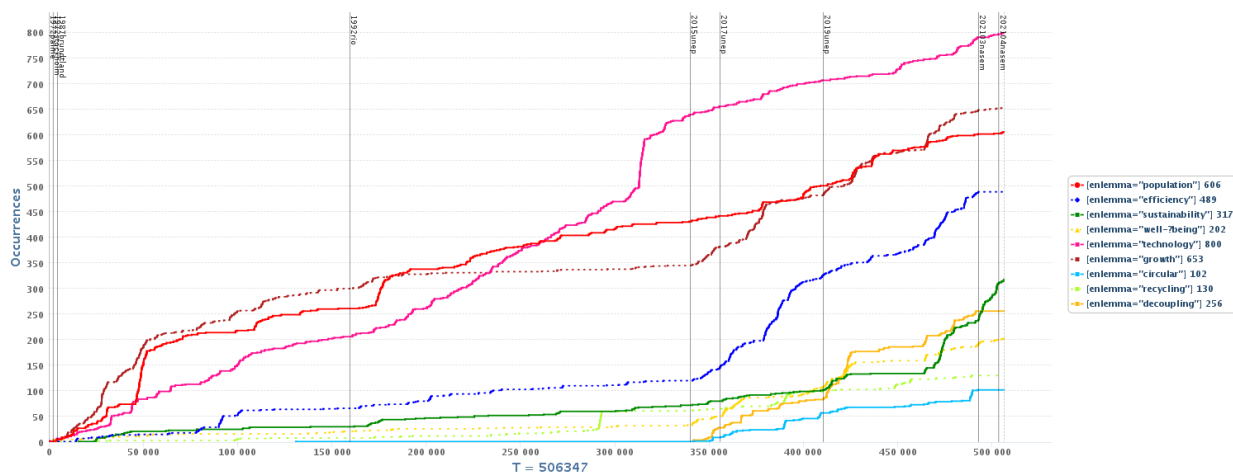


Figure 11. Évolution (Progression) de *population* en regard de *growth* et de 7 autres notions liées à une forme de progrès matériel (technology, efficiency, sustainability, decoupling, well-being, recycling, circular).

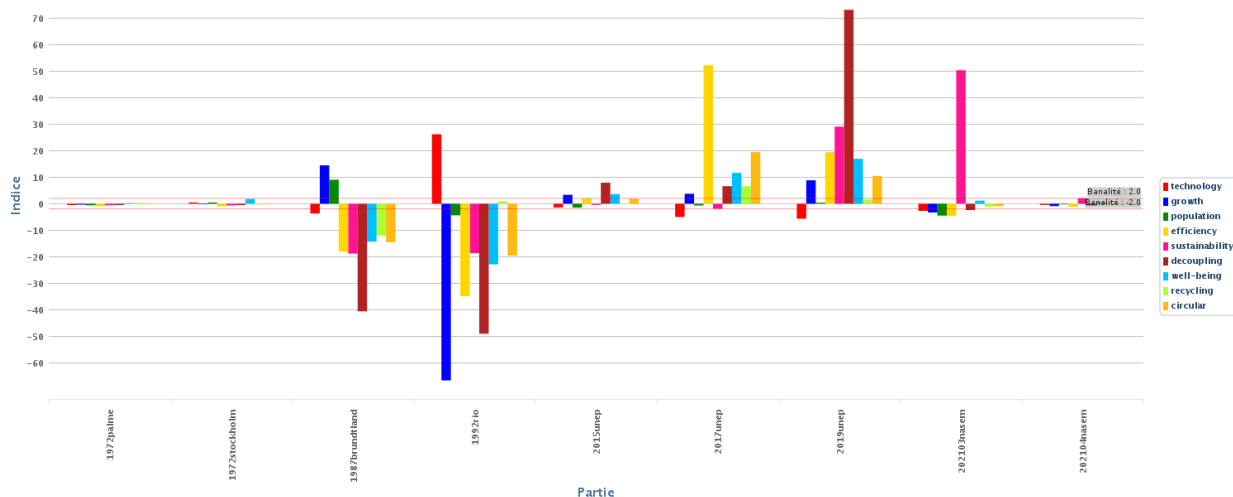


Figure 12. Répartition (Spécificités) de *population* en regard de *growth* et de 7 autres notions liées à une forme de progrès matériel (technology, efficiency, sustainability, decoupling, well-being, recycling, *circular*).

3.4. De la syntaxe à la sémantique : observation systématique des constructions dans lesquelles s’ancrent les notions à l’étude

Dans cette section, nous examinons de façon systématique les syntagmes construits avec un mot précis donné (d’abord *health*, puis *pollution*)⁴⁷, et non plus des familles de mots ou des thématiques. En effet, ce travail portant sur des régularités syntaxiques est plus efficace en travaillant sur une catégorie grammaticale donnée. Il se limite à ce qui est exprimé à travers les relations syntaxiques immédiates, ce qui est à garder à l’esprit à la lecture des sections suivantes : les décomptes portent sur des expressions explicites, pas sur la présence sémantique de l’idée, qui pourrait passer par d’autres formes d’expression plus diffuses. Mais les hiérarchies quantitatives et ordres de grandeur révèlent déjà certaines orientations et certains équilibres (ou déséquilibres).

⁴⁷ Ces mots seront pointés par les requêtes suivantes : [enlemma="health"%c] et [enlemma="pollution"]. (Notons que le modifieur %c est nécessaire dans un seul des cas, l’outil TreeTagger associé plusieurs lemmes à “health” mais un seul à “pollution”).

3.4.1. Les constructions d'emploi du nom *health*

3.4.1.1. Un contexte dominant : les effets sur la santé, en lien avec l'environnement

Le nom *health* est très souvent employé dans des contextes qui parlent d'*impact(s)* (ou *risk(s)*, *effect(s)*, *problem(s)*, *hazard(s)*) pour la santé (*health* et souvent *human health*), en particulier en combinaison syntagmatique avec *environment(al)*, principalement en coordination (*and*)⁴⁸. Ce motif lexico-syntaxique représente 191 occ. des 726 occurrences du nom *health*, soit plus du quart des usages de *health* (26,3 %), avec 48 réalisations différentes (**Tableau 8**).

Tableau 8. Liste des 48 réalisations du motif lexico-syntaxique de l'impact sur la santé (avec compléments optionnels *human* et *environnement(al)*) : relevé alphabétique et nombre d'occurrences dans le corpus (fréquence totale : 191 occurrences)

Expression	Fréquence
effect on human health	1
effects on environment and human health	1
effects on health and the environment	1
effects on human health	2
environmental and health effects	2
environmental and health impacts	5
environmental and human health impacts	2
environmental health hazards	3
environmental health impact	2
environmental health problems	2
hazards to health and environment	1
hazards to health and the environment	1
hazards to human health	1
hazards to human health and the environment	1
health and environmental effects	1
health and environmental impact	2
health and environmental impacts	5
health and environmental risks	2
health and environment effects	1
health and environment impacts	1
health and environment risks	1
health effects	10
Health effects	1
health hazard	1
health hazards	3
health impact	2
health impacts	64
Health impacts	4
Health Impacts	2
health problem	4
health problems	8
health risk	8
Health risk	1
health risks	14
human health and the environment hazards	1

⁴⁸ Requête CQL (en stratégie *longest*) :

```
(((([enlemma="human"]?[enlemma="health"%c]([enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="environment(al)"]?)))([enlemma="environment(al)"]?[enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="human"]?[enlemma="health"%c]))[enlemma="effect|impact|risk|problem|hazard"])([enlemma="effect|impact|risk|problem|hazard"])[word="on|to"%c]([enpos="DT|PP\$"]?[enlemma="environment(al)"]?[enlemma="and"]?)?[enpos="DT|PP\$"]?[enlemma="human"]?[enlemma="health"%c][enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="environment(al)"]?))
```

human health effects	2
human health impacts	6
human health risk	1
human health risks	1
impact on health and	2
impact on health and the environment	1
impact On human health and the environment	1
impacts on human health	3
impacts on human health and the environment	5
impacts on the environment and human health	1
risks to human health	1
risks to human health and the environment	4
risk to human health and the environment	2

Dans cet ensemble, on peut remarquer que les expressions qui combinent *health* et *environment(al)* comptabilisent 49 occurrences⁴⁹. Il ne s'agit pas d'une expression figée, 25 manières différentes de combiner les mots sont attestées. Ces expressions se trouvent à toutes les périodes puisque les 49 occurrences se répartissent comme suit (**Figure 13**): 6 dans 1987brundtland (toutes différentes), 31 dans 1992rio (19 formes différentes), 1 dans 2015unep, 3 dans 2017unep (dont 2 *environmental and health impacts*, dans une phrase de l'introduction reprise dans la synthèse) et 8 dans 2019unep (sous 4 formes différentes, dont les deux *environmental and human health impacts*). Notons que cette dernière construction, *environmental and human health impacts*, ouvre une ambiguïté sur le rattachement syntaxique de *environmental* : *environmental impacts* semble attendu vu le paradigme des expressions analogues, mais *environmental health* peut aussi être perçu, ce qui pourrait refléter l'élargissement de la conception de la santé. Nous y reviendrons plus loin.

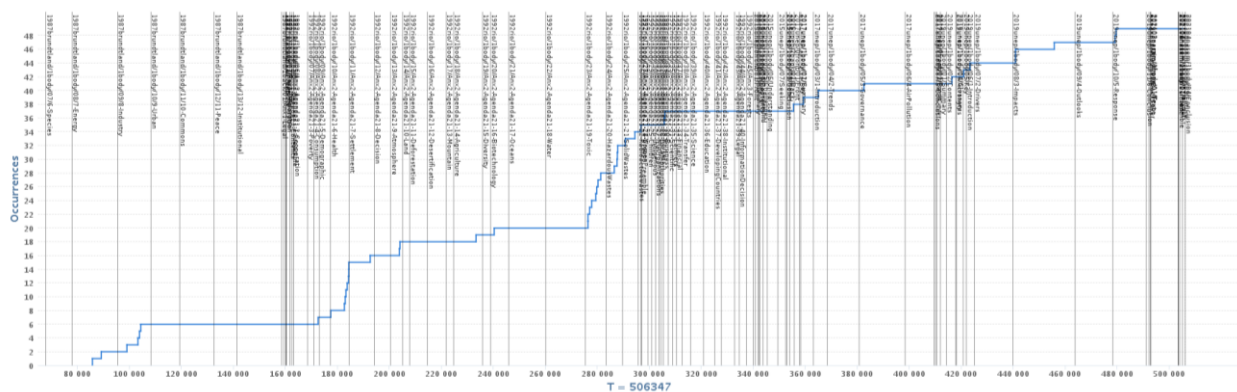


Figure 13. Évolution (Progression) des 49 expressions du motif de l'impact sur la santé et l'environnement. La courbe montre que ces expressions se répartissent sur une large période (de 1987brundtland à 2019unep), mais apparaissent de façon plus intense dans certains chapitres de 1987brundtland et de 1992rio (au niveau des hautes marches du graphique).

On retient que **la santé est le plus souvent mentionnée comme exposée** (à des risques, impacts, etc.) (cela concerne plus du quart des occurrences). Il s'agit généralement de santé humaine, mais avec un **lien à l'environnement réaffirmé tout au long du corpus, et peut-être formulé de façon encore plus étroite dans les années les plus récentes**. Cela lui dessine une place potentiellement centrale dans les questions environnementales.

⁴⁹ Requête CQL :

`((([enlemma="human"]?[enlemma="health"%c][enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="environme nt(al)?"]])|([enlemma="environment(al)"?][enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="human"]?[enlem ma="health"%c]))[enlemma="effect|impact|risk|problem|hazard"]|([enlemma="effect|impact|risk|problem|haz ard"]|[word="on|to"%c]([enpos="DT|PP\$"]?[enlemma="environment(al)"?][enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="human"]?[enlemma="health"%c])|([enpos="DT|PP\$"]?[enlemma="human"]?[enlemma="heal th"%c][enlemma="and"]?[enpos="DT|PP\$"]?[enlemma="environment(al)"?])))`

3.4.1.2. De quelle santé s'agit-il ?

Santé humaine

Les mentions de la santé font majoritairement référence à la santé personnelle des êtres humains : soit directement avec **human health** (100 occurrences), soit à travers des expressions récurrentes comme *public health* (20 occ.) ou *primary health care* (10 occ.) ou des qualificatifs évaluatifs usuels (*good* ou *poor health*, 3 occ. chacun, *ill health* 2, *proper health* 1).

Une attention particulière est accordée aux **femmes, aux mères et aux enfants** (*reproductive health* 10 occ., *maternal health* 3, *health of children, youth (and) women* 2, etc., pour au total une vingtaine d'occurrences).

Enfin tout au long de son Agenda 21, **1992rio va insister sur l'association entre preventive and curative health** (6 occurrences⁵⁰ réparties sur les chapitres 3 (pauvreté), 5 (démographie), 6 (santé) et 24 (femmes)). L'intérêt pour des solutions préventives aussi bien que curatives est intéressant, mais reste marginal par rapport à l'ampleur de la tâche.

Santé sociale

La santé est également considérée au **niveau sociétal, mais dans une moindre mesure** (une quinzaine d'occurrences), avec des expressions comme *occupational health and safety (of workers)* (3 occurrences), ou *urban health* (3 occ.), *health of city inhabitants*, *regional health* (1 occurrence chacune). 1972stockholm envisage de façon large "the physical, mental and social health of man" (Introduction, §3). Au chapitre 6 (sur la santé) de son Agenda 21, 1992rio parle de *community-based health* (1 occ.) et *community health* (2 occ.) ; et 1987brundtland (p. 88, au chapitre 4 sur les questions de population) désigne une *national health* :

First, although generally speaking national wealth buys national health, some relatively poor nations and areas, such as China, Sri Lanka, and the Indian state of Kerala, have achieved remarkable success in lowering infant mortality and improving health through increases in education, especially of women; the establishment of primary health clinics;

etc. (c'est nous qui soulignons).

Santé écologique

Mais à un troisième niveau, la santé concerne également les autres êtres vivants et tout l'environnement.

Dans 1987brundtland et 1992rio, cette santé environnementale est formulée de façon relativement détaillée et concrète, elle se décline comme *animal health* (3 occ.), mais aussi (occurrences uniques) : *crop health / health of the crop*, *health of forest ecosystems / health of trees and forests*, *health of the oceans*, *livestock health*. Par la suite, les rapports UNEP se préoccupent de la santé globale de milieux comme les terres agricoles (*soil health*, 4 occ.) et les milieux aquatiques (*health of water related ecosystems, health of water systems*).

On observe de fait que la perspective devient résolument générale dans les textes plus récents.

Ainsi, *ecosystem health* (qui apparaît 11 fois en tout) a 3 occurrences dès 1992rio, au chapitre 5 sur la démographie : la santé de l'environnement est un facteur (*factor*) du bon développement de la population. Mais la **relation entre santé humaine et santé de l'écosystème est perçue de façon plus symétrique et équilibrée à partir des rapports UNEP** (2015, 2017, 2019), où l'on présente les deux réalités comme analogues et liées, avec des co-bénéfices (*co-benefits*), par exemple dans les trois extraits suivants :

⁵⁰ En ajoutant deux variantes : *prevention and curative health practices, preventative and curative health*.

positive links between goals on health, poverty, ecosystem health and climate change (Lim et al., 2013) (2015unep, p. 34)

The strong interlinkages that exist between human well-being and ecosystem health (2015unep, p. 41)

human well-being is intrinsically linked to the health of the natural ecosystem (2015unep, p. 13)

Dans leurs parties de synthèse, 2015unep et 2017unep appellent à agir pour la santé des ressources naturelles :

restoring and maintaining the health of the natural resource base is a necessary condition

Dans les termes mêmes, et d'un point de vue quantitatif, on relève 78 mentions (de 11 formes différentes) qui coordonnent santé et environnement⁵¹ (**Tableau 9**).

Tableau 9. Liste des coordinations entre santé et environnement. Relevé alphabétique et nombre d'occurrences dans le corpus

Expression	Fréquence
environmental and health	15
environmental and human health	3
environment and human health	2
health and environment	8
health and environmental	13
health and the environment	9
human and environmental	1
human and environmental health	2
human health and environment	1
Human health and environmental	1
human health and the environment	23

Les constructions coordonnant santé et environnement (**Tableau 9**) se répartissent de 1987 à 2019 : 13 dans 1987brundtland, 49 dans 1992rio, 1 dans 2015unep, 6 dans 2017unep, 9 dans 2019unep. Mais dans les occurrences plus récentes, la coordination se déplace et les catégories ne sont plus santé d'une part et environnement de l'autre, mais une même santé qui concerne d'une part l'être humain et d'autre part son environnement. On relève ainsi :

- deux occurrences de *human and environmental health* dans 2017unep
- deux occurrences de *environmental and human health*, mais aussi *human well-being and ecosystem health* dans 2019unep.

La coordination suggère que l'homme et l'environnement sont vus comme deux sujets entiers et complémentaires de santé.

Avec les tout derniers documents, la santé devient planétaire et unifiée / inclusive, ce qui ouvre sur des questions éthiques. Ainsi lit-on dans 202104nasem (section sur la planète) :

Planetary health. The health of nature, our planet, and people is tightly connected

et aussi

“ one health ” — the intimate connections between human health and the health of other animals and the environment.

⁵¹ Requête CQL (en stratégie *longest*) :

`((([enlemma="human"]?[enlemma="health"%c])[enlemma="human"])([enlemma="and"])[enpos="DT|PP\$"]?[enlemma="environment(al)"]?[enlemma="health"%c]?)([enlemma="environment(al)"]?[enlemma="health"%c]?[enlemma="and"])[enlemma="human"]?[enlemma="health"%c]))`

(c'est nous qui soulignons).

Emploi métaphorique

De façon figurée, la santé s'étend même à des abstractions, en particulier l'économie, inscrite dans le social et l'environnement : deux occurrences de *health of the economy* (dans 2015unep, p. 6 et 17), une de *health of national economies* (1987brundtland, chapitre 9, p. 212).

Le contexte élargi peut se révéler embrasser d'autres objets plus ou moins abstraits, comme dans 2015unep, p. 6 (Summary) :

Outdated strategies that take inadequate account of the Earth's resources and natural capacities, in particular, often have major unintended consequences for the health of the economy, society and natural environment, usually by shifting the problem being addressed to another country, social group or environmental medium, or by stunting the development of another sector.

De cette typologie issue d'un inventaire des syntagmes nominaux construits avec le nom *health*, il ressort que, dans notre corpus, **la santé est de façon prototypique la santé humaine, mais qu'une conception plus environnementale, plus globale et intégrative, se développe** jusqu'à prendre une place centrale dans les derniers documents (issus de la NASEM).

3.4.2. Les constructions d'emploi du nom *pollution*

3.4.2.1. Le pollué : qu'est-ce qui est affecté par la pollution ?

La pollution peut être décrite de façon très générale comme pollution de l'environnement (*environmental pollution*, 14 occurrences réparties dans nos quatre plus longs textes). Mais elle est beaucoup plus souvent évoquée en lien avec trois grands objets : l'air, l'eau, et la terre (les sols). La textométrie nous permet d'opérer un bilan quantitatif pour chacun de ces objets, par delà une grande variété de formulations. Ces relevés se sont basés sur un examen méticuleux de certaines constructions appropriées (*pollution of X*, adjectif ou/et nom + *pollution*)⁵² observées en concordances (triées sur le contexte, qui permettent de considérer exhaustivement et efficacement toutes les occurrences du mot *pollution*) (**Figure 14**) ou en index (liste d'expressions hors contexte, qui attire l'attention sur les constructions dominantes et récurrentes).

[Encart explicatif Concordance] La concordance d'un mot pivot P est le relevé exhaustif de ses contextes, présenté selon une disposition de tableau (chaque contexte dans une ligne du tableau), disposition qui s'avère efficace pour une lecture synthétique et comparative des contextes immédiats. Chaque contexte est identifié par une indication de localisation dans le corpus (par exemple dans notre corpus : le texte, la page, le chapitre à l'intérieur du texte). Le mot pivot (ou expression composée) est isolé dans une colonne centrale, de sorte que ses voisins immédiats se présentent de façon verticalement superposée. Il est possible de trier sur toute colonne : lorsque le tri porte sur une colonne de contexte, il se fait alphabétiquement sur le mot le plus proche du pivot, puis sur le mot une position plus loin, puis deux positions plus loin, etc., ce qui regroupe automatiquement les constructions analogues et permet de les faire ressortir visuellement.

⁵² Exemples de requêtes CQL utilisées pour générer des listes d'expressions potentiellement intéressantes :
[enpos="JJ.?[NNS?"]+[enlemma="pollution"]
[enlemma="pollution"][word="of"][enpos="DT|PP\\$|CC|JJ.?"]*@[enpos="NNS?"]

ref	Contexte gauche	Pivot	Contexte droit
1992rio, p. 240, Anx2-Agenda21-17-Oceans	as well as contingency plans for degradation and	pollution	of anthropogenic origin, including spills of oil and oth
1987brundtland, p. 019, Overview	the overuse of soil and chemicals, the	pollution	of both water resources and foods with these chemi
1992rio, p. 275, Anx2-Agenda21-18-Water	spread scarcity, gradual destruction and aggravated	pollution	of freshwater resources in many world regions, along
1987brundtland, p. 052, 2-Development	alinization and alkalinization of irrigated lands, nitrate	pollution	of ground-water, and pesticide residues in food. Ecol
1987brundtland, p. 105, 5-Food	land and watershed protection areas; and nitrate	pollution	of ground-water aquifers due to the often subsidized
1987brundtland, p. 175, 8-Industry	death 'of Lake Erie; the progressive	pollution	of major rivers like the Meuse, Kibe, and Rhine;
1987brundtland, p. 216, 10-Commons	1. The Geosynchronous Orbit 2. The	Pollution	of Orbital Space 3. Nuclear Power in Orbit 4. Towards
1987brundtland, p. 228, 10-Commons	by the regional conferences. 3. The	Pollution	of Orbital Space 70. Debris in orbit is a growing threa
1987brundtland, p. 054, 2-Development	. We are the first to feel the	pollution	of our waters, as the Ojibway peoples of my own hom
1972palme, p. 008, Statement	uncommon. To this is added the increasing	pollution	of regional and high seas. We expect that the draft co
1992rio, p. 287, Anx2-Agenda21-18-Water	of freshwater supply, degraded water quality and	Pollution	of surface and groundwater sources. Major problems
1987brundtland, p. 225, 10-Commons	regulations to 'prevent, reduce and control	pollution	of the marine environment from dumping'. It also rec
1992rio, p. 245, Anx2-Agenda21-17-Oceans	technologies to deal with the major types of	pollution	of the marine environment from land-based sources
1972stockholm, p. 003, II-Principles	States shall take all possible steps to prevent	pollution	of the seas by substances that are liable to create haz
1987brundtland, p. 220, 10-Commons	the exploitation of their coastal resources or the	pollution	of their waters by powerful countries or companies. T
1987brundtland, p. 106, 5-Food	, and Latin America; and waste and	pollution	of water almost everywhere. Within 40-70 years, glob

Figure 14. Contextes du mot pollution affichés en Concordance, avec tri sur le contexte droit, de façon à consulter de façon méthodique les expressions "pollution of X".

L'air

Les mentions de pollution affectant l'air (ou l'atmosphère) sont les plus massives : elles totalisent 157 occurrences⁵³. Un qualificatif peut préciser le type d'air (*indoor, outdoor, ambient*) ou la portée de la pollution (*transboundary, local, city, urban*).

L'eau

La pollution de l'eau est également largement évoquée (57 mentions), avec une certaine diversité de perspectives puisqu'il peut s'agir des sources et des nappes phréatiques (*pollution of ground-water aquifers* et cinq autres variantes de formulation, 6 occurrences), des fleuves et rivières (2 occurrences), des mers et océans (principalement *marine pollution*, et quatre autres variantes, totalisant une vingtaine d'occurrences).

Les sols

Dans les documents de notre corpus, la pollution des sols semble rarement évoquée : deux *land pollution*, deux *soil pollution*. Indirectement, on peut rapprocher encore une mention de *nutrient pollution* et une autre de *pollution of both water resources and foods*. Il y a donc peu d'indices lexicaux de ce type de pollution.

Nos différents relevés montrent donc que **la pollution de l'air apparaît davantage prise en compte que celle de l'eau et surtout celle des sols**. On peut l'observer aussi au niveau de l'organisation textuelle : le chapitre 4 de 2017unep est explicitement consacré à la pollution de l'air, et les chapitres du rapport Brundtland ou de l'Agenda 21 de 1992 adoptent des points de vue secteur par secteur (la nourriture, l'énergie, les industries, la ville..., pour le premier, l'atmosphère, la terre (*land*), l'agriculture, les océans, l'eau,... pour le second). Si donc la problématique des sols et de l'agriculture est bien présente dans notre corpus, elle ne semble pas aussi directement massivement formulée en termes de pollution.

⁵³ Il est possible de formuler une requête CQL opérant le décompte total :
[enlemma="air|atmospheric|indoor"] [enlemma="pollution"]

3.4.2.2. Le pollueur : quelles sources de pollution sont évoquées ?

On ne relève qu'une occurrence de la formulation très générale de pollution d'origine humaine (*pollution of anthropogenic origin*, 1992rio, p. 240, Agenda 21 ch. 17), mais une recherche sur *anthropogenic* trouve 13 occurrences dont 6 *anthropogenic* [...] *emissions* et autres *anthropogenic activities, interference, pressures, inputs*, etc. dans des contextes à connotation ou orientation négative.

Industrie, énergies, tissu urbain

La pollution est industrielle (16 occurrences, avec une demi-douzaine de variantes), urbaine (3 dont 2 *urban-industrial air pollution*), et quelquefois plus précisément liée à la combustion d'énergie (4 occurrences : *energy-related pollution, pollution caused by atmospheric pollutants from the combustion of fossil fuels, pollution from cook stoves, pollution from cooking fuels*). Soit donc une vingtaine d'occurrences pour cette pollution industrielle.

Déchets

La pollution issue des déchets (14 occurrences) est le plus souvent mentionnée comme *waste-related pollution* (7 occurrences) ou *end-of-pipe pollution* (3 occ.).

Agriculture, élevage, pêche

Nous avons décompté une dizaine d'occurrences de pollution pouvant se rapporter aux activités d'exploitation agricole (dites souvent *land-based*) ou maritime (souvent *sea-based*).

Transports

Du côté des moyens de transport, la pollution la plus fréquemment évoquée est celle des navires (*ships, cargo ships* et *vessels*) ; nous ne relevons qu'une seule mention directe de pollution automobile (dans 1987brundtland au chapitre 7, sur l'énergie, p. 150). Au total, sont dénombrées 8 occurrences pour cette pollution des transports.

Au total, ces décomptes suggèrent que la problématique du pollueur est moins abordée dans les documents que la description du pollué. Les activités industrielles et les déchets sont majoritaires.

3.4.2.3. Le paradigme de la pollution : à quoi la pollution est-elle associée (coordonnée) ?

Nous portons l'attention sur les entités (noms communs) qui s'articulent à la pollution à travers l'expression d'une coordination ("pollution and X", "X and pollution", en prévoyant aussi que des qualificatifs ou modificateurs puissent s'insérer devant les noms)⁵⁴.

Le nom qui apparaît le plus souvent coordonné à pollution est santé (*health*), d'ailleurs toujours dans l'ordre *pollution and health* (16 occurrences dont 10 exactement sous la forme précédente). Nous avons noté la corrélation entre les deux notions au niveau des chapitres, puis des paragraphes, elle se manifeste donc aussi jusqu'à l'étroit niveau du syntagme.

⁵⁴ Requêtes CQL :

[enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enlemma="pollution"] [word="and"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enpos="NNS ?"]

[enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enpos="NNS ?"] [word="and"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enlemma="pollution"]

Puis, pour un nom XXX trouvé grâce aux requêtes précédentes, décompte avec

(([enlemma="pollution"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[word="and"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enlemma="XXX"])([enlemma="XXX"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[word="and"] [enpos="JJ. ?|NNS ?|DT|PP\ \$|IN"]*[enlemma="pollution"])

Le deuxième nom entrant le plus en relation de coordination avec pollution est *resource(s)*, avec également 16 occurrences (4 *pollution and resource*, 3 *resource efficiency and pollution*, 2 *pollution reduction and more sustainable resource*, etc.). La coordination apparaît donc comme un contexte privilégié pour manifester un **lien linguistique entre pollution et ressources**.

Ces observations s'interprètent de façon assez directe : pollution, ressources et santé montrent des liens importants, car la pollution résulte de l'extraction et du traitement des ressources, avec des impacts sur la santé.

3.4.2.4. Les actions sur la pollution

Pour étudier l'expression des actions sur la pollution, nous nous sommes appuyés sur des relations prédicatives tant verbales⁵⁵ que nominales.

D'une façon générale, la pollution est quelque chose qu'il faut aborder (**address** [...] *pollution*, 5 occurrences). Mais on trouve aussi et surtout l'expression de plusieurs actions concrètes majeures :

Contrôle

La première action qui ressort quantitativement est la nécessité de contrôle, avec 67 occurrences (60 de *pollution control(s)* et 7 de **control** [...] *pollution*).

Réduction

Une autre grande action est tout ce qui peut opérer une décroissance de la pollution, avec le lexique **reduction/reduce**, *combat*, *abatement*, *mitigation* ; on relève 43 mentions de ce type.

Prévention

Enfin, on trouve 8 occurrences de *pollution prevention*.

Si la problématique de la pollution est présente dans tous les documents, il apparaît cependant que les solutions économiques et politiques restent limitées (la question de la pollution de l'air est symptomatique à cet égard). Dans ces textes, les actions sont plus des objectifs que des constats, des vœux que la description de réalités, comme on peut le vérifier à la lecture des contextes de ces constructions verbales liées à la pollution, et comme en témoigne aussi d'ailleurs la place dominante des modalisations par le conditionnel (*should*, mais aussi *could* et *would*), le futur (*will*), le potentiel (*can*), notée dans nos premières observations. Par ailleurs, en apportant dans ce cadre la composante risques, se poserait la question de quoi et comment mesurer, quels seuils risques-bénéfiques sont envisagés. Pour y répondre, l'analyse de corpus supplémentaires et plus spécifiques sur santé et sur pollution serait nécessaire.

4. Discussion et conclusions.

4.1. Bilan et perspectives socio-écologiques

À partir des textes fondateurs de la période 1970 et après cinquante années pendant lesquelles les institutions internationales ont cherché des réponses aux crises sociales, économiques et environnementales, notre corpus présente globalement les principaux points suivants:

1. La thématique des **ressources** ressort comme une première préoccupation économique et politique pour la période analysée.

⁵⁵ Requête CQL : @[enpos="VV.?"][enpos!="VV.?"|NNS?" & word!="{P}"]{0,3}[enlemma="pollution"]

2. L'analyse des contextes lexicaux montre que **ressources/énergie et pollution** sont étroitement liées, leur origine étant multiple (industrie, agriculture, transports) ; leur présence s'étend sur toute la période analysée ;
3. le lien entre **pollution et santé** s'affirme tout au long des discours. Les cycles extraction-production-consommation-déchets ont des conséquences sur la santé des personnes et des milieux. Ces conséquences sont d'autant plus marquées que les politiques et les éléments des discours institutionnels deviennent plus flous et permissifs (ex. les dérégulations) à partir des années 1990.
4. le lien entre **ressources et santé** est exprimé de façon moins directe, ce qui pourrait indiquer une moindre reconnaissance des ressources comme facteur de justice et santé sociale ; cependant on peut le reconnaître dans les liens répétés entre environnement et santé (impacts d'un environnement dégradé sur la santé).

Par ailleurs, avec une population de 8 milliards atteinte à la fin de 2022 et une pression sans précédent sur les ressources de la planète avec l'augmentation du tandem pollution-déchets, nous avons cherché à savoir comment la question de la démographie avait été abordée ces dernières décennies. Le corpus permet de vérifier une préoccupation constante des textes précoces pour la démographie, mais qui perd de sa dimension concrète dans les textes plus récents pour s'abriter derrière des questions moins sensibles politiquement ou qui focalisent l'attention médiatique (ODD, efficacité énergétique et plus généralement des ressources, climat, bien-être). C'est le paradoxe et le défi des temps présents (Meadows et al., 2005), sans oublier le défi de la sobriété qui doit y être associé (et que le corpus ne fait pas apparaître).

On peut considérer que la substance du corpus analysé repose en bonne partie sur un ensemble de cinq composantes – ressources/rareté, pollutions, santé, développement social et économique, et justice – qu'il faudrait davantage articuler politiquement. Mais divers obstacles s'y opposent, à commencer par le poids croissant de l'économie et de la finance dans les équilibres et évolutions politiques, mais aussi les limites du solutionnisme technologique et du dogme de la croissance infinie, et encore des discours trop souvent enfermés dans le développement durable, l'économie circulaire, le découplage ressources-PIB et la gouvernance systémique, voir **Annexe 1**). C'est le passage du discours à sa mise en œuvre qui ne fonctionne pas, car la volonté politique exprimée dans le corpus n'arrive pas à dépasser sa propre inertie institutionnelle et celle, plus forte, du système économique et financier qui en découle (Negrutiu et al., 2023).

Il faut attendre 2021 pour que les Nobel expriment clairement l'idée que le récit qui se dégage est problématique. Ils expriment des positions assez radicales, en faisant valoir qu'on ne peut plus faire l'impasse sur les relations de plus en plus marquées entre pollution globale et santé. Car c'est cette relation qui illustre le mieux les modes de développement, les styles de vie, et *in fine* des modèles de sociétés devenus problématiques.

Ils cherchent à identifier et à définir des priorités (voir aussi Hulme, 2020). Ainsi, la santé commune (*planetary health*) ressort comme une des idées-forces. En effet, depuis les années 1970, l'évolution des éléments de langage, de crime écologique (*ecocide*, 1972palme, p. 6) à la santé écologique et ensuite à la santé commune, indique d'une part une charge émotionnelle croissante de nos rapports à la Terre, passant des rapports guerriers à des rapports de paix (Serres, 2016), et d'autre part une paix pour laquelle la démarche santé-soin se décline à l'ensemble du vivant et de la planète ("maison commune" et siège de nos ressources vitales). L'extension de la notion de santé au social est aussi une manière de penser la santé économique. Ainsi, la santé finit par réunir un ensemble allant des individus au planétaire.

Nos résultats suggèrent que la santé commune pourrait être l'aboutissement du cheminement sémantique rapporté ici (Collart Dutilleul et al., 2023). Elle permettrait de renforcer un ensemble de potentialités législatives et politiques (en particulier des instruments pour la mise en valeur des communs et pour évaluer des coûts sociaux et écologiques réels) et possède des éléments d'acceptabilité sociétale plus parlants que les éléments de langage exprimés par Olof Palme et le processus de Stockholm dans les années 1970. Regardons de plus près ce concept-récit.

En effet, la définition de la santé commune intègre la santé de la civilisation humaine et l'état des systèmes naturels dont elle dépend (*Planetary Health* ; The Lancet-Rockefeller Foundation et Whitmee et al., 2015). Cette santé inclusive « nature-société-individus » rend les trois catégories de santé indivisibles, interdépendantes et réciproques. Il faut noter que les fondements juridiques des trois catégories de santé ont été élaborés depuis plus de 50 ans.

1. La santé des personnes (individuelle) selon l'OMS (1948) "n'est pas seulement l'absence de maladie ou d'infirmité, c'est un état de bien-être complet (physique, mental, social)".
2. La santé sociale signifie l'accès équitable pour tous aux ressources sociales, culturelles, naturelles et constitue un vecteur de cohésion sociétale. Le Conseil économique et social des Nations Unies (1994) précise que "les droits civiques, culturels, politiques, sociaux sont universels, interdépendants, indivisibles". Cette santé sociale doit favoriser l'accès à des services et ressources essentiels comme la santé, l'eau, l'alimentation ou encore l'éducation.
3. La santé des milieux naturels implique le maintien de l'organisation, des fonctions et de l'autonomie du système dans le temps (Rapport et al., 1998). L'Assemblée générale des Nations Unies défend depuis 1948 le droit à un environnement sain, elle établit une relation entre la qualité (santé) de l'environnement et les droits fondamentaux. La santé des milieux est donc un puissant déterminant de santé sociale et personnelle.

74 années plus tard, la dégradation des milieux naturels (donc des ressources essentielles) est reconnue comme une source de risque majeur pour les activités économiques et est objet de confrontations géopolitiques sur les ressources (WEF, 2020). Or la santé humaine, parce qu'elle est socialisée, implique toujours une forme d'organisation économique, une organisation soumise aux besoins et aux conditions de vie. L'analyse d'Eloi Laurent (2018) intègre ainsi l'économie et l'écologie par la problématique de la santé. Cette économie-là est un sous-système de l'écosphère, elle est intégrée dans les contraintes naturelles et reflète de manière adéquate les flux d'énergie/matière (Røpke, 2004 ; Rees, 2015).

Le Rapport Lancet / Oslo University (Ottersen et al., 2014) résume cela pour conclure que "Dans une société basée sur la justice sociale et la responsabilité écologique, la santé devient condition - résultat - indicateur de soutenabilité. Autrement dit, la santé doit être adoptée comme valeur universelle et objectif social et politique partagés par tous".

On voit l'intérêt sociétal, la puissance politique de la notion de santé, passerelle entre temps long et échéances courtes. En tant que valeur universelle, indissociable de l'accès équitable aux ressources vitales, elle reflète l'essence de la dignité humaine et des biens communs, mais aussi la capacité des sociétés pour fonder la paix sociale sur l'inséparabilité du bien-vivre des personnes et du soin accordé aux ressources dont elles dépendent (Ostrom, 2009 ; Bourgeron et al., 2018 ; Li et al., 2020). Cette santé-là est un passe-partout des liens sociaux et des liens naturels. Elle fournit la légitimité et l'acceptation des intérêts à long terme contre des avantages immédiats.

Ainsi, la santé commune s'imposerait comme le récit qu'appellent de leurs vœux les Nobel. C'est un récit du lien social et du temps long, qui éclaire l'ambivalence des rapports que les êtres humains ont à la santé et aux ressources, et qui met la responsabilité partagée au cœur des engagements politiques. Ces engagements alimentent une liberté sociale nécessaire pour mettre l'économie dans la société et l'écologie dans le social (Charbonnier, 2020).

Dans ce contexte, les besoins fondamentaux, et la santé qui en dépend, constituent l'aboutissement de la gestion responsable du patrimoine des ressources à tous les niveaux (Collard Dutilleul et al., 2023). La question des communs, mise en avant dans 1987brundtland et réaffirmée dans 202104nasem, mais en clair-obscur dans le reste du corpus⁵⁶, demeure l'horizon institutionnel à atteindre et pour lequel ressources - pollution / déchets – santé – justice constituent une matrice universelle. Cette matrice a été mise à l'épreuve des crises successives des cinq dernières

⁵⁶ On relève 35 *commons* (au pluriel, avec ou sans majuscule) dans notre corpus : 31 dans 1987brundtland et 4 dans 202104nasem.

décennies. La textométrie nous a aidés à examiner dans les discours la voie dans laquelle nos sociétés semblent pouvoir s'engager sans pour autant le savoir ou le dire encore clairement.

4.2. Bilan et ouvertures méthodologiques

Cette étude textométrique illustre comment, à partir d'un questionnement de recherche, la textométrie permet de construire de nouveaux observables pour nourrir la réflexion.

Le travail méthodologique et scientifique commence dès le choix des textes, car la composition du corpus détermine et pondère les observations subséquentes. Dans une première étape, nous avons arrêté une sélection de neuf documents relativement généraux ; une prochaine étape consisterait à étendre ce corpus vers des textes développant davantage les sujets centraux pour notre réflexion, autour de la santé globale et de la pollution.

L'outil logiciel (TXM) est compatible avec un large éventail de formats, et il est robuste vis-à-vis des imperfections des données (coquilles : mots inconnus, phrases incomplètes, etc.), il ne bloque pas et fournit des observations en l'état. Par conséquent, l'investissement dans la préparation des données numériques procède d'un équilibre entre d'une part le niveau de régularité et de précision attendu, et d'autre part le temps et l'expertise nécessaires pour la production plus ou moins automatisée et vérifiée d'une édition numérique adaptée à l'outil d'analyse. Ici, nous aurions pu faire moins : par exemple traiter les textes en bloc sans faire apparaître les divisions en chapitres, ne pas faire de corrections dans les textes issus de reconnaissance ou d'extraction automatique. Ou nous aurions pu faire plus : compléter la correction des mots mal reconnus, restituer les paragraphes, etc.

Comme l'atteste la diversité des sections détaillant la partie sur le parcours d'analyse et les résultats, les possibilités d'analyse sont multiples, et chacune offre de nombreux choix de paramétrage : il faut donc élaborer un parcours de recherche qui s'appuie à la fois sur l'expérience des façons d'aborder un terrain nouveau (une analyse factorielle d'ensemble, un examen du lexique quantitativement dominant, l'observation systématique de contextes, la vérification des interprétations dans les textes...), et sur des intuitions liées à la connaissance préalable des textes eux-mêmes et affinées voire rectifiées au fil des observations (le choix de mots servant de point d'entrée, l'élaboration de thématiques...). Il est difficile de conclure l'analyse, au sens où une observation en appelle une autre, et la mise en relation d'une moisson de résultats suggère à nouveau des pistes complémentaires. Ainsi, les observations que nous avons présentées rendent compte d'un point de vue sur le corpus, d'une "lecture" textométrique parmi d'autres, d'un moment dans un parcours de recherche, et n'épuisent absolument pas ce qui pourrait être dit ou mis en évidence à partir de ces textes, par nous-mêmes ou par d'autres.

Les observations apportent des éléments lexicaux et quantitatifs factuels pour étayer la réflexion : confirmation de dominantes (comme les cinq thématiques suivies ici) mais aussi de non-dits (avec l'exemple de la sobriété), dynamiques temporelles (l'usure du développement et l'essor de la question climatique), postures de différents émetteurs (par exemple, dans les rapports UNEP, l'adoption d'un style objectivant avec effacement de l'interlocution, et l'évitement de termes négatifs), contextes d'emploi précisant les façons d'aborder un sujet à travers l'organisation textuelle du propos et le choix des formulations (par exemple l'articulation confirmée, textuellement et lexicalement marquée, entre santé et pollution).

Il reste que l'analyse textométrique produit des résultats, et non des réponses ; elle propose des faits, mais toujours situés et en partie relatifs aux choix de la recherche, et jamais des démonstrations. Aussi la réflexion scientifique dépasse ce que le recours à l'outil pourrait avoir de technique. Elle s'élabore sur un terrain plus large, et en apportant une interprétation personnelle, experte et critique, sur la base de données partagées, elle s'ouvre à la discussion scientifique et à la mise au point collective de propositions nouvelles.

L'approche textométrique est accessible par divers logiciels, dont certains développés au sein de la communauté scientifique et plusieurs en *open-source*, comme le logiciel TXM utilisé ici. Dans le

contexte des humanités numériques, elle peut très concrètement être expérimentée par chacun, sur ses propres données et selon ses problématiques de recherche personnelles.

Remerciements

Nous sommes reconnaissants à Pablo Jensen pour avoir imaginé notre improbable rencontre. Nous remercions vivement les ingénieurs du pôle Humanités numérique de l'IHRIM pour leur soutien lors de la préparation du corpus numérique : Alain Collignon pour son travail minutieux de comparaison et d'évaluation des convertisseurs de PDF à TXT, Serge Heiden pour ses conseils experts dans la transformation de sources numériques pour TXM ainsi que pour l'implémentation d'utilitaires TXM intégrant l'outil PDFBox, et Alexei Laventiev pour son aide dans la mise au point des feuilles de style XSL pour améliorer la qualité du corpus produit à l'import (richesse, maniabilité). Nous remercions par ailleurs Pierre Beccaria pour ses retours sur une version préliminaire de ce texte et son attention particulière à une présentation claire des résultats issus des calculs textométriques.

Références

Bibliographie scientifique

BIERMANN Frank, HICKMANN Thomas, SENIT Carole-Anne, et al., 2022, « Scientific evidence on the political impact of the Sustainable Developmental Goals », *Nature Sustainability*, n° 5, p. 795-800.

BINDER Claudia R., HINKEL Jochen, BOTS Pieter W.G., PAHL-WOSTL Claudia, 2013, « Comparison of Frameworks for Analyzing Social-Ecological Systems », *Ecology and Society*, n° 18, p. 26-44.

BOURGERON Patrick, KLISKEY Andrew, ALESSA Lilian, LOESCHER Henry, KRAUZE Kinga, VIRAPONGSE Arika, GRIFFITH David L, 2018, « Understanding large-scale, complex, human–environmental processes: a framework for social–ecological observatories », *Frontiers in Ecology and the Environment*, n° 16, S1, p. S52-S66.

BRAND Ulrich, MURACA Barbara, PINEAULT Eric, SAHAKIAN Marlyne, SCHAFFARTZIK Anke, NOVY Andrea et al., 2021, « From planetary to societal boundaries: an argument for collectively defined self-limitation », *Sustainability: Science, Practice and Policy*, vol. 17, n° 1, p. 265-292.

CHARBONNIER Pierre, 2020, *Abondance et liberté, une histoire environnementale des idées politiques*, Paris, La Découverte.

COLLART DUTILLEUL François, 2011, « Les enjeux alimentaires du commerce des ressources naturelles », dans I. Negrutiu et al. éd., *Les ressources*, St Etienne, Presse Universitaires de St Etienne, p. 63-80.

COLLART DUTILLEUL François, 2017, *La Charte de La Havane - Pour une autre mondialisation*, Paris, Dalloz, coll. Tiré à part.

COLLART DUTILLEUL François, HAMANT Olivier, NEGRUTIU Ioan, RIEM Fabrice, 2023, *Manifeste pour une santé commune*, Paris, Utopia.

HEIDEN, Serge, MAGUE Jean-Philippe, PINCEMIN Bénédicte, 2010, « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », dans S. Bolasco, I. Chiari, L. Giuliano éd., *Statistical Analysis of Textual Data. Proc. of JADT 2010*, Rome, Edizioni Universitarie di Lettere Economia Diritto, p. 1021-1032.

HULME Mike, 2020, « One planet, many futures, no destination », *One Earth*, n° 2, p. 309-311.

- LAURENT Eloi, 2018, *Et si la santé guidait le monde ? Une société socio-écologique*, Paris, Les Liens qui Libèrent.
- LEBART Ludovic, PINCEMIN Bénédicte, POUDAT Céline, 2019, *Analyse des données textuelles*, Québec, Presses de l'université du Québec.
- LEBART Ludovic, SALEM André, 1994, *Statistique textuelle*, Paris, Dunod.
- LI Ting, DONG Yuxiang, LIU Zhenhuan, 2020, « A review of social-ecological system resilience: Mechanism, assessment and management », *The Science of the Total Environment*, n° 723: 138113, DOI:10.1016/j.scitotenv.2020.138113.
- MEADOWS Donella, RANDERS Jorgen, MEADOWS Denis L, 2005, *A synopsis: The Limits to Growth: The 30-Year Update*, Londres, Earthscan Editions.
- NARAIN Divya, BULL Joseph W, ALIKHANOVA Shahzoda, EVANS Megan C, MARKHAM Robert, MARON Martine, 2022, « A step change needed to secure a nature-positive future – Is it in reach? », *One Earth*, n° 5, p. 589-592.
- NEE Émilie (dir.), 2017, *Méthodes et outils informatiques pour l'analyse des discours*, Presses universitaires de Rennes.
- NEGRUTIU Ioan, 2022, « A Compass for Resource Justice and Planetary Health: Food Systems and Global Pollution », *Resources, Conservation & Recycling*, n° 181, 106229, <https://doi.org/10.1016/j.resconrec.2022.106229>.
- NEGRUTIU Ioan, ESCHER Gérard, WHITTINGTON Jason, OTTERSEN Ole Petter, GILLET Philippe, STENSETH Nils Christian, 2023, « The time boundary 2025-2030: the global resources and planetary health toolbox », *Proceedings of the Romanian Academy series B*, vol. 25, n° 2, p. 117-135. <https://acad.ro/sectii2002/proceedingsChemistry/doc2023-2/Art.4.pdf>.
- OSTROM Elinor, 2009, « A general framework for analyzing sustainability of social-ecological systems », *Science*, n° 325, p. 419-422, <http://dx.doi.org/10.1126/science.1172133>.
- OTTERSEN Ole Petter, DASGUPTA Jashodhara, BLOUIN Chantal, BUSS Paulo, CHONGSUVIVATWONG Virasakdi, FRENK Julio et al., 2014, « The Lancet–University of Oslo Commission on Global Governance for Health – The political origins of health inequity: prospects for change », *The Lancet*, n° 383, p. 630-667.
- PINCEMIN Bénédicte, 2020, « La textométrie en question », dans V. Magri, dir., *Linguistique et traitements quantitatifs*, *Le français moderne*, vol. 88, n°1, p. 26-43.
- PINCEMIN Bénédicte, 2022, « Sémantique textométrique », dans A. Biglari et D. Ducard, dir., *La sémantique au pluriel. Théories et méthodes*, collection Rivages linguistiques, Presses universitaires de Rennes, p. 373-396.
- POUDAT Céline, LANDRAGIN Frédéric, 2017, *Explorer un corpus textuel*, Louvain-la-Neuve: De Boeck.
- RAPPORT David J, CONSTANZA Robert, MCMICHAEL Anthony J, 1998, « Assessing ecosystem health », *Trends in Ecology & Evolution*, n° 13, p. 397-402.
- REES, William E, 2015, « Economics versus the economy », <http://www.greattransition.org/publication/economics-vs-the-economy> (consulté le 24 octobre 2023).
- RØPKE Inge, 2004, « The early history of modern ecological economics », *Ecological Economics*, n° 50, p. 293-314.
- SCHMID Helmut, 1994, « Probabilistic Part-of-Speech Tagging Using Decision Trees », dans *International Conference on New Methods in Language Processing*, p. 44–49.

SERRES Michel, 2016, *Darwin, Bonaparte et le samaritain. Une philosophie de l'histoire*, Paris, Le Pommier.

STEFFEN Will, RICHARDSON Katherine, ROCKSTRÖM Johan, CORNELL Sarah E., et al., 2015, « Planetary boundaries: Guiding human development on a changing planet », *Science*, vol. 347, n° 6223, DOI: 10.1126/science.1259855.

UN Economic and Social Council, 1994, « Draft Principles On Human Rights And The Environment. p. 1-5. Report No.: E/CN.4/Sub.2/1994/9, Annex I. UN International pact on economic, social, and cultural rights », <https://www1.umn.edu/humanrts/instree/1994-dec.htm>. (consulté le 24 octobre 2023).

WEF World Economic Forum, 2020, « World Economic Forum, Nature Risk Rising: Why the Crisis Engulfing Nature Matters for Business and the Economy », http://www3.weforum.org/docs/WEF_New_Nature_Economy_Report_2020.pdf (consulté le 24 octobre 2023).

WHITMEE Sarah, HAINES Anthony, BEYRER Chris, BOLTZ Frederick, CAPON Anthony G et al., 2015, « Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation-Lancet Commission on planetary health », *The Lancet*, n° 386, p. 1973-2028.

WHO World Health Organization, 1946, « Constitution, as adopted by the International Health Conference », New York, 19-22 June, 1946; entered into force on 7 April 1948. <http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf?ua=1> (consulté le 24 octobre 2023).

Textes du corpus

Cet inventaire est organisé suivant l'ordre chronologique.

Les fichiers utilisés pour le corpus ont été téléchargés entre le 12 et le 13 septembre 2022. Les liens ont été vérifiés le 26 octobre 2023 et fonctionnent encore.

Olof Palme statement, 1972 : *Statement by Prime Minister Olof Palme in the Plenary Meeting, June, 6, 1972*. Swedish Delegation to the UN Conference on the Human Environment. Available from:

http://www.olofpalme.org/wp-content/dokument/720606a_fn_miljo.pdf

Stockholm declaration, 1972 : *Stockholm Declaration*, United Nations. Available from:

<https://wedocs.unep.org/bitstream/handle/20.500.11822/29567/ELGP1StockD.pdf>

Brundtland report, 1987 : *Our Common Future. Report of the World Commission on Environment and Development*, United Nations. Available from:

<https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf>

Rio, 1992 : *Report of the United Nations Conference on Environment and Development, Rio de Janeiro, 3-14 June 1992. Volume 1, Resolutions adopted by the Conference*. United Nations. Available from:

<https://documents-dds-ny.un.org/doc/UNDOC/GEN/N92/836/55/PDF/N9283655.pdf?OpenElement>

UNEP International Resource Panel, 2015 : *Policy Coherence of the Sustainable Development Goals. A natural resource perspective*. Available from:

<https://wedocs.unep.org/handle/20.500.11822/9720>

UNEP International Resource Panel, 2017 : *Assessing Global Resource Use. A systems approach to resource efficiency and pollution reduction*. Available from:

<https://www.resourcepanel.org/reports/assessing-global-resource-use>

UNEP International Resource Panel, 2019 : *Global Resources Outlook 2019: Natural Resources for the Future We Want*, B. Oberle, S. Bringezu, S. Hat eld-Dodds et al. Available from: [https:// www.resourcepanel.org/reports/global-resources-outlook](https://www.resourcepanel.org/reports/global-resources-outlook)

NASEM, mars 2021 : *Progress, Challenges, and Opportunities for Sustainability Science: Proceedings of a Workshop in Brief*, National Academies of Sciences, Engineering, and Medicine 2021, Washington, DC, The National Academies Press. Available from: <https://doi.org/10.17226/26104>

NASEM, avril 2021 : *Our Planet, Our Future. An Urgent Call for Action. Nobel Prize Laureates and Other Experts Issue Urgent Call for Action After ‘Our Planet, Our Future’ Summit. Statement*, National Academies of Sciences, Engineering, and Medicine, April 29, 2021. Available from: <https://www.nationalacademies.org/news/2021/04/nobel-prize-laureates-and-other-experts-issue-urgent-call-for-action-after-our-planet-our-future-summit>

Annexes

Annexe 1. Tableau de synthèse du corpus.

The synthesis of 50 years of institutional agendas concentrates on progress, or rather progression, in achieving, or rather building the discourse on, environmental and socio-economic goals. These include the 1972 Stockholm declaration process, the ensuing Brundtland report 1987, the Rio conference in 1992, the UNEP thinking in the years 2015-2019, and the Nobel laureates questioning sustainability science in 2020-2021.

“We have never had to deal with problems of the scale facing today’s globally interconnected society. No one knows for sure what will work, so it is important to build a system that can evolve and adapt rapidly.” Elinor Ostrom (Nobel Laureate 2009).

Tableau A1.1. Tableau de synthèse des 9 textes du corpus.

Référence	Idées-cadre	Observations
<p>1. Palme 1972 (a premonitory plaidoyer on the Stockholm declaration) The principle of shared responsibility, fairness, and solidarity on the environment and resources. The balance between national independence and international interdependence concerning the commons and the common future.</p>	<ul style="list-style-type: none"> - future of affluence and abundance, need of growth - poverty and starvation - future of scarcity - resource inequality, population, poverty, and pollution/waste nexus - conflicts of interest with social change and direction of change, the issue of aims and values. 	<p>The term “ecocide” is used and focus on the balance pollution and energy supply.</p> <p>Action plan to assess the state of the global environment.</p> <p>The need for intergovernmental control of pollution and natural resource perspective, including the future of energy resources.</p> <p>The need for an international “machinery” and fund on the environment.</p>

Référence	Idées-cadre	Observations
<p>2. Stockholm declaration 1972 Environmental law guidelines and principles for the preservation of the human environment and enjoyment of basic human rights and needs for decent human existence. The responsibility of citizens, communities, enterprises and institutions.</p>	<p>26 principles. 1. Resources and ecosystems; development planning 2. Pollution, toxic substances, hazards to human health, liability and compensation 3. Economic and social development: quality of life 4. Demographic policies without prejudice to basic human rights 5. Science and technology on environmental risks 6. Education and mass media in the responsible conduct by individuals, enterprises, and communities 7. States and organizations, international law and standards.</p>	<p>Enacted by U.N. General Assembly Resolutions 2994/XXVII, 2995/UVII and 2996/XXII of 15 December 1972). Growing evidence of man-made harm in many regions of the earth: dangerous levels of pollution in water, air, earth and living beings; major and undesirable disturbances to the ecological balance of the biosphere; destruction and depletion of irreplaceable resources; and gross deficiencies, harmful to the physical, mental and social health of man,</p>
<p>3. Brundtland report 1987 From one earth to one world (socioecological thinking). Interrelationships between people, resources, environment, and development; Producing more with less. Legal principles for common action.</p>	<p>1. Common concerns on environment and development. 2. Sustainable development as equity and common interest (social dimension). 3. International economy, the environment and development, global risks. 4. Common challenges, population and human resources, food security. 5. Resources for development: species and ecosystems (biodiversity), economic value. - action at national and international levels 6. Energy for environment and development 7. Global industrial growth and impacts. 8. The urban challenge.</p>	<p>Managing the commons - oceans, space, arctic zones, - peace, security, development - principles of institutional and legal change.</p>

Référence	Idées-cadre	Observations
<p>4. Rio conference 1992</p> <p>'Earth Summit' to produce a broad agenda and a new blueprint for international action on environmental and development issues. Differentiated responsibilities to conserve and restore the health and integrity of the Earth's ecosystem. Growth, economic and trade.</p>	<p><i>27 principles and Agenda 21</i></p> <p><i>The principles of the Stockholm declaration are recalled and enforced:</i></p> <p><i>Principle 3.</i> The right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations.</p> <p><i>Principle 4.</i> Environmental protection shall constitute an integral part of the development process</p> <p><i>Principles 12, 15 and 16.</i> Trade policy measures for environmental purposes or the polluter-payer issue versus international trade and investment. The precautionary approach.</p> <p><i>Principle 23.</i> The environment and natural resources of indigenous people.</p> <p><i>Agenda 21</i> details the actionable aspects imbedded in the principles, with focus on planning and management.</p>	<p>Global partnership built on the premises of General Assembly resolution 44/228 of 22 December 1989; in agreement with the Report of the United Nations Conference on the Human Environment, Stockholm, 1972.</p> <p>The difficult, subtle, and sometimes hidden balance between public and private interests and means enters the diplomatic spheres and agendas.</p>
<p>5. UNEP 2015</p> <p>Considering SDGs as the unfinished MDG agenda, UNEP advocates increased efficiencies of resource use in production systems along with reduction of wastes and unnecessary consumptive patterns. Examines the resource interlinkages and potential trade-offs in the SDGs. Necessary transformations across all nations in production systems and in the consumption of services derived from natural resources.</p>	<p><i>The nexus between Natural Resource System and Socio-Economic System implies sustainable resource management and fair allocation (via institutions).</i></p> <ol style="list-style-type: none"> 1. Eradication of poverty identified as the greatest global challenge. 2. Ensure equity and dignity for all within Earth's safe and restored operating space. 3. Resource requirements and ecosystem health are linked, directly or indirectly. 4. Trends in population growth, lifestyle changes and economic activity are powerful drivers behind escalating resource use and related environmental degradation. 5. Policies need to anticipate and resolve possible resource conflicts across SDGs. 6. Systemic approaches, strategic planning to resources and ensuing tradeoffs must substitute to sectoral policies. 7. Synergies rather than trade-offs. Synergies as strategy to reduce pressures on the Natural Resources System. 8. Greatly increase information and knowledge about the current state of resources. 	<p>Resource constraints have always generated questions of equitable access to raw materials and distribution of derivative economic benefits at local, national, and global levels. In the current situation of emerging global resource scarcity, significant improvements to both the use efficiency and distributional equity of natural resources are essential.</p> <p>Solution pathways need to be beneficial and effective both on local and global levels, avoid geographic problem shifting, and make stakeholder interests converge.</p>

Référence	Idées-cadre	Observations
<p>6. UNEP 2017 Reducing overconsumption and waste of natural resources, and providing secure access to natural resources and food must be addressed simultaneously. The UNEP mantra of sustainability: better and more efficient production and use of natural resources are the most cost-efficient and effective ways to reduce impacts on the environment and advance human well-being.</p>	<p><i>Achieving environmental and socio-economic goals.</i></p> <ol style="list-style-type: none"> 1. Growing material use is driven by expanding populations, consumption trends in mainly developed economies and the transformation of developing economies. 2. Essential infrastructure (energy, buildings, transportation, water supply, sanitation and waste management) and food supply sectors significantly contribute to global resource-use pollution and environment-related impacts on human health, while shaping social equity in basic provisioning. 3. Systems approaches, decoupling economy, targets and indicators, reform of financial instruments, policy learning, capacity building and knowledge sharing are instrumental in restoring and maintaining the health of the natural resource base while offering high-quality services as an alternative to material consumption. 	<p>The links between human well-being, economic prosperity and environmental resilience, all these three dimensions to be considered in policymaking. Global assessment of resource use, with focus on delivering services rather than material products can inform the setting of long-term orientation goals, and corrects market failures.</p>
<p>7. UNEP 2019 Over the past five decades, global population has doubled, the extraction of materials has tripled and gross domestic product has quadrupled, accounting for more than 90 per cent of our biodiversity loss and water stress, and approximately half of our climate change impacts. Emphasis on primary resources, namely land / cropland, water, and biomass.</p>	<p><i>Targeting a sustainable scenario.</i></p> <ol style="list-style-type: none"> 1. Historical and current patterns of natural resource use have increasingly negative impacts on the environment and human health. 2. Realizing the Sustainable Development Goals – while staying within the planetary boundaries requires an urgent and systemic transformation of how we use and manage natural resources. 3. Policymakers and decision makers have tools at their disposal to analyze the demographic and socioeconomic forces driving the extraction and use of natural resources globally, achieving decoupling, also supporting economic growth and human well-being. <p>In brief,</p> <ul style="list-style-type: none"> - indicators and targets - national plans - policy mixes and circular economy - sustainable financing - education and changing behaviour - international cooperation. 	<p>Natural resources and environmental impacts have become the limiting factor of production and shifts from improvements in labour productivity are required to focus on resource productivity. Global trade in materials translated in displacement of all types of environmental and health impacts from the consuming high-income countries to the middle- and low income countries.</p>

Référence	Idées-cadre	Observations
<p>8. Nasem 2020 (published in March 2021) <i>sustainability science</i> progress and gaps (concepts, methodologies, and scales). Can systemic change arise from local initiatives? Who has agency to decide what is a “good transformation”? The need for a compelling vision to provide direction, forming networks, and weakening dominant structural forces.</p>	<p><i>On the reality of power structures, time – space scales, and the fact that social-environmental science remains weak in construction and testing.</i></p> <p>1. <i>Measurement, the space for informed conversation:</i> still far from a practical measure of sustainable development, namely resource deployment and human well-being (equity and justice). However, multi-indicator dashboards and remotely sensed information are currently available for decision makers helping develop politics of early warning signals, transparency, and accountability).</p> <p>2. <i>Justice-focused equity.</i> Breaking down the economic, social and political drivers of lock-in.</p> <p>3. <i>Adaptation governance and Transformative change, linkages between knowledge and action</i> are processes with phases at all scales with barriers arising from power dynamics and vested interests, and boundaries of science relative to politics.</p>	<p>The reality: 75 percent of the global population live in poverty, with 10 percent in extreme poverty. Ability to make and remake relationships across both biophysical and social processes may have immediate practical implications. SDGs serve more as aspiration than practical level guidance. Confronted with unease in discussing equity in the context of finite resources and fast versus slow tradeoffs building beyond false choices, implies multiple players and assets and tackling distributional issues across scales.</p>

Référence	Idées-cadre	Observations
<p>9. NASEM statement 2021</p> <p>Facing new challenges at unprecedented scale.</p> <p>Long term, scientifically based decision-making.</p>	<p>Global crisis: vast and uncontrolled experiment on Earth's life-support system.</p> <p>1. <i>The global commons and land use</i> raise issues concerning all the living and non-living systems on Earth that societies use but that also regulate the state of the planet. <i>Planetary Health</i>. the intimate connections between human health and the health of other animals and the environment becomes the actionable narrative of the commons.</p> <p>2. <i>The social capital</i> - addressing destabilizing inequalities in the world. Reducing inequality raises social capital and positively impacts collective and difficult long-term decision-making.</p> <p>3. <i>Guiding the technological revolution</i>. On aggregate, technological advancements so far have accelerated us down the path toward destabilizing the planet. Need to support social goals and the importance of infrastructure projects.</p> <p>4. <i>Time and urgency - 2020-2030</i>, a decisive decade. The transition process, including education and media, will we do it fast enough?</p> <p>5. <i>What to measure?</i> Beyond GDP, measures of the true well-being of people and nature (i.e., planetary health, the resilience of societies and the resilience of Earth's biosphere).</p> <p>6. <i>Finance and Business</i>. Investors and companies must adopt principles of recirculation and regeneration of materials and apply science-based targets for all global commons and essential ecosystem services. Economic, environmental, and social externalities should be fairly priced.</p>	<p>Without transformational action this decade (2021-2030), humanity is taking colossal risks with our common future.</p> <p>A disconnect between the urgency indicated by the empirical evidence and the response from electoral politics.</p>

Annexe 2. Vocabulaire caractéristique des neuf textes du corpus (Spécificités)

Le calcul de spécificités porte ici sur le corpus divisé selon ses neuf textes : il met en évidence les mots particulièrement présents dans chaque texte, au regard de leur fréquence générale d'emploi à l'intérieur du corpus entier. C'est une approche non seulement relative au corpus, mais aussi statistique : il ne s'agit pas directement d'une présence quantitative, mais bien d'une évaluation du caractère remarquable de la fréquence du mot, non attribuable au hasard, semblant donc marquer un choix lexical.

Tableau A2.1. Mots (lemmes) les plus spécifiques de 1972palme (12 max. par catégorie et indice de spécificité supérieur ou égal à 2)⁵⁷

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
conference fact solidarity future environment declaration program document delegation draft laissez-faire prophecy	Swedish international common simple preparatory decisive immense general careful <i>anonymous</i> ⁵⁸ <i>terrifying</i> <i>unparalleled</i>	fear know center attack share act hope work propose wish lay demand	must will want	we our this my it I they us	all not yes nowadays always rightly wrongly

Dans 1972palme, la parole est impliquée et personnelle (première personne), le ton est ferme, volontaire et interpellant (cf. modaux, interjections). Il s'agit à partir d'un ressenti et d'un point de vue (*fear, know*), de se mobiliser collectivement (*solidarity, common, share*) pour notre avenir (*future, hope, wish*).

N.B. : *ecocide* n'apparaît qu'une seule fois dans le corpus, et c'est dans le texte 1972stockholm (indice de +2,4 mais c'est au-delà du douzième nom commun spécifique).

⁵⁷ Par souci de transparence et d'accès aux données fondant l'interprétation, nous donnons ici les résultats bruts du calcul, au sens où nous donnons tous les mots correspondant aux critères de sélection chiffrés, sans filtrer ceux qu'on pourrait juger moins intéressants, par exemple le caractère évident de *Swedish* dans le contexte de 1972palme.

⁵⁸ En italiques, les mots dont la seule et unique occurrence est dans le texte considéré. Il y a un lot de mots ainsi ex-aequo, pour compléter la liste à 12 nous avons choisi ceux qui nous paraissent les plus parlants.

Tableau A2.2. Mots (lemmes) les plus spécifiques de 1972stockholm (12 max. par catégorie et indice de spécificité supérieur ou égal à 2)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
principle environment man life jurisdiction earth posterity state mankind co-operation domination harm	human common environmental man-made essential social adverse irreversible future present adequate <i>irreplaceable</i>	co-ordinate safeguard improve bear protect transform reach <i>propel</i> <i>entrust</i> <i>remedy</i> <i>imperilled</i> <i>condemn</i>	must shall	his their him	all wisely continuously heedlessly gravely wrongly

1972stockholm se présente comme un texte au ton plus solennel et formel (*man, shall, principle*), posant sans détour (*must*) les notions fondamentales de la préoccupation socio-écologique (*environment* et *environmental, earth, co-operation* et *domination, posterity, future* et *irreversible, safeguard, protect, improve* et *transform*).

Tableau A2.3. Mots (lemmes) les plus spécifiques de 1987brundtland (12 max. par catégorie et indice de spécificité supérieur ou égal à 3)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
nation citation problem weapon NGO aid arm specie world space crisis number	nuclear industrial military political tropical new ecological industrialized major poor foreign next	dump live grow become threaten believe come suffer go see spend rise	must can would could will may want might	they it we our now I today these them you itself us	many not most few much more no too even so little only

Ce qui frappe pour le rapport Brundtland, c'est le ton impliqué, l'abondance des marques évaluatives et des modaux. Le discours rentre dans le concret, dans les détails, et, à l'échelle du monde, il semble particulièrement préoccupé par la question des conflits et rapports armés.

Tableau A2.4. Mots (lemmes) les plus spécifiques de 1992rio (12 max. par catégorie et indice de spécificité supérieur ou égal à 3)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
development organization information programme implementation management training community activity government cooperation mechanism	international national regional relevant appropriate sound non-governmental indigenous local marine indicative concessional	strengthen include promote develop implement review establish decide undertake take facilitate ensure	should (shall 2,2)	their those	environmentally particularly all fully personally well particular socially appropriately duly mutually

Le discours de 1992rio frappe par son caractère formellement structuré et systématique (comme on le voit avec la déclinaison méthodique des échelles : *international, national, regional, local*) et le lexique très général de l'action politique (*information, management, training, promote, strengthen, decide, ensure...*).

Tableau A2.5. Mots (lemmes) les plus spécifiques de 2015unep (12 max. par catégorie et indice de spécificité supérieur ou égal à 2)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
goal resource trade-off SDGs outcome strategy synergy industrialisation policy nexus target consumption	natural sustainable resilient socio-economic current social economic inclusive radial null radical positive	achieve target need avoid lead design recognise fail magnify address end decouple	will		simultaneously directly intrinsically significantly indirectly away too

Le premier rapport UNEP donne une place-clé aux objectifs du développement durable, sous l'angle de l'activité économique (*resource, industrialisation, consumption*). Il adopte un ton générique et objectivant qui semble effacer les marques de positionnement personnel et d'évaluation.

Tableau A2.6. Mots (lemmes) les plus spécifiques de 2017unep (12 max. par catégorie et indice de spécificité supérieur ou égal à 3)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
city material infrastructure footprint air emission tonne efficiency pollution resource consumption economy	resource-efficient multiple premature circular final inclusive natural urban global strategic fossil ambitious	reduce provision steer illustrate see use achieve approach show demonstrate extract track			around globally mainly lastly disproportionally co-beneficially fast annually

Le 2e rapport UNEP rentre dans le détail des impacts du modèle socio-économique industriel et de la civilisation urbaine. La question de la pollution est centrale. Le ton reste général et impersonnel, technique et factuel, objectivant.

Tableau A2.7. Mots (lemmes) les plus spécifiques de 2019unep (12 max. par catégorie et indice de spécificité supérieur ou égal à 3)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
extraction figure impact income cent capita decoupling material scenario climate ton trend	historical global net domestic upper-middle relative natural high resource-related particulate absolute low	model show use impact decrease drive project increase compare extract leapfrog lead	can	this	mainly additionally approximately finally below extensively onwards globally still typically around broadly

Le 3e rapport UNEP poursuit le discours factuel et objectivant (effacement des marques de personne), mobilisant de plus en plus nombreuses figures et autres diagrammes et éléments chiffrés. À la question des ressources s'articule celle des scénarii climatiques. Les nuances portent sur les chiffres, les ordres de grandeurs, c'est une discussion technique autour des modèles et projections.

Tableau A2.8. Mots (lemmes) les plus spécifiques de 202103nasem (12 max. par catégorie et indice de spécificité supérieur ou égal à 2)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
sustainability equity workshop science governance transformation theory justice knowledge measurement power resilience	adaptive transformative intergenerational social-ecological white research-based procedural normative social inclusive distributional transformational	say note understand explain suggest comment state conclude intertwine explore discuss happen	can might	she he we her his our	not rather just

La NASEM veut envisager le sujet au plus large, dans toutes ses dimensions, y compris de justice sociale. Les chercheurs s'expriment personnellement (cf. pronoms personnels et possessifs) et ce sont les termes du débat qui sont rapportés (cf. les verbes).

Tableau A2.9. Mots (lemmes) les plus spécifiques de 202104nasem (12 max. par catégorie et indice de spécificité supérieur ou égal à 2)

Noms communs	Adjectifs qualificatifs	Verbes (hors auxiliaires)	Modaux (y compris <i>to want</i>)	Pronoms personnels, Possessifs, Déictiques	Adverbes, Interjections, Intensité
laureate society ice pandemic inequality stewardship planet epoch humanity summit urgency misinformation	global collective planetary Holocene destabilizing irreversible societal long-term technological great transformational scientific	destabilize wake embed value embrace undermine come exacerbate <i>lighten</i> <i>cascade</i> <i>applaud</i> <i>celebrate</i>	must will	our we us now it	highly deeply tightly

Dans son document d'avril 2021, c'est la communauté scientifique qui s'exprime dans le document de la NASEM (première personne du pluriel). Sans détours (*urgency, must, will*), elle insiste sur la nécessité de changer d'échelle (*embrace, highly, deeply, tightly*), dans le temps (*epoch, Holocene, long-term, irreversible*) et dans l'espace (*planet* et *planetary, global*), et de prendre le parti du collectif, avec ses dimensions d'équité sociale et de solidarité (*society* et *societal, inequality, collective*).