

To cite this article: François, Marius (2025). Exploration des corpus oraux en terminologie par le prisme de la néologie « première » en anglais dans le domaine des événements climatiques extrêmes. In *Terminology Science & Research / Terminologie : Science et Recherche* 28, 57–84. Available at: <https://journal-eaft-aet.net/index.php/tsr/issue/archive>.

Exploration des corpus oraux en terminologie par le prisme de la néologie « première » en anglais dans le domaine des événements climatiques extrêmes

Marius François

CeRLA, Université Lumière Lyon 2, <https://orcid.org/0009-0007-8188-2023>

Résumé :

Cet article s'intéresse à l'intersection jusque-là peu considérée entre les corpus oraux et la terminologie via la définition et la considération d'une néologie « première ». Ce travail s'appuie sur une méthodologie fondée sur la méthode du « corpus-pivot » afin d'identifier des termes nouveaux dans le discours de scientifiques impliqués dans le projet interdisciplinaire XAIDA. Nous mobilisons à la fois une approche quantitative et qualitative afin d'analyser plus en profondeur les néologismes premiers et tirer des conclusions préliminaires sur l'utilisation des corpus oraux à des fins d'analyse terminologique.

Mots-clefs : corpus oraux, néologie terminologique, néologie première, exploration outillée

Abstract:

This article investigates the hitherto underexplored intersection between oral corpora and terminology via the definition and consideration of “primary neology” in oral contexts. This work relies on a methodology based on the “pivot-corpus” method in order to identify new terms in the discourse of scientific experts involved in the interdisciplinary XAIDA project. We employ both quantitative as qualitative approaches to analyze primary neologisms in greater depth and draw preliminary conclusions regarding the use of oral corpora for terminological analysis purposes.

Keywords: oral corpora, terminological neology, primary neology, tool-assisted exploration

1 Les corpus oraux : une source d'information peu exploitée en terminologie

En dépit des constats dressés par certains terminologues (notamment Desmet, 2007, p. 10), des impulsions (par exemple Candel, 1984, 1993 ; de Vecchi, 1999 ; Delavigne & de Vecchi, 2021 ; Delavigne et al., 2022), des prescriptions (Cabré, 2000, p. 14) ou des vœux formulés (notamment Gaudin, 1995, p. 232, 2005, p. 88,)), les corpus oraux constituent un matériau de recherche sous exploité, voire complètement marginalisé par la discipline :

Les progrès technologiques permis par la puissance de calcul des logiciels ne sont pas accompagnés de travaux, plus coûteux en moyens humains et en temps, sur l'oral, dimension cruciale pour comprendre la genèse et la diffusion des terminologies [...]. (Gaudin, 2005, p. 89)

Les raisons d'une telle marginalisation sont plurielles (difficultés méthodologiques posées par la collecte de données orales [voir Erlos, 2012, entre autres], accès limité aux ressources spécialisées [voir Diagne, 2022, entre autres]). Sur le plan théorique d'abord, la Théorie Générale de la Terminologie (TGT), largement inspirée des travaux d'Eugen Wüster (1979), postule une stabilité des concepts ainsi que le principe de biunivocité, en somme un terme pour un concept, un concept pour un terme (Humbley, 2004, p. 11). Postulons ici, en nous alignant sur ce que Seghezzi (2007, p. 5) a écrit, que cette vision figée du terme de spécialité « dépourvu de valeur communicative » (*ibid.*, notre traduction) inhibe la prise en compte de « toute caractéristique contextuelle » (*ibid.*, notre traduction), ce qui ne laisse donc guère de place pour la considération influences potentielles du canal de communication sur l'usage des termes. Qui dit usage indifférencié prédit sans doute une proximité parfaite entre expression écrite et expression orale.

Si l'on met sur le même plan écrit et oral, quel intérêt saurait représenter l'étude de ce dernier à une époque où l'on dispose de véritables « autoroutes de l'information » nous permettant un « accès facilité à d'énormes masses textuelles » (Van Campenhout & Temmerman, 2011, p. 224) ? D'autant plus que les ressources orales spécialisées, ou bien ne sont pas disponibles (Humbley, 2018, p. 222 ; Ciapuscio, 2007, pp. 95-96), ou bien requièrent un processus de retranscription fastidieux (Gaudin, 2005, p. 89, García Antuña, 2015, p. 51, Ylönen, 2018, p. 365) contenant de nombreuses étapes avant de pouvoir envisager une phase d'analyse (Lerat, 1995, Cabrera Castro, 2017, p. 40). Boulanger (2010) ira même jusqu'à qualifier la « dimension orale » d'« insaisissable », quand Tallarico (2022) indiquera que c'est « la partie orale de la néologie qui échappe le plus souvent au repérage ».

Le présent article propose de rendre compte des principes méthodologiques et des premiers résultats obtenus dans le cadre de la compilation et de l'exploration d'un corpus oral spécialisé à des fins d'analyse terminologique, par le prisme de la néologie spécialisée. Nous avons, pour ce faire, utilisé des données orales tirées d'interventions organisées au cours de la *Summer School* du projet européen XAIDA¹, un consortium interdisciplinaire mêlant sciences climatologiques et intelligence artificielle dont l'objectif est la détection et l'attribution d'événements climatologiques extrêmes au changement climatique. Ces données orales ont été comparées avec d'autres données, écrites celles-là, issues des références bibliographiques mentionnées par les intervenants ainsi que de leur production scientifique écrite post *Summer School*, afin de statuer sur le statut néologique de certains termes et de mesurer leur degré d'implantation. Le détail des données ici évoquées sera repris dans les paragraphes consacrés.

La première partie de l'article présentera certains des travaux évoqués en début d'introduction, en complément d'autres écrits sur des sujets proches. Nous présenterons ensuite nos objectifs et postulats de recherche, relatifs notamment à la définition d'une néologie dite première. Nous passerons ensuite en revue notre méthodologie en présentant, entre autres, nos corpus de travail et nos outils. Nous exposerons quelques résultats et tâcherons de décrypter certains des mécanismes à l'œuvre dans le contexte observé. L'article sera conclu par une synthèse des éléments à l'étude ainsi que par des pistes de réflexion à suivre.

1.1 Les travaux existants

Bien que la recherche terminologique n'ait jusque-là que peu considéré les corpus oraux comme matériel de recherche, il est néanmoins possible de trouver quelques exceptions. Nous avons défini notre champ d'analyse autour de différents supports de publication, tels que *Terminology*², *Neologica*³, *META*⁴, *Terminalia*⁵, entre autres.

Les travaux des équipes catalanes de l'*Observatori de Neologia* ont notamment taillé une place de choix à la néologie orale. Nous revenons, dans les quelques lignes qui suivent, sur certains d'entre eux.

¹ <https://xaida.eu/>

² <https://benjamins.com/catalog/term>

³ <https://classiques-garnier.com/neologica.html>

⁴ <https://www.erudit.org/fr/revues/meta/>

⁵ <https://revistes.iec.cat/index.php/Terminalia/index>

Estopà (2008) détaille la création et le fonctionnement de l'*Observatori de Neologia de l'IULA*. La chercheuse précise qu'au corpus écrit initialement considéré s'est ajouté, à la fin des années 1990, un corpus oral d'émissions radiophoniques dépouillé manuellement afin d'étudier les phénomènes néologiques oraux, en espagnol et en catalan. Les émissions considérées constituent un observatoire privilégié de la communication orale, que l'autrice qualifiera de « plus spontanée » (Estopà, 2009, p. 552) et dont elle se servira afin d'étudier le phénomène lexicogénique de composition en catalan. Les observations de la chercheuse tendent à rapprocher les phénomènes néologiques écrits et oraux, dans un contexte général, précisons-le, en ce sens que la part occupée par la composition parmi les néologismes relevés est comparable (de l'ordre d'environ 2,5 %) et que les structures morphosyntaxiques mobilisées sont également comparables (prévalence des structures NOM + NOM, VERBE + NOM, etc.). Dans le même contexte, Domènech & Estopà (2009) mettent l'accent sur la production néologique orale en catalan et dégagent les tendances les plus saillantes. Les néologismes formels constituent la majorité des néologismes relevés ($\approx 60\%$), devant les emprunts à l'espagnol et à l'anglais ($\approx 30\%$) et les cas de néologie sémantique ($\approx 8\%$) (*ibid.*, pp. 53), des tendances comparables à celles observées en contexte écrit. Les chercheuses notent également des cas de néologismes sémantiques plus fréquents en contexte spontané, dont la plupart demeurent toutefois hapaxiques, ou n'ont au maximum que deux occurrences ($\approx 95\%$) (*ibid.*, p. 59). Enfin, les autrices notent que la quasi-totalité des néologismes oraux correspondent à des cas de néologie expressive (*ibid.*, p. 62). Du reste, Freixa (2012) propose un état de l'art exhaustif des travaux en néologie orale menés depuis 1995 en contexte catalano-hispanophone. Nous noterons toutefois que ce premier tour d'horizon ne fait pas mention explicitement d'analyse de données spécialisées. Les prochains travaux recensés mobilisent, quant à eux, des corpus oraux techniques, sans pour autant que l'accent soit nécessairement mis, cette fois-ci, sur les phénomènes néologiques.

Marcellesi (1973) propose une étude de ce qu'elle nomme « la langue orale de l'informatique » au sein de laquelle elle mobilise, entre autres, un corpus oral d'entretiens et de conversations avec des experts du domaine (*ibid.*, p. 68). Ces données lui permettent de constater un écart conséquent entre écrit et oral spécialisé, ce dernier présentant une diversité beaucoup plus importante (*ibid.*, p. 64). Cette dernière s'exprime, par exemple, via une utilisation plus poussée d'abréviations (*macro* pour *macro-instruction*, *ibid.*, p. 70) de sigles (*BH*, *EOJ*, *MVC*, etc., *ibid.*, p. 69) de dérivés de termes anglais intégrés au système morphologique du français (*canceler*, *bufforiser*, etc.,

ibid., p. 70), de certains patrons de composition ou encore d'une prévalence plus importante d'une terminologie anglaise en comparaison avec l'écrit.

Lanqueur (2016), quant à lui, inclut, entre autres, une douzaine de retranscriptions d'émissions radiophoniques (en plus d'un corpus écrit) dans une démarche d'évaluation du passage des mots d'un registre courant à un registre spécialisé, dans le domaine de l'économie, et compare les phénomènes néologiques observés dans les deux contextes. L'observation majeure que nous retiendrons de ce travail est que le corpus écrit est moins riche en néologismes. L'auteur impute cela au contexte de production qui permet, selon lui, de « dispose[r] de plus de temps pour la réflexion, ce qui fait que le respect des règles de la langue est plus rigoureux ».

De Vecchi (1999), dans le cadre de la communication en entreprise, indique que l'écrit ne saurait être considéré comme « seul véhicule de connaissances » et que « l'oral a sa place et ne saurait être ignoré » (*ibid.*, p. 13), notamment car il se constitue réservoir d'un ensemble de mots que l'auteur qualifie de « tabou » (*ibid.*, p. 143) qui contribue au maintien du lien social entre les employés, fonctionne comme un marqueur d'identité, d'appartenance à un groupe (*ibid.*, p. 295), révèle des usages polysémiques spécialisés (*ibid.*, pp. 292-293), entre autres choses.

Rappelons également les articles de Candel, publiés respectivement en 1984 et en 1993 qui appréhendent, pour le premier, les marqueurs métalinguistiques aux moyens desquels des chercheurs en sciences physiques tendent de « mieux faire passer [leur] message, mieux le formuler ou le "reformuler" » et, pour le second, la variation spécialisée en contexte oral scientifique.

Condamines (2010) évoque une expérience qui s'appuie sur des documents écrits produits par des experts de deux départements du CNES (Centre National d'Études Spatiales) et retranscriptions de communications entre ces mêmes experts, dans une optique d'analyse des différentes acceptions du terme *satellite*.

Delavigne, Picton & Thibert (2022) considèrent des transcriptions d'interactions orales afin de saisir les caractéristiques des négociations effectuées entre interlocuteurs de niveaux d'expertise différents. Schneider & Estopà (2020) proposent, en guise d'ouverture de leurs travaux de détection et d'observation de la néologie spécialisée du

domaine médical, d'inclure l'oralité, car elles la considèrent comme un canal important de diffusion d'informations spécialisées dudit domaine.

Seghezzi (2013), elle, cherche à introduire les données orales spécialisées dans une optique plus large d'étude des phénomènes susceptibles d'être à l'origine d'une variation terminologique. L'autrice cite les travaux de Ciapuscio (1997), qui stipulent que l'oral constitue une plateforme privilégiée de la reformulation terminologique au fil de l'eau (notre traduction et notre adaptation). De ses observations, notons que du point de vue de la « masse » de variation terminologique, il n'existe vraisemblablement pas de différence notable entre l'écrit et l'oral spécialisés (*ibid.*, p. 73). Seghezzi évoque toutefois, au niveau formel, une présence accrue de variantes terminologiques orales moins lexicalisées, plus transparentes et qui, selon ses mots, pourraient mettre en exergue une volonté de se rapprocher du public (*ibid.*).

Garcia Antuña (2015), enfin, aborde les détails d'un projet terminographique du domaine de la maroquinerie en espagnol, travail basé en partie sur un corpus oral, un recours rendu nécessaire, du fait de la non-exhaustivité des sources écrites. Son recours aux données orales lui permet notamment d'obtenir quantité d'informations importantes, relatives aux composants, aux pièces ainsi qu'aux processus de fabrication, non mentionnées dans les données écrites.

De manière relativement similaire, l'analyse de la terminologie sensorielle de la dégustation de vin par Gautier & Bach (2017) enjoint au recours à un corpus oral spécialisé. En effet, la terminologie sensorielle considère à la fois la « dimension expérientielle (...) et la dimension sensible » (*ibid.*, p. 488). Les interactions orales constituent, selon Leroyer (2015), la majorité des interactions professionnelles dans le domaine de la dégustation de vin. Le corpus oral se constitue alors lieu privilégié pour « observer le fonctionnement des termes *in vivo* » (Gautier & Bach, 2017, p. 496) via, par exemple, l'observation de la négociation du sens entre plusieurs locuteurs (*ibid.*, p. 492, Leroyer, 2015, p. 258) comme de la variation (*ibid.*) ou de l'introduction d'hapax, d'idiosyncrasies qui intègrent parfois le discours. À cet effet, Gautier & Bach évoquent par exemple le terme allemand *Schnüffelwein*, observable en premier lieu uniquement en contexte oral et qui a ensuite pénétré le discours écrit des experts (2017, p. 492).

Les travaux susmentionnés ont été répartis en deux catégories : nous avons d'abord évoqué les travaux catalans qui s'intéressent aux phénomènes néologiques en langue

générale. Ensuite, nous avons fait état des études qui ont considéré l'oral spécialisé, dont certaines, minoritaires (Lanseau, 2016, Gautier & Bach, 2017, ouverture de Schneider & Estopà, 2020), mettent l'accent sur la néologie spécialisée orale. Notre étude vise à se démarquer en approfondissant les tendances observables relatives aux phénomènes néologiques spécialisés en contexte anglophone.

2 Objectifs et postulats de recherche

2.1 Une distance lexicale/terminologique notable entre l'oral et l'écrit ?

Notre postulat de recherche principal porte sur l'existence d'une différence entre l'oral et l'écrit. Pour l'illustrer, nous avons cherché à démontrer son existence via une Analyse Factorielle de Correspondances (AFC), une technique d'analyse statistique mise au point par Benzécri (1973). Elle permet une représentation graphique de la distance lexicale entre plusieurs sous-corpus. Elle « donne une visualisation des distances entre les mots et les sous parties d'un corpus » (Beaudouin, 2016, p. 27) que nous considérons ici comme point d'entrée dans nos données. Elle n'a pas pour autre vocation que de constituer une analyse introductive ayant pour objectif de confirmer ou d'infirmer l'existence d'une distance lexicale ou terminologique entre nos ensembles de communications orales et écrites.

2.2 Observer et décrire la création néologique lexicale en contexte oral spécialisé

Boulangier (2010) écrivait que « le néologisme naît généralement dans la parole ; il rejoint l'écrit plus ou moins rapidement selon les besoins ». Cette première citation pourrait être mise en miroir de ce que Humbley (2018) écrivait :

Idéalement, on observe plusieurs états de langue : avant l'apparition du néologisme, ses premières attestations et son évolution immédiatement après sa création. **L'acte même de création n'est que rarement documenté** [gras ajouté], d'où la nécessité de disposer de modèles pour se faire une idée de ce qui se passe.

Au vu des éléments précédemment évoqués concernant le manque de ressources orales spécialisées ainsi que le peu d'études terminologiques les mobilisant, nous souhaitons, à travers notre démarche, vérifier s'il est possible de trouver des exemples de néologie spécialisée morphologique née à l'oral. Quelles matrices terminogéniques sont à l'œuvre ? Quelles situations de communication incitent à la création terminologique ?

2.3 Définir les contours d'une néologie « première »

Pour ce faire, nous avons souhaité définir les contours d'une néologie dite première. Le choix de l'adjectif résulte de notre hypothèse que le terme scientifique est amené à apparaître dans un contexte oral, avec un nombre d'occurrences relativement faible : puisque sa création est récente, son utilisation peut demeurer sporadique un temps. Après une période de temps donnée plus ou moins longue, rappelons à cet effet que Cabré & Nazar (2012) soulignaient l'imprévisibilité de l'évolution de l'utilisation d'un terme, le néologisme spécialisé est amené, ou non, à transitionner vers un contexte écrit.

Un néologisme premier possède donc une première occurrence datable à l'oral et est utilisé pour exprimer un concept spécialisé, dans un objectif heuristique ou de transposition. Ainsi, notre définition de néologie première incorpore-t-elle deux types de néologismes spécialisés. Les premiers sont les néonymes d'origine (Rondeau, 1981) ou *primary term formation* (Sager, 1989), en somme la création heuristique au cours de laquelle une nominalisation est choisie pour mettre en discours un concept scientifique nouveau (Condamines et al., 2012). Les seconds correspondent à ce que Humbley (2006, 2018, p. 253) définit comme néologismes d'interface et/ou de transposition : un expert d'un domaine donné cherche à rendre son discours compréhensible pour un expert d'un autre domaine ou pour quiconque possédant un niveau de connaissances supposé moins élevé que l'émetteur du discours dans le même domaine.

3 Éléments méthodologiques

3.1 Méthode de l' « événement-pivot »

L'approche méthodologique que nous avons mise en place s'inspire des travaux de Paquin & Drouin (2010) et Lafrance, Drouin & L'Homme (2017). Ces deux études considèrent un événement ayant supposément eu un impact majeur sur la terminologie d'un domaine donné. Paquin et Drouin définissent comme pivot les attaques du 11 septembre 2001 afin d'analyser leurs conséquences sur la terminologie du terrorisme tandis que Lafrance, Drouin & L'Homme s'intéressent aux changements terminologiques du domaine de l'informatique résultant de la diffusion de *Facebook* au grand public. Les deux études définissent également un corpus « pré-événement » (dans notre cas, les références bibliographiques) et un corpus « post-événement » (dans notre

cas, les productions scientifiques écrites post *Summer School*). Nous avons pris le parti de nous approprier cette méthode en définissant non seulement un événement-pivot mais aussi un corpus-pivot. Ce faisant, nous constituons un observatoire de néologie première : nous sommes en mesure d'observer si un terme a été employé pour la première fois dans le contexte de la *Summer School*.

3.2 Constitution des corpus de travail

3.2.1 Le projet XAIDA

Notre recherche s'appuie sur l'analyse de la production scientifique de XAIDA, un projet financé par l'Union européenne et lancé en 2021. Seize organisations de recherche sont mobilisées dans une démarche multidisciplinaire, le consortium réunit aussi bien des chercheurs spécialisés en modélisation climatique que des chercheurs spécialisés en intelligence artificielle. Leur objectif concret est de mettre au point des modèles de prédiction climatique plus efficaces et plus précis à l'aide de l'intelligence artificielle, afin de détecter des événements météorologiques extrêmes⁶ et de les attribuer, ou non, au changement climatique.

Ce projet a retenu notre attention pour plusieurs raisons. Tout d'abord, il constitue un premier point de croisement de deux domaines techniques extrêmement dynamiques. Nous faisons l'hypothèse que cette rencontre est susceptible d'être à l'origine d'un croisement des intérêts de recherche et des concepts mobilisée, ce qui constitue un potentiel terreau fertile pour la néologie terminologique.

Par ailleurs, comme évoqué précédemment, une *Summer School* a été organisée par le consortium en 2022. Les plus de vingt interventions qui l'ont composée nous ont été rendues accessibles et ont constitué la base de notre corpus oral.

Enfin, le projet étant encore actif, il nous est relativement aisé de faire appel aux experts et nous avons pu réaliser une série d'entretiens par visioconférence avec certains d'entre eux dont nous présenterons certains éléments dans les prochaines sections.

⁶ Un événement météorologique extrême est, selon la définition de l'Organisation Météorologique Mondiale (OMM), « rare, à un lieu et un moment de l'année particuliers, présente des caractéristiques extrêmes en matière d'ampleur, de localisation, de durée et de degré. Les propriétés de ce qui est désigné par "événement météorologique extrême" peuvent donc varier en fonction des normes locales ». Définition repérée sur : <https://wmo.int/topics/extreme-weather>.

3.2.2 Le corpus oral

Le premier corpus que nous avons défini est notre corpus-pivot, le corpus oral : *SummerSchool 2022*. Il comprend l'ensemble des 23 présentations réalisées par plus de 30 intervenants. Sa durée est de 27 heures et 30 minutes pour un total de 265 095 mots⁷. Pour des raisons de confidentialité, nous n'avons été autorisés à retranscrire lesdites interventions que manuellement, ce que nous avons réalisé via le logiciel *CLAN*⁸.

3.2.3 Le corpus écrit de référence

Nous avons ensuite compilé un corpus de références bibliographiques écrites, *RefBiblio*, qui est composé des 329 références bibliographiques mentionnées par les intervenants au cours de leurs présentations, car nous postulons qu'elles constituent un réservoir conceptuel comme linguistique dans lequel les intervenants puisent explicitement afin de mettre au jour leurs propres travaux. Parmi celles-ci, on peut noter une majorité d'articles de recherche, du rapport de l'année 2001 du *Groupe d'experts intergouvernemental sur l'évolution du climat* (GIEC), de quelques ouvrages collectifs (M. Lee & E.-J. Wagenmakers, 2013, *Bayesian Cognitive Modeling: A Practical Course*, Cambridge University Press, entre autres). Il servira ici de corpus d'exclusion, en somme une ressource spécialisée écrite permettant de confirmer ou d'infirmer préliminairement le statut néologique d'une unité (Sablayrolles, 2006, p. 145). Il comprend environ 3 151 049 mots et couvre une période s'étendant de 1963 à 2022. Bien que ce corpus soit volumineux, il n'est bien entendu pas exhaustif.

3.2.4 Le corpus d'étude de l'implantation

Notre troisième et dernier corpus, le corpus d'étude de l'implantation, *Post22*, tient son nom de Quirion (2000, p. 55). Selon les mots de l'auteur, il permet « l'examen d'une terminologie en voie d'installation » (*ibid.*). Dans notre cas, il servira à observer une éventuelle transition des candidats-termes relevés, du contexte oral au contexte écrit. Il est donc composé de tous les articles publiés par les intervenants depuis la *Summer School 2022*. Il totalise 896 888 mots et c'est un corpus ouvert que nous alimentons au gré des publications des différentes parties prenantes de la *Summer School 2022*.

⁷ Le total des mots déterminé pour ce corpus, comme pour ceux qui suivent, a été réalisé via TXM.

⁸ <https://dali.talkbank.org/clang/>

TABLEAU 1. Récapitulatif de la taille des différents corpus à l'étude

| Nom du corpus | Nombre de mots |
|---------------------------|----------------|
| <i>Summer School 2022</i> | 286 237 |
| <i>RefBiblio</i> | 3 151 049 |
| <i>Post22</i> | 1 527 909 |

3.3 Le logiciel *NeoTriTerm-M*

Afin d'identifier des cas potentiels de néologismes premiers, une condition préalable facilement observable est requise : la première occurrence datable en contexte oral. Elle implique donc, dans le cadre de nos travaux, une première condition : l'absence dans *RefBiblio*. Une seconde condition peut être ajoutée : la présence du candidat-terme dans *Post22*, afin d'observer, ou non, une potentielle implantation en contexte écrit.

NeoTriTerm-M, l'outil que nous avons développé dans le cadre de nos recherches doctorales, fonctionne avec nos trois corpus⁹. Son fonctionnement repose sur la relation absence/présence que nous avons décrite. Si un candidat-terme extrait du corpus oral est présent dans *RefBiblio*, il ne recevra aucune coloration (il n'est pas un candidat à la néologie première) dans le tableur de sortie. Si le candidat-terme est absent de *RefBiblio* et de *Post22*, il sera coloré en jaune (il est susceptible d'être un néologisme premier mais n'est vraisemblablement pas encore implanté à l'écrit). Enfin, si un candidat-terme est absent de *RefBiblio* mais présent dans *Post22*, il sera considéré comme (préliminairement) intégré à l'écrit et sera coloré en vert.

Schématiquement, ce logiciel met en œuvre la synthèse de notre approche méthodologique (recours au corpus d'exclusion, au corpus d'étude de l'implantation, au corpus-pivot, considération de la néologie première).

3.4 Extractions terminologiques

3.4.1 Extraction manuelle

L'obligation de retranscription manuelle de notre corpus oral nous a incité à considérer une extraction manuelle des termes utilisés. Nous avons donc retenu des termes simples

⁹ Il mobilise également un extrait de la partie orale du *COCA Corpus* (<https://www.english-corpora.org/coca/>) afin de calculer la spécificité de manière plus précise. Un corpus de référence écrit ne permettra pas de filtrer certains mots propres à la communication orale comme *yeah* ou *right* (interjection).

comme complexes : des éléments lexicaux dont la nature de terme présumée résulte de notre position de néophyte des domaines concernés. Nous avons pu ainsi obtenir, à partir des différentes interventions de la *Summer School*, un total de 4230 termes. Après exclusion des doublons, leur nombre final a été ramené à 2627.

Étant donné que nous n'avons encore pas pu optimiser le fonctionnement de *NeoTriTerm-M*, celui-ci requiert des ressources computationnelles importantes, notamment du point de vue de la longueur du traitement des données, nous n'avons pas pu réaliser un traitement unique de l'ensemble des candidats-termes considérés. Nous avons été contraints de les répartir en sept différents sous-ensembles de tailles comparables. Le tableau suivant récapitule le traitement qu'en a réalisé notre outil :

TABLEAU 2. Récapitulatif de la taille des différents sous-ensembles de candidats-termes relevés manuellement et résultats de traitement par *NeoTriTerm-M*

| Sous-ensemble | CT | CT Abs. RefBiblio | CT Abs. RefBiblio + Prés. Post 22 | CT exclus |
|---------------|------|-------------------|-----------------------------------|-----------|
| SE n°1 | 436 | 144 | 9 | 313 |
| SE n°2 | 439 | 99 | 19 | 321 |
| SE n°3 | 349 | 91 | 14 | 244 |
| SE n°4 | 347 | 74 | 7 | 266 |
| SE n°5 | 347 | 87 | 16 | 244 |
| SE n°6 | 349 | 86 | 16 | 247 |
| SE n°7 | 360 | 79 | 9 | 272 |
| Totaux | 2627 | 630 | 90 | 1907 |

Sur l'ensemble des 2627 candidats-termes extraits manuellement, 1907, soit environ 73 % n'ont pas été créés au cours de la *Summer School* et ne constituent donc pas des candidats à la néologie première (par exemple *clustering algorithm*, *stratospheric polar vortex*, *signal-to-noise ratio* ou *statistical non-stationarity*). Les 720 restants, soit environ 27 %, nécessitent une analyse supplémentaire.

3.4.2 Extraction semi-automatique via *TermoStat Web 3.0*

En complément de l'extraction manuelle, nous avons réalisé une extraction semi-automatique via *TermoStat*¹⁰. Il est probable, du moins envisageable, que le logiciel détecte des termes qui nous auraient échappé au cours de la retranscription. Dans tous les cas, il nous permet d'obtenir des données chiffrées (fréquence, spécificité, entre

¹⁰ Nous avons ainsi traité les données après avoir reçu l'autorisation nécessaire auprès de la direction du projet XAIDA.

autres) concernant l'emploi des termes qui nous seront utiles pour de futurs traitements. Nous avons obtenu 2868 candidats-termes. Les résultats sont consignés comme suit :

TABLEAU 3. Récapitulatif de la taille des différents sous-ensembles de candidats-termes extraits via *TermoStatWeb 3.0* et résultats de traitement par *NeoTriTerm-M*

| Sous-ensemble | CT | CT Abs. <i>RefBiblio</i> | CT Abs. <i>RefBiblio</i> + Prés. Post 22 | CT exclus |
|---------------|-------------|--------------------------|--|-------------|
| SE n°1 | 483 | 140 | 21 | 322 |
| SE n°2 | 483 | 124 | 22 | 337 |
| SE n°3 | 481 | 60 | 18 | 403 |
| SE n°4 | 481 | 57 | 6 | 418 |
| SE n°5 | 478 | 12 | 2 | 464 |
| SE n°6 | 462 | 2 | 1 | 459 |
| Totaux | 2868 | 395 | 70 | 2403 |

Ici, la proportion de termes exclus par notre outil est supérieure à celle que nous avons obtenue via l'extraction manuelle : 2403 candidats-termes n'ont pas été retenus, soit environ 84 %. Les sections suivantes sont consacrées à l'étude de quelques-unes des unités triées par notre logiciel.

Les termes remplissant la condition binaire absence/présence (colorés en vert par le logiciel), soit 90 pour l'extraction manuelle et 70 pour l'extraction outillée, ont été analysés et aucun ne relevait effectivement de néologie première.

3.4.3 Termes remplissant la condition binaire retenus

Le tableau suivant présente un échantillon de quelques trente termes-candidats extraits du corpus oral et colorés en jaune par *NeoTriTerm-M*, signifiant ainsi qu'ils remplissent notre condition d'absence de *RefBiblio*. De plus, nous avons vérifié manuellement qu'ils n'étaient pas présents dans d'autres ressources non comprises dans notre corpus de référence. Notons également qu'à l'instant où nous rédigeons ces mots, aucun n'a (encore) franchi la frontière entre l'oral et l'écrit. Le tableau contient également les mesures de spécificité calculées par *NeoTriTerm-M* et *TermoStatWeb 3.0*, mesures qui s'appuient sur le même calcul de spécificité (Lafon, 1980) et qui seront considérées dans nos analyses.

TABLEAU 4. Présentation de quelques candidats-termes relevant de la néologie première

| CT | Spécificité (TermoStat) | Spécificité (NeoTriTerm -M) |
|---|----------------------------|-----------------------------------|
| analogue temperature map | 1,8 | 0,73 |
| anchor constraint-based fingerprint | 1,8 | 0,73 |
| anchor regression algorithm | 1,8 | 0,73 |
| anchor regression input | 1,8 | 0,73 |
| anchor-based fingerprint | 1,8 | 0,73 |
| anthropogenic sea surface temperature fingerprint | 1,8 | 0,73 |
| backwards impact-driven approach | 1,8 | 0,73 |
| circulation drivers storyline | 1,8 | 0,73 |
| circulation-induced prediction | 1,8 | 0,73 |
| circum-hemispheric upper-level wind anomaly | 1,8 | 0,73 |
| circum-hemispheric wave pattern | 1,8 | 0,73 |
| circum-hemispheric wave structure | 1,8 | 0,73 |
| circum-hemispheric zonal flow pattern | 1,8 | 0,73 |
| convection-permitting model uncertainty | 1,8 | 0,73 |
| counterclockwise wave breaking | N/A ¹¹ | 0,73 |
| detection-attribution model | 1,8 | 0,73 |
| disruptive surface wind | 1,8 | 0,73 |
| hourly storm rainfall rate | 1,8 | 0,73 |
| internal intervariability | 1,8 | 0,73 |
| omega-shaped circulation | 1,8 | 0,73 |
| out-of-distribution heatwave | 1,8 | 0,73 |
| physically-based statistical bias correction | 1,8 | 0,73 |
| physicizing layer | 1,8 | 0,73 |
| process-based covariate approach | 1,8 | 0,73 |
| seeder-feeder circulation | 1,8 | 0,73 |
| spatial large-scale circulation | 1,8 | 0,73 |
| spatially compounding extreme weather event | N/A | 0,73 |
| stratospheric storyline | 1,8 | 0,73 |
| univariate impact indicator | 1,8 | 0,73 |

¹¹ Dans ce cas précis, le logiciel TermoStat n'a pas considéré le syntagme car breaking est erronément considéré comme un verbe plutôt que comme un nom du fait de la présence du suffixe -ing. Il s'agit d'une erreur d'étiquetage.

4 Discussion

Les prochains paragraphes seront consacrés à l'étude de quelques néologismes premiers issus de nos extractions et aux résultats d'AFC.

4.1 Résultats d'AFC

Nous avons initialement prévu de réaliser une Analyse Factorielle de Correspondances via *TXM*. Pour ce faire, le logiciel nécessite un minimum de quatre sous-corpus. Nous avons donc divisé le corpus *Post22* en trois sous-ensembles de tailles comparables. À ces sous-corpus a été comparé le corpus *SummerSchool 2022*. Les quatre sous-corpus sont ainsi constitués :

TABLEAU 5. Récapitulatif de la taille des différents sous-corpus pour l'Analyse Factorielle de Correspondances

| Sous-corpus | Nombre de mots contenus |
|--------------|-------------------------|
| Écrit 1 | 371 885 |
| Écrit 2 | 404 914 |
| Écrit 3 | 445 118 |
| Oral | 265 095 |
| Total | 1 487 012 |

L'AFC s'appuie sur un nombre x d'unités lexicales (mots grammaticaux exclus), triées par ordre de fréquence décroissant¹². En complément de cette première approche lexicale, et dans un souci d'ancrage de notre démarche dans une perspective terminologique, nous avons également souhaité réaliser deux AFC complémentaires ayant pour base les 1000 termes simples (c'est-à-dire composés d'un seul mot) les plus spécifiques à notre corpus combiné (*SummerSchool 2022* + *Post22*), obtenus via *TermoStatWeb 3.0* et *AntConc*¹³.

Notre première AFC nous a permis d'obtenir les résultats suivants :

¹² Cela correspond, dans le logiciel *TXM*, au paramètre *Vmax*. Si celui-ci est fixé à cent, l'AFC réalisée s'appuiera sur les cent unités lexicales les plus fréquentes dans les différents sous-corpus (*TXM 2019* : 103).

¹³ Le manuel du premier outil mobilise la formule de Lafon (1980) « afin de cerner le vocabulaire spécifique à un sous-corpus par rapport à l'ensemble d'un corpus ». *AntConc* propose, quant à lui, une extraction (très) relativement similaire¹³ des termes « anormalement fréquents » dans un corpus fourni, par rapport à un corpus de référence (*Manuel AntConc 2024* : 7). Le programme définit une valeur de keyness (que nous franciserons en spécificité) via l'application d'une log-vraisemblance (ou Log-Likelihood en anglais), sur la base d'une valeur critique choisie par l'utilisateur. Pour nos expérimentations, nous avons conservé la valeur critique par défaut égale à 3,84, ce qui correspond à un risque $p < 0,05$.

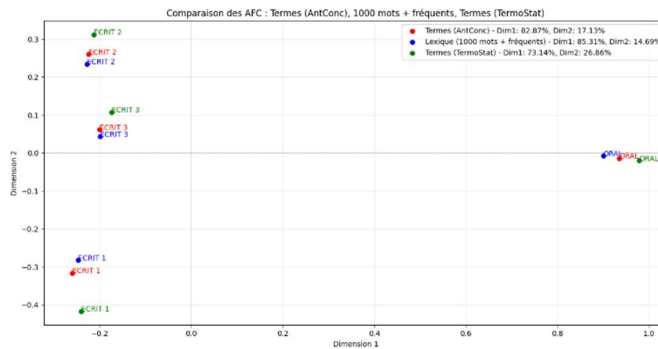


FIGURE 1. Analyses Factorielles de Correspondances non normalisées, basées sur le lexique et sur des listes de termes

Ces premières analyses permettent de mettre en avant une distance relativement faible entre l’AFC classiquement lexicale et l’AFC terminologique. Nous ne considérerons que cette dernière, car nous l’estimons plus juste et plus fine dans le cadre de notre étude car basée sur la spécificité plutôt que sur la fréquence¹⁴.

Concernant la différence entre l’oral spécialisé et l’écrit spécialisé, nous constatons ici une différence notable. Les trois sous-corpus écrits sont relativement proches (la « Dimension 2 » n’occupant qu’une importance relative, comprise entre 14,16 % et 24,79 %) et le sous-corpus oral beaucoup plus à la marge. La terminologie (comme le lexique) constitue un réservoir linguistique qui distingue clairement la communication orale de la communication écrite. Cela laisse supposer que chaque mode d’expression possède un encodage propre.

4.2 Le terme *counterclockwise wave breaking*

Le candidat-terme *counterclockwise wave breaking* a été mobilisé pour la première fois par Gabriele Messori (GM), chercheur du projet *XAIDA* et intervenant au cours de la *Summer School 2022*, dans l’extrait suivant :

- (1) | {}¹⁵ obviously |¹⁶ waves can break two ways | you can have cyclonic wave breaking or anticyclonic wave breaking | or clockwise or **counterclockwise wave breaking** if you're not familiar with this terminology | {} and here what they've done is that

¹⁴ Et ce quand bien même il apparaît que la variable « fréquence » possède une incidence très importante sur le résultat des calculs de spécificité.

¹⁵ Le symbole « {} » marque une hésitation du locuteur.

¹⁶ Le symbole « | » marque la fin d’un tour de parole.

they've looked at the | anomalies in the frequency of occurrence of this wave breaking | [GM_300622]

Gabriele Messori construit ici la première occurrence d'un antonyme au terme *clockwise wave breaking*, utilisé dans un article de recherche de Tim Wollings, en 2010 :

The normal north to south gradient in stream function has been reversed by a large-scale anticyclonic (**clockwise**) **wave-breaking** [gras ajouté] pattern.

Le candidat-terme *counterclockwise wave breaking* constitue donc un antonyme d'une alternative d'ores et déjà existante (bien qu'hapaxique sur le moteur de recherche *Google*), alternative fondée sur une analogie relative au sens de rotation des aiguilles d'une montre.

Concernant la proposition originale de Wollings, la mention de *clockwise* imbriquée entre parenthèses ouvre une piste de réflexion. Elle nous laisse penser que Wollings propose cette alternative afin de rendre le concept plus accessible à son lectorat.

Le contexte d'énonciation dans lequel Gabriele Messori propose sa création nous fait faire l'hypothèse qu'il s'inscrit dans une démarche similaire à celle de Wollings. En effet, la présence de la conjonction *or*, ainsi que la mention *if you're not familiar with this terminology* sous-tendent que le terme proposé constitue une alternative au terme plus couramment usité. Cela pose d'ailleurs, au passage, la question du statut de la conjonction de coordination *or* en contexte oral. S'agit-il ici de proposer deux termes équivalents mobilisables indifféremment ou s'agit-il de hiérarchiser entre le terme et une de ses variantes, attribuant ainsi le rôle possiblement joué par les parenthèses à *or* ? Rappelons, concernant ce dernier point, que l'oral est logiquement davantage dépourvu en matière de signes graphiques (parenthèses, guillemets, entre autres) permettant, dans certains cas, d'identifier des néologismes, techniques ou non (Sablayrolles, 2012, p. 41). La mise au point d'une typologie des équivalents oraux de ces indices graphiques écrits constitue une piste de recherche intéressante.

Le syntagme *anticyclonic wave breaking* est largement usité dans le domaine. Il est surprenant de ne constater l'apparition d'une variante plus simple qu'en 2022. Postulons ici que Gabriele Messori s'emploie à créer un pont entre ses connaissances notionnelles et linguistiques propres et celles qu'ils supposent de son auditoire. Il ne s'agit pas ici d'une création néologique à visée didactique. Malgré l'ambiguïté que peut revêtir le nom *Summer School*, le public n'est pas uniquement constitué d'étudiants en

situation d'apprentissage. L'objectif visé par les membres du projet XAIDA était, avant tout, de mettre au point et de partager des techniques et de théories concernant l'attribution d'événements météorologiques extrêmes au changement climatique, d'informer quant au danger que celui-ci représente, entre autres¹⁷. Il ne s'agit pas non plus d'un néologisme de vulgarisation, le discours n'est pas adressé au grand public.

L'entretien mené avec Gabriele Messori a confirmé notre hypothèse concernant le rôle d'interface que joue ce terme :

Ce n'est pas un terme (*counterclockwise*) que j'utiliserais à l'écrit, dans un article, parce qu'on utilise *cyclonic* et *anticyclonic*. Là, je l'ai utilisé seulement pour me faire comprendre d'un public d'étudiants. [Ent_GM_040424]

Le chercheur a conscience que son auditoire, composé de doctorants en sciences climatologiques, certes, mais également d'experts en intelligence artificielle, demeure relativement hétérogène d'un point de vue notionnel comme linguistique. Il cherche donc à rester le plus intelligible possible auprès de son public en mobilisant un terme plus aisément compréhensible par les récepteurs du discours (Faber, 2012, p. 234). Le terme, créé par incrémentation, via l'ajout d'un élément nouveau (*counterclockwise*) à un élément ancien (*wave breaking*), constitue, du moins nous en faisons l'hypothèse, un néologisme de transposition.

Concernant la pérennisation du syntagme, le chercheur a également confirmé que son objectif n'était pas de promouvoir l'emploi du terme en contexte scientifique. Il précise qu'il est personnellement susceptible de le mobiliser à l'écrit uniquement dans le cadre de la rédaction d'un ouvrage cette fois-ci didactique.

4.3 Le terme *hourly storm rainfall rate*

Nous allons ici nous intéresser à notre second candidat-terme issu de notre corpus oral : *hourly storm rainfall rate*. Le syntagme a été employé par Davide Faranda, autre membre du projet XAIDA, dans le contexte suivant :

¹⁷ « The objectives were clearly formulated in the proposal of the project. First of all, to share and devise techniques to tackle the problem of attributing meteorological extreme events to climate change. Secondly, to raise awareness of the possible consequences of changing climate on vulnerable societies and ecosystems, often located in developing countries, through rigorous attribution. Finally, the open source and global nature of the techniques taught in the school could then be used by scientists in developing countries independently for local attribution studies ». Citation extraite de la page dédiée à l'organisation de la *Summer School 2022* du projet XAIDA, disponible à l'adresse suivante : <https://xaida.eu/summer-school/>

(2) | the extreme three **hourly storm rainfall rates** | and rainfall amounts during- during the 2020 hurricane season | by ten and five percent respectively | [DF_270622]

Il sera intéressant de constater, en premier lieu, que le syntagme *hourly rainfall rate* existe, avec environ 17 100 occurrences sur *Google*. Le terme complexe *rainfall rate* existe également, avec environ 330 000 occurrences en ligne et une occurrence dans notre corpus *RefBiblio*. Il est même lexicalisé dans le glossaire de la société américaine de météorologie (*American Meteorological Society, AMS*) comme suit : *A measure of the intensity of rainfall by calculating the amount of rain that would fall over a given interval of time if the rainfall intensity were constant over that time period*¹⁸.

La création de *hourly rainfall rate*, par incrémentation de *hourly*, précise la durée sur laquelle la mesure est effectuée. L'ajout de *storm* traduit la volonté du chercheur de préciser l'origine des chutes de pluie qu'il souhaite mesurer, les précipitations produites par des tempêtes. Aussi, il convient de rappeler que la pluie peut être le fruit de plusieurs phénomènes météorologiques et n'est pas l'apanage des tempêtes. Il existe en effet différents types de précipitations, par exemple les précipitations orographiques, causées par « la montée d'air humide par-dessus une barrière orographique » (Grand Dictionnaire Terminologique, 2011)¹⁹. Rappelons également que les membres du projet *XAIDA* portent un intérêt particulier aux événements météorologiques/climatologiques extrêmes, les chutes de pluie résultant d'épisodes de tempête sont d'une intensité différente et ont des conséquences autrement plus graves. Les précipitations issues de la tempête Alex de 2020 ont entraîné plus de dix décès en France (Melun et al., 2022) et ont engendré des dommages chiffrés à plus d'un milliard d'euros (IGEDD, 2023). Il convient donc de distinguer les chutes de pluie lourdes de conséquences de celles qui enjoignent au simple usage d'un parapluie. C'est précisément l'intention que nous prêtons à Davide Faranda et qu'il a eu l'occasion de corroborer lors de notre entretien :

Pour moi, ça existe ce que je dis après : *hourly rainfall rate*. Pour moi, ce que j'ai ajouté *in-between* c'est *storm*. Je voulais dire que ce qui m'intéressait ici, ce ne sont pas nécessairement tous les *hourly rainfall rate* mais c'est *hourly rainfall rate* conditionné aux tempêtes. [Ent_DF_140324]

Il précise toutefois qu'il a mis en discours ce néologisme de manière inconsciente. Cela situerait donc le nouveau terme dans une catégorie de néologisme énonciatif pratique.

¹⁸ https://glossary.ametsoc.org/wiki/Rainfall_rate. Consultée le 10 janvier 2025.

¹⁹ <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26516682/precipitation-orographique>. Consultée le 10 janvier 2025

4.4 Les termes *circum-hemispheric* + NOMS

Nos analyses nous ont également permis d'observer un patron de création terminologique, toujours chez Gabriele Messori. En effet, les quatre extraits suivants montrent des exemples de création par incrémentation pour lesquels le chercheur ajoute l'adjectif *circum-hemispheric*, qui constitue donc ici l'élément nouveau, à des syntagmes nominaux d'ores et déjà lexicalisés :

(3) and you see these vectors here pointing almost exactly | south to north over the north atlantic so | which are part of | a larger scale | **circum-hemispheric upper-level wind anomaly** |[GM_300622]

(4) | they do {} a decent job also relating them to these | {} these large-scale anomalies | for other things | there is a big spread | so i think [] did some analysis | some preliminary analysis on these | wave | {} these **circum-hemispheric wave patterns** | and i think different models for example | they favor | different wave numbers |[GM_300622]

(5) | there was also a heatwave in Japan | ok and this can be related to the fact that there was an almost **circum-hemispheric wave structure** | {} | in the northern hemisphere mid latitudes | with a kind of regular or alternance of [] low and high pressure centers |[GM_300622]

(6) | most of the red regions are also | dotted meaning that they | saw a clear connection between jet stream intensity and cyclone clustering | so again linking large scale picture of how this **circum-hemispheric zonal flow pattern** varies | with regional anomalies in | cyclone clustering or cyclone frequency | [GM_300622]

Ici, les quatre syntagmes nominaux constituant l'élément ancien sont ou bien présents dans *RefBiblio*, ou bien dans *Google*, ou bien les deux :

TABLEAU 6. Récapitulatif du nombre d'occurrences des différents syntagmes nominaux auxquels ont été ajoutés *circum-hemispheric* au cours de la *Summer School*

| Syntagme nominal | Nombre d'occurrences (<i>RefBiblio</i>) | Nombre d'occurrences (<i>Google</i>) |
|---------------------------------|--|---|
| <i>upper-level wind anomaly</i> | 0 | 3 880 |
| <i>wave pattern</i> | 36 | 5 220 000 |
| <i>wave structure</i> | 4 | 421 000 |
| <i>zonal flow pattern</i> | 0 | 21 000 |

De la même manière que l'incrémentation du nom *storm* dans le syntagme *hourly rainfall rate* circonscrit le type de précipitations qui sont considérées par le chercheur, l'utilisation de *circum-hemispheric* délimite, cette fois-ci à un hémisphère, l'espace dans lequel est observé le reste du syntagme. L'entretien mené avec Gabriele Messori a confirmé cette analyse :

Cela pourrait être parce que je me focalisais sur l'hémisphère nord, par exemple, et donc je dis *circum-hemispheric* plutôt que *circum-global*. [...] Si j'avais parlé de l'hémisphère nord et de l'hémisphère sud (ensemble), j'aurais utilisé *circum-global*. [Ent_GM_040424]

Le chercheur nous a également indiqué qu'il pourrait tout à fait envisager d'utiliser ces termes dans un contexte écrit. Au moment où nous rédigeons ces lignes, aucun n'a cependant été réutilisé.

4.5 La question de la termicité pour évaluer le degré de spécificité d'un néologisme premier

Si l'on considère le calcul de spécificité de Lafon (1980) utilisé par *TermoStat*, rappelons les mots de Drouin (2003, p. 101) :

In this research, we are interested only in **items that appear in the analysis corpus more often than can be predicted from the global corpus** [gras ajouté] (SP+). To isolate them amongst the set of specific items, we use a test-value threshold of +3.09, which means that probability of finding the observed frequency is less than 1/1 000. In other words, the SP- items have a test-value lower than or equal to -3.09, while the SP+ items have a test-value equal or higher to +3.09 and the SP0 items obtain a test-value between these two thresholds.

Plus la spécificité est élevée, plus il en est considéré probable qu'une unité lexicale appartienne à un domaine spécialisé et constitue ainsi un bon candidat à la termicité²⁰ (ou *termhood* en anglais), au sens défini par Kageura et Umino (1996, p. 270) :

On the other hand, termhood refers to the degree that a linguistic unit is related to (or more straightforwardly, represents) domain-specific concepts. Thus 'termhood' is not only relevant to complex linguistic units, but also to simple units.

Le seuil de significativité fixé à 3,09 implique donc que *TermoStat* n'affiche pas les unités simples ou complexes avec un score de spécificité inférieur. Puisque la fréquence est

²⁰ C'est le sens que nous prêtons aux mots de L'Homme (2005 : 1124), qui écrit de la spécificité qu'elle est un « indice fort du statut terminologique des unités ».

une variable de premier plan, largement constitutive de la valeur finale de spécificité (Drouin, 2004, p. 346 ; voir également nos observations sur l'AFC) et que nous considérons des candidats-termes créés au cours des interventions de la *Summer School*, avec des fréquences que nous postulons faibles, voire hapaxiques²¹, le seuil est susceptible d'entraver l'extraction de néologismes premiers²². C'est d'ailleurs le cas avec les termes dont nous venons de disséquer. En effet, *counterclockwise wave breaking* ne possède qu'une spécificité calculée par *NeoTriTerm-M* de 0,73, ce qui correspond vraisemblablement à une spécificité calculée par la version sans seuil de *Termostat* de 2,91. La spécificité de *hourly storm rainfall rate* n'est évaluée qu'à 0,73, le même score que pour les syntagmes *circum-hemispheric* + NOMS.

Ces observations vont dans le sens de la vision scalaire de la termicité, telle que rappelée par Humbley (2018, p. 81) : « [...] un élément peut ainsi être plus ou moins néologique, plus ou moins terminologique » et s'inscrivent dans l'assertion suivante de Cabré (2023, pp. 46-47) :

We postulate that a **lexical unit is by itself neither terminological nor general** but that **it is general by default and acquires special or terminological meaning when this is activated by the pragmatic characteristics of the discourse** [gras ajouté].

La notion ici évoquée par Cabré de « caractéristiques pragmatiques du discours » est centrale lorsqu'il s'agit de considérer la terminologie en contexte oral. Puisque notre approche s'éloigne de la terminologie wüsterienne, en ce sens qu'elle considère le contexte d'énonciation, il convient de rappeler que l'expression orale peut constituer un mode d'expression qui plonge le locuteur dans une forme d'urgence (Suffys, 2000, p. 33).

À cet effet, Stumpf (2021), oppose deux notions : *language of immediacy* (que l'on peut apparenter à l'expression orale) et *language of distance* (que l'on rattachera plus aisément à l'expression écrite). Cette urgence que nous évoquons constitue un phénomène néologène. Sablayrolles (2006, p. 87) considère les créations, en français, de *examination* et *horribilité* comme ayant eu lieu « dans l'urgence, alors que la phrase a déjà commencé à être proférée ». Postulons qu'elle puisse être exacerbée par l'usage, quasi généralisé au sein de *XAIDA*, de l'anglais comme *lingua franca* : Sager (1997, cité

²¹ Sans que cela ne les constitue *de facto* mauvais candidats à la termicité.

²² Cartier & Sablayrolles (2009 : 178) indiquaient que « le dépouillement manuel traditionnel [avait] encore de beaux jours devant lui ». La situation, seize ans plus tard, souligne la validité de leur observation.

dans Humbley ,2018), avait écrit que la création terminologique dépendait, en autres critères, des compétences linguistiques du locuteur à son origine.

Le contexte d'énonciation oral plonge donc le locuteur dans une forme d'urgence qui ébranle ses capacités à produire un discours, qui plus est dans une langue qu'il ne maîtrise que relativement, rendant donc envisageable le recours à des créations *ex abrupto*.

Concernant ces créations, elles pourront être d'ordre morphologique en mobilisant l'incrémentation d'un élément technique nouveau (*storm, circum-hemispheric*) au sein d'un syntagme d'ores et déjà usité, comme elles pourront mobiliser une unité non spécialisée (*counterclockwise*) afin de remplacer une unité considérée (trop) spécialisée, voire opaque (*anticyclonic*) dans un syntagme préexistant (*wave breaking*), via donc un procédé sémantique d'analogie aboutissant à la mise en discours d'une variante moins lexicalisée et qui trahit la volonté des experts de « se rapprocher » de leur auditoire, observation également réalisée par Seghezzi (2013). Pour reprendre les mots cités précédemment de Cabré, l'unité lexicale *counterclockwise* obtient donc un « sens terminologique », jusque-là quiescent, du fait des « caractéristiques pragmatiques (urgence, anglais *lingua franca*) du discours ». Notons aussi que la composition apparaît comme matrice terminogénique de premier plan dans les contextes étudiés, concomitamment aux observations faites pour la langue générale, notamment par (Estopà, 2009).

5 Conclusion et pistes de réflexion

Cet article visait à présenter succinctement les quelques croisements entre analyse terminologique et corpus oraux ainsi qu'à aborder ces derniers par le prisme d'une description de la néologie spécialisée qu'ils contiennent.

Après avoir détaillé certaines caractéristiques définitoires de la notion de néologie première, nous nous sommes employés à décrire notre approche méthodologique qui tire son originalité de la considération d'un corpus-événement-pivot oral en relation avec deux autres corpus écrits de référence et d'implantation.

Nos premiers résultats nous ont permis d'observer la distance lexicale comme terminologique entre les corpus oraux et écrits. L'étude plus approfondie de

counterclockwise wave breaking, *hourly storm rainfall rate* et des syntagmes *circum-hemispheric* + NOMS nous a permis d'observer plusieurs néologismes premiers créés à des fins discursives différentes : l'interface entre les connaissances ou le besoin énonciatif, inconsciemment ou consciemment, mais témoignant quoi qu'il en soit d'une volonté d'établir une proximité avec les récepteurs du discours (Humbley 2008 : 37). Nous pouvons également nous poser la question du devenir de la néologie d'interface orale spécialisée en contexte écrit. Se prédestine-t-elle uniquement à transitionner vers une néologie à proprement parler didactique, voire de vulgarisation, ou peut-on envisager qu'elle réalise sa mue vers une variante qui sera perçue comme telle ?

Notons également que nous ne disposons d'aucun exemple de création *ex-nihilo*. Tous les néologismes premiers observés l'ont été via l'incrémentation d'un adjectif (*counterclockwise*, *circum-hemispheric*) ou d'un nom (*storm*) à des syntagmes en usage, parfois même recensés dans des ressources spécialisées. Ces observations tendent à confirmer l'importance du modèle incrémental, que Humbley (2018, p.102) souligne en contexte écrit. L'auteur cite également les travaux de Daille et al. (1996) qui montrent que les termes composés d'au moins deux unités sont les plus fréquents en contexte anglophone.

Ces termes ne possédaient qu'une seule et unique occurrence *in corpus* et un score de spécificité relativement faible, en conséquence, et surtout en deçà du seuil de significativité couramment employé. Cela souligne au passage, si besoin était, que les mesures statistiques, dont fait partie la spécificité, ne sauraient être totalement substituées à l'analyse qualitative par le terminologue. Il est également possible d'imaginer que la néologie première soit, en général, repérable via une fréquence et une spécificité faibles. L'extraction manuelle, éventuellement complétée par une extraction semi-automatique apparaît donc comme un moyen relativement efficace de détecter des cas de néologie première.

Toutefois, aucun néologisme premier heuristique n'a pu être observé et aucun des cas observés ne s'est vraisemblablement implanté en contexte écrit. L'enrichissement progressif de *Post22* pourra potentiellement y remédier. Toutefois, n'excluons pas que certains termes voient leur usage circonscrit à l'oralité (ce qui expliquerait partiellement les distances terminologiques observées via l'AFC). La néologie sémantique n'a pas été considérée ici, il conviendra donc de l'examiner plus en détail.

De la même manière, la néologie première ne recouvre ici qu'un pan de la néologie spécialisée qui intègre également des changements conceptuels appliqués à une désignation existante comme des variations de désignation pour un concept inchangé. Il sera donc nécessaire de considérer, à terme, des unités jusque-là absentes de notre champ d'analyse.

Références

- Beaudouin, V. (2016). Retour aux origines de la statistique textuelle : Benzécri et l'école française d'analyse des données. *13es Journées Internationales d'analyse statistique des données textuelle*, Nice, France, 17-27. <https://hal.science/hal-01376938/>
- Benzécri, J.-P. (1973). *L'analyse des Données : 2 L'Analyse des correspondances*. Paris, Dunod.
- Boulanger, J.-C. (2010). Sur l'existence des concepts de néologie et de néologisme : propos sur un paradoxe lexical et historique. In M.-T. Cabré, *Actes del I Congrés internacional de neologia de les llengües romàniques*, Girona, (pp. 31-73). https://boulanger.recherche.usherbrooke.ca/document-article-boulanger_2010
- Cabré, M.-T., & Nazar, R. (2012). Towards a new approach to the study of neology. *Neologica*, (6), 63-80. DOI: 10.15122/ISBN.978-2-8124-4232-2.P.0063
- Cabré, M.-T. (2000). Terminologie et linguistique : la théorie des portes. *Terminologies nouvelles*, 21, 10-15.
- Cabré, M.-T. (2023). *Terminology. Cognition, language and communication*. John Benjamins Publishing Company. <https://doi.org/10.1075/ivitra.36>
- Cabrera-Castro, T. (2017). Applications of Simultaneous-Interpreting Corpora in Terminology Research. *New voices in Translation Studies*, 17, 31-51. <https://doi.org/10.14456/nvts.2017.12>
- Candel, D. (1984). Une approche de la langue des physiciens. *Langue française*, 64, 93-108. <https://doi.org/10.3406/lfr.1984.5207>
- Candel, D. (1993). Français oral de spécialisé et terminologie. In Y. Gambier & F. Gaudin (eds.), *Socioterminologie, Le Langage et l'homme*, (pp. 259-272), XXVIII-4.
- Cartier, E., & Sablayrolles, J.-F. (2009). Néologismes, dictionnaires et informatique. *Cahiers de lexicologie*, 93(2), 175-192. DOI : 10.15122/ISBN.978-2-8124-4345-9.P.0179.
- Ciapuscio, G. (1997). Los científicos explican: la reformulación del léxico experto en la consulta oral. *Cadernos do Instituto de Letras*, 18, 37-47.
- Ciapuscio, G. (2007). Epistemic modality and academic orality: Pilot study for COTECA (Corpus Textual del Español Científico de la Argentina). In G. Parody, *Working with Spanish Corpora*, Londres: Continuum, 90-115.
- Condamines, A. (2010). Variations in terminology: Application to the management of risks related to language use in the workplace. *Terminology*, 16(1), 30-50. DOI : 10.1075/term.16.1.02con
- Condamines, A., Dehaut, N., & Picton, A. (2012). Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Études outillées en corpus. *Cahiers de lexicologie*, 1(100), 161-184. DOI : 10.15122/isbn.978-2-8124-4147-9.p.0161.
- Daille, B., Habert, B., Jacquemin, C., & Royauté, J. (1996). Empirical Observation of Term Variations and Principles for their Description. *Terminology*, 3(2), 197-257. <https://doi.org/10.1075/term.3.2.02dai>

- Delavigne, V. , & De Vecchi, D. (2021). *Termes en discours. Entreprises et organisations*. Presses de la Sorbonne Nouvelle.
- Delavigne, V., Picton, A., & Thibert, E. (2022). Socioterminologie et terminologie textuelle : l'expertise en questions. In F. Neveu, S. Prévost, A. Steuckardt, G. Bergounioux, B. Hamma (Éds.), *Congrès Mondial de Linguistique Française*, Orléans, France. <https://doi.org/10.1051/shsconf/202213804012>
- Desmet, I. (2007). Éléments pour une théorie variationniste de la terminologie et des langues de spécialité. *Cahiers du RIFAL*, 26, 3-13.
- De Vecchi, D. (1999). *La terminologie en entreprise : formes d'une singularité lexicale*. [thèse de doctorat, Université de Paris 13].
- Diagne, A. (2022). Constitution et Usage de Corpus pour Langues Terminologiquement Peu Documentées. *Revue Africaine et Malgache de Recherche Scientifique (RAMReS) : Littérature, Langues et Linguistique*, 14, 34-44. <https://publication.lecames.org/index.php/lit/article/viewFile/18157/1322>
- Domènech Bagaria, O., Estopà Bagot, R. (2009). La neología en textos orales. *Revista de Investigación Lingüística*, 12, 39-64.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9(1), 99-115. DOI: 10.1075/term.9.1.06dro
- Drouin, P. (2004). Spécificités lexicales et acquisition de la terminologie. In G. Purnelle, C., Fairon, A., Dister (Éds.), *Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles*, Louvain-la-Neuve, Belgique, 345-352. https://olst.ling.umontreal.ca/wp-content/uploads/olst/pdf/JADT_2004.pdf
- Dury, P. (2021). Les acronymes et les sigles spécialisés, des termes comme les autres ? Étude en diachronie courte dans le domaine médical. *Cahiers de lexicologie*, 1(118), 227-246. DOI : 10.48611/isbn.978-2-406-12006.3.p.0227.
- Erlös, F. (2012). L'activité prise au mot : le traitement des néologismes en entreprise. In P. Dury, J. Carlos de Hoyos, J. Makri-Morel, F. Maniez, V. Renner, B. Villar Diaz (Dir.), *La néologie en langue de spécialité – Détection, implantation et circulation des nouveaux termes*, Journées du CRTT 2012, 13-36.
- Estopà, R. (2008). L'observation de mots nouveaux en espagnol et en catalan au cours de deux décennies : l'Observatori de Neologia de l'IULA. *Neologica*, 2, 169-184.
- Estopà, R. (2009). Les composés populaires de la langue catalane : étude diachronique de la néologie. *Meta*, 54(3), 551-565. <https://doi.org/10.7202/038314ar>
- Faber, P. (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston : Mouton de Gruyter.
- Freixa, J. (2012). La néologie hispanique : analyse d'une éclosion. *Histoire Épistémologie Langage*, 34(2), 9-28. https://www.persee.fr/doc/hel_0750-8069_2012_num_34_2_3248
- García Antuña, M. (2015). La investigación en terminología. La relevancia del corpus oral para determinadas lenguas de especialidad. *Cuadernos Iberoamericanos*, 2, 47-55.
- Gaudin, F. (1995). Champs, clôtures et domaines : des langues de spécialité à la culture scientifique. *Meta*, 40(2), 229-237. <https://doi.org/10.7202/002821ar>
- Gaudin, F. (2005). La socioterminologie. *Langages*, 157(1), 81-93. <https://doi.org/10.3917/lang.157.0081>

- Gautier, L., Bach, M. (2017). La terminologie du vin au prisme des corpus oraux de dégustation/présentation (français-allemand) : entre émotions, culture et sensorialité. *Études de linguistique appliquée*, 188(4), 485-509.
- Humbley, J. (2004). La réception de l'œuvre d'Eugen Wüster dans les pays de langue française. *Les Cahiers du CIEL*, 33-51.
- Humbley, J. (2006). La néologie : interface entre ancien et nouveau. In R. Greenstein (Ed.), *Langues et cultures : une histoire d'interface*, Publications de la Sorbonne (pp. 91-103).
- Humbley, J. (2008). Issues of distance and proximity in neologisms, as instanced in e-commerce. *ASp, la revue du GERAS*, 53-54, 25-38. <https://doi.org/10.4000/asp.325>.
- Humbley, J. (2018). *La néologie terminologique*. Limoges : Lambert-Lucas.
- Kageura, K., & Umino, B. (1996). Methods of automatic term recognition. *Terminology*, 3(2), 259-289. <https://doi.org/10.1075/term.3.2.03kag>
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots. Les langages du politique*, 1, 127-165.
- Lafrance, A., Drouin, P., & L'Homme, M.-C. (2017). Découverte de la néologie et de la nécrologie dans la terminologie anglaise de l'informatique à l'aide d'une méthode outillée. *Neologica*, (11), 185-206. DOI : 10.15122/isbn.978-2-406-06995-9.p.0185
- Lanseur, S. (2016). Quel corpus pour une étude de néologie terminologique dans les textes de vulgarisation scientifique à propos de l'économie en Algérie ? In É. Comby (Éd.) *Corpus de textes : composer, mesurer, interpréter*, ENS Éditions (pp. 14-25). <https://doi.org/10.4000/books.enseditions.7348>.
- Lerat, P. (1995). *Les langues spécialisées*. Paris : Presses Universitaires de France.
- Leroyer, P. (2015). Turning the Corpus into a Functional Component of the Dictionary: The Case of the Oenolex Wine Dictionary. *Procedia - Social and Behavioral Sciences*, 198, 257-265.
- L'Homme, M.-C. (2005). Sur la notion de « terme ». *Meta*, 50(4), 1112-1132. <https://doi.org/10.7202/012064ar>
- Marcellesi, C. (1973). Le langage des techniciens de l'informatique : quelques aspects de leur vocabulaire écrit et oral. *Langue française*, 17, 59-71.
- Paquin, A., & Drouin, P. (2010). Néologie du terrorisme : médiatisation et intégration. *Neologica* (4), 105-118. DOI : 10.15122/isbn.978-2-8124-4230-8.p.0105
- Quirion, J. (2000). *Aspects évaluatifs de l'implantation terminologique*. [Thèse de doctorat, Université de Montréal].
- Rondeau, G. (1981). *Introduction à la terminologie, 2^e édition*. Québec : G. Morin.
- Sablayrolles, J.-F. (2006). La néologie aujourd'hui. In C. Gruaz (Dir.), *À la recherche du mot : De la langue au discours*, Lambert-Lucas, Limoges, France (pp. 141-157).
- Sablayrolles, J.-F. (2012). Extraction automatique et types de néologismes : une nécessaire clarification. *Cahiers de lexicologie*, 1(100), 37-53. DOI : 10.15122/isbn.978-2-8124-4147-9.p.0037
- Sager, J. C. (1989). Term Formation. *Lebende Sprachen*, 34, 154-161.
- Sager, J. C. (1997). Term Formation. In S.E. Wright, G. Budin (Éds.), *Handbook of Terminology Management*, Vol. 1, Amsterdam and Philadelphia, Benjamins, (pp. 25-41).
- Schneider, C., & Estopà, R. (2020). Methodology to build and exploit a representative corpus for neological study in the field of medicine. *Terminàlia*, 22, 29-39. DOI : 10.2436/20.2503.01.154

- Seghezzi, N. (2007). El papel de la oralidad en Terminología. *Interlingüística*, ISBN, 18, 1035-1047.
- Seghezzi, N. (2013). Variación terminológica: de la escritura a la oralidad. *Debate Terminológico*, 9, 62-80. <https://files.core.ac.uk/download/pdf/303982357.pdf>
- Stumpf, S. (2021). Occasional word formations in written and spoken German. *Neologica*, 15, 151-169. DOI : 10.48611/isbn.978-2-406-11896-1.p.0151
- Suffys, S. (2000). Un oral, des « orals », et autres voies orales.... *Recherches*, 33, 29-59. https://www.revue-recherches.fr/wp-content/uploads/2016/01/R33_29-59_Suffys.pdf
- Tallarico, G. (2022). Néologismes expressifs et ludiques dans le vocabulaire de la pandémie. *Repères DoRiF*, 25, Le lexique de la pandémie et ses variantes. <https://www.dorif.it/reperes/giovanni-tallarico-neologismes-expressifs-et-ludiques-dans-le-vocabulaire-de-la-pandemie/>
- Van Campenhoudt, M., & Temmerman, R. (2011). Les corpus et la recherche en terminologie et en traductologie. *Meta*, 56(2), 223-225. <https://doi.org/10.7202/1006173ar>
- Wüster, E. (1979). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Infoterm.
- Ylönen, Sabine. 2018. Oral Discourse in Scientific Research. In J. Humbley, G. Budin, C. Laurén (Éds.), *Languages for Special Purposes: An International Handbook*, Berlin, Boston: De Gruyter Mouton, 364-380.