

TermCD

TERMinologie,
Communication
et Discours

TermCD

TERMinologie,
Communication
et Discours

3

2025

TermCD

TERMinologie, Communication et Discours

Anno III - 3 (2025)

ISSN digitale 3034-8668

ISBN 979-12-5535-581-6

www.termcd.eu

Direction éditoriale | Direzione editoriale | Editors in chief

Maria Teresa Zanola, Università Cattolica del Sacro Cuore-Accademia della Crusca

Manuel Célio Conceição, Universidade do Algarve

Claudio Grimaldi, Università degli Studi di Napoli "Parthenope"

Comité éditorial | Comitato editoriale | Editorial Board

Paolo D'Achille, Accademia della Crusca-Università degli Studi Roma Tre

Ieda Maria Alves, Universidade de São Paulo

Paola Puccini, Alma Mater Studiorum-Università degli Studi di Bologna

Roma Kriauciūniene, Vilniaus Universitetas

Manuel González González, Universidade de Santiago de Compostela

Comité scientifique | Comitato scientifico | Scientific Board

Rosa Agost, Universitat Jaume I

Maria Helena Carreira, Université Paris 8

Concetta Cavallini, Università degli Studi di Bari "Aldo Moro"

Anne Condamines, Université de Toulouse

Rute Costa, Universidade Nova de Lisboa

Patrick Drouin, Université de Montréal

Isabel Margarida Duarte, Universidade do Porto

Wael Farouq, Università Cattolica del Sacro Cuore

Paolo Frassi, Università degli Studi di Verona

Aurélia Gaillard, Université Bordeaux Montaigne

Laurent Gautier, Université Bourgogne Europe

John Humbley, Université Paris Cité

Christine Jacquet-Pfau, Collège de France

Ofelia Palermo, Nottingham Trent University

Francesca Piselli, Università degli Studi di Perugia

Étienne Quillot, Délégation générale à la langue française et aux langues de France

Pedro Sousa, Universidade do Algarve

Annalisa Zanola, Università degli Studi di Brescia

Comité de rédaction | Comitato di redazione | Editorial Committee

Silvia Calvi, Università Cattolica del Sacro Cuore

Klara Dankova, Università Cattolica del Sacro Cuore

Fátima Noronha, Universidade do Algarve

Maria Chiara Salvatore, Università degli Studi di Napoli "Parthenope"

Silvia Domenica Zollo, Università degli Studi di Napoli "Parthenope"

Quest'opera è distribuita con Licenza Creative Commons CC BY-SA 4.0.

© 2026 **EDUCatt - Ente per il Diritto allo Studio Universitario dell'Università Cattolica**

Largo Gemelli 1, 20123 Milano | tel. 02.7234.2235 | fax 02.80.53.215

e-mail: editoriale.dsu@educatt.it (produzione); librario.dsu@educatt.it (distribuzione)

web: libri.educatt.online

Questo volume è stato stampato nel mese di settembre 2026

presso la Litografia Solari - Peschiera Borromeo (Milano)

su carta Fedrigoni Arcoset usomano certificata FSC*, di pura cellulosa ecologica E.C.F.
utilizzando il carattere Adobe Garamond Pro.

NUMÉRO THÉMATIQUE

*Terminologie et inderdisciplinarité :
défis et perspectives de recherche futures,*
dirigé par Maria Teresa Zanola
et Claudio Grimaldi

Table des matières

Introduction <i>Étienne Quillot</i>	7
La vie scientifique et culturelle de REALITER (2012-2023) <i>Maria Teresa Zanola, Claudio Grimaldi</i>	9
A terminologia da COVID Longa no Português Brasileiro e aspectos de sua variação <i>Ieda Maria Alves, Beatriz Curti-Contessoto, Márcia de Souza Luz-Freitas, Ana Maria Ribeiro de Jesus, Lucimara Alves da Conceição Costa</i>	21
La terminologie médicale en tant que pratique métalinguistique <i>Doina Butiurcă, Cristina-Alice Toma, Elena Museanu</i>	29
Traduction spécialisée et post-édition : pour une modélisation du processus de gestion des textes <i>Silvia Calvi, Klara Dankova</i>	39
<i>Un mare di termini: storia di un progetto terminologico come catalizzatore di collaborazioni interdisciplinari</i> <i>Ana Lourdes de Hériz, Joachim Gerdes, Veronica Monticelli, Martina Odello, Velia Pascale, Micaela Rossi, Laura Santini</i>	59
Imprécisions terminologiques et marketing : le cas de la bière <i>doppio malto</i> <i>Lorenzo Devilla, Nicla Mercurio</i>	71
Terminologie et traduction au prisme de l'interdisciplinarité : des débuts de la Terminologie textuelle à la Traduction automatique neuronale <i>Cécile Frérot</i>	83
Terminologie de l'endométriose et interdisciplinarité : propositions méthodologiques et étude de cas <i>Julie Humbert-Droz</i>	95

La terminologie de la télémédecine : enjeux et perspectives d'un domaine interdisciplinaire	109
<i>Carolina Iazzetta</i>	
<i>TermTestQA</i> : evaluación del contenido y de la calidad de la información generada por ChatGPT en el campo de la terminología	121
<i>Cristina Varga</i>	
En droit pénal français, il y a <i>prévenus</i> et <i>prévenus</i> : un cas de polysémie interdisciplinaire pas comme les autres	133
<i>Anca-Marina Velicu</i>	

Introduction

L'interdisciplinarité est consubstantielle à la terminologie, tant dans la production de ressources linguistiques que dans le discours sur ces ressources. La terminologie « ressource linguistique » procède de la description des arts et métiers, des techniques et des sciences. La terminologie « objet de recherche » croît et prospère dans l'échange avec d'autres disciplines pour produire des connaissances et des objets scientifiques nouveaux et originaux. La terminologie n'est donc ni autonome, ni isolée, elle est interdisciplinaire.

La XIX^e Journée scientifique REALITER, qui a lieu à Paris le 13 octobre 2023, dans les locaux de la Délégation générale à la langue française et aux langues de France du ministère de la Culture, en a fait la démonstration en explorant le thème de « Terminologie et interdisciplinarité : défis et perspectives de recherche futures », dont plusieurs communications sont réunies dans ce numéro de *TermCD*.

Cette XIX^e Journée scientifique offrait aussi l'occasion de commémorer la fondation de REALITER en 1993 à Paris, quand la Délégation générale à la langue française accueillait pour la première fois universitaires, chercheurs, représentants de centres de recherches et d'organismes de politique linguistique et d'aménagement terminologique, des hommes et des femmes venus de tous les points du monde roman, animés par l'idée que le développement de la terminologie panlatine serait, entre eux, le plus solide et le plus prometteur des liens de coopération. Maria Teresa Zanola et Claudio Grimaldi retracent ainsi les actions et les apports du réseau au cours de la dernière décennie.

Interdisciplinarité entre terminologie et médecine font de la terminologie une discipline médicale : la terminologie est au cœur des politiques de santé et de la relation singulière entre le soignant et le patient. Ainsi, Ieda Maria Alves et ses consœurs témoignent du caractère clé de l'analyse des variations terminologiques en portugais du Brésil dans l'identification et le traitement du covid long. Julie Humbert-Droz passe au crible le discours sur l'endométriose, illustrant l'importance de disposer d'un vocabulaire partagé, reconnu et validé dans la prise en charge de cette maladie. Carolina Iazzetta ausculte la transformation de la pratique médicale, qui passe de l'entretien dans le cabinet à la consultation à distance, concomitamment à l'apparition de notions et de dénominations floues, qui peinent à expliciter clairement les enjeux de cette mutation. Enfin, Doina Butiurcă, Alice Toma et Museanu interrogent le rôle de la fonction métalinguistique du langage dans la genèse de la terminologie médicale.

La terminologie est au service de toutes les sciences. Ana Lourdes de Hérizet et ses collègues du CERTEM décrivent comment la conception et la mise en œuvre d'une ressource terminologique peut faire de celle-ci la colonne vertébrale d'un vaste projet

multidisciplinaire sur la mer. Anca-Marina Velicu analyse finement les défis que pose en droit français la polysémie du terme « prévenu » dans la procédure pénale et dans le droit pénitentiaire. Enfin, Lorenzo Devilla et Nicla Mercurio dissèquent par le prisme de la terminologie les stratégies mercatiques de vente de la bière *doppio malto*.

L'interdisciplinarité en terminologie, c'est aussi de nouveaux outils et développements scientifiques et techniques qui ébranlent la pratique de la terminologie et sa relation aux autres disciplines linguistiques. À l'instar des humanités numériques, la terminologie devient numérique. Cécile Frérot revient aux sources du développement de la terminologie textuelle fondée sur l'ingénierie des connaissances et la linguistique de corpus et rappelle l'importance de la prise en compte des outils d'analyse de corpus lors la formation des traducteurs et des terminologues. La traduction via l'intelligence artificielle (IA) bouleverse profondément traduction et terminologie. Silvia Calvi et Klara Dankova analysent la qualité du traitement de la terminologie par différents outils de traduction d'IA, pour proposer une modélisation pour la gestion des textes. Cristina Varga, elle aussi, s'intéresse à la qualité des traductions produites grâce à l'IA en proposant, à partir de la constitution d'une base de données de questions types relatives à la terminologie, un modèle d'évaluation des grands modèles de langue.

Étienne Quillot

La vie scientifique et culturelle de REALITER (2012-2023)

MARIA TERESA ZANOLA, CLAUDIO GRIMALDI¹

1. Introduction

L'expérience associative internationale dans le domaine roman constitue le cœur des activités du Réseau panlatin de terminologie REALITER, fondé en 1993 dans le but de promouvoir le développement harmonisé des langues néolatines à partir de leur origine commune. La création du Réseau s'inscrit dans une initiative conjointe de la Délégation générale à la langue française et de l'Union latine, organisation engagée à l'époque dans la valorisation du patrimoine identitaire du monde latin. Les 13 et 14 décembre 1993, sous l'impulsion de Bernard Quémada, éminent linguiste et lexicographe, un groupe de travail est constitué à Paris afin de mettre en place une Commission panlatine de terminologie. Ce projet prend forme lors de cette première réunion sous la dénomination officielle de Réseau panlatin de terminologie (REALITER), appellation conjointement adoptée comme acronyme commun aux différentes désignations dans les langues du Réseau : *Red panlatina de terminología* en espagnol, *Rede panlatina de terminologia* en portugais, *Rede panlatina de terminoloxía* en galicien, *Réseau panlatin de terminologie* en français, *Rete panlatina di terminologia* en italien, *Reteaua panlatina de terminologie* en roumain et *Xarxa panllatina de terminologia* en catalan (Grimaldi, 2025).

REALITER regroupe aujourd'hui un vaste ensemble de chercheurs, enseignants-chercheurs, centres de recherche, institutions publiques, universités et associations œuvrant dans le domaine de la terminologie pour les sept langues actuellement représentées au sein du Réseau. Ces langues (catalan, espagnol, français, galicien, italien, portugais et roumain) sont utilisées aussi bien pour la communication interne et externe que comme langues de travail dans les activités scientifiques du Réseau (Zanola, 2014a). En raison de cette organisation et de la nature de ses actions, REALITER constitue un organisme de coopération unique entre pays et communautés de langues néolatines, réunissant des acteurs engagés dans la terminologie, la terminographie, la planification linguistique et l'harmonisation terminologique.

¹ La présente contribution est le résultat d'une réflexion commune. Cependant, Claudio Grimaldi a rédigé les paragraphes « Les débuts de REALITER : fondation, objectifs et principes fondateurs » et « Les lexiques REALITER : objectifs, méthodologie et portée », et Maria Teresa Zanola les paragraphes « Les Journées scientifiques REALITER (2012-2016) : thématiques et acquis » et « Les Journées scientifiques REALITER (2017-2022) : un élargissement progressif des thématiques de recherche ». Les deux auteurs ont corédigé l'Introduction et la Conclusion.

Les Journées scientifiques annuelles organisées par le Réseau ont constitué un espace structurant de réflexion sur la relation entre terminologie et politiques linguistiques, entendues comme un ensemble de choix relatifs à la gestion des langues dans des contextes de contact et de diversité. Au fil des années ces rencontres ont permis d'établir des références méthodologiques solides (cf. §4 et §5) pour le travail terminologique et de renforcer la reconnaissance institutionnelle de la terminologie dans l'espace roman.

Le rôle de la terminologie comme instrument de linguistique appliquée capable de répondre aux défis contemporains de la communication, se manifeste tant dans la production de données fiables que dans son usage comme levier de développement technologique. Les travaux menés au sein du Réseau associent linguistes, terminologues et spécialistes de domaine, dont la contribution est essentielle à la construction de savoirs terminologiques adéquats pour la communication de nature technique. La collaboration entre institutions, chercheurs et professionnels a, d'ailleurs, favorisé la production d'ouvrages terminologiques multilingues destinés aussi bien aux spécialistes qu'au grand public². Parallèlement, les utilisateurs quotidiens de ces ressources – traducteurs, experts de secteurs spécialisés, consultants, documentalistes, enseignants, journalistes et professionnels de la communication – en sont devenus des destinataires privilégiés.

Dans tous les domaines relevant des politiques linguistiques, de l'enseignement-apprentissage des langues étrangères, de la communication institutionnelle et professionnelle, de la diffusion et de la promotion des langues romanes spécialisées, ainsi que du multilinguisme, la recherche terminologique – à travers lexiques, glossaires, volumes thématiques et contributions scientifiques (Grimaldi, Zanola, 2021) – constitue un atout majeur (Zanola, 2018 ; Zanola, Resi, 2025). Ces recherches se caractérisent par une approche à la fois diachronique, synchronique et pluridisciplinaire, renforçant les liens avec la lexicologie et la lexicographie.

Les travaux de REALITER ont favorisé un échange terminologique et néologique particulièrement fécond entre les langues romanes, contribuant au développement de la terminologie dans cet espace linguistique. L'efficacité des contributions des membres italiens du Réseau met en évidence la richesse des relations entre terminologie et politique linguistique dans l'ensemble des pays de langues romanes.

Sur la période 2012-2023, les études terminologiques ont permis l'élaboration de principes méthodologiques communs, de cadres théoriques partagés et de pratiques issues de contextes internationaux, dont l'applicabilité à la langue italienne a été évaluée. Elles ont également contribué à la diffusion de documents de référence et à la production de glossaires multilingues, à partir de l'italien (par exemples, ceux des cols et des manches, de l'enseignement EMILE, des fibres textiles, de la ville intelligente, de l'énergie géothermique, des systèmes photovoltaïques), dans des domaines d'intérêt sociétal. Le rôle de la formation universitaire s'est révélé déterminant, favorisant l'émergence d'une nouvelle génération de chercheurs.

² La liste mise à jour des lexiques panlatins REALITER réalisés au sein des années est disponible sur le site du Réseau (URL : <<https://www.realiter.net/fr/lexiques-realiter>>, consulté le 19-11-2025).

Ce mouvement de réflexion et de recherche a contribué à la promotion de la diversité linguistique, en diffusant la terminologie spécialisée auprès des citoyens et de l'ensemble des usagers de la langue. Cette dynamique repose sur la qualité constante des travaux menés par REALITER à travers les centres universitaires et les chercheurs italiens, qui ont soutenu le développement d'analyses terminologiques d'un point de vue théorique, appliqué et communicationnel, souvent dans le cadre de recherches multilingues.

2. Les débuts de REALITER : fondation, objectifs et principes fondateurs

REALITER est né de la volonté de favoriser un développement harmonisé des langues romanes, en tenant compte de leur origine commune, de la proximité de leurs mécanismes de formation lexicale et de l'usage de formants similaires. Dès la première réunion fondatrice de 1993, le Réseau s'est doté d'objectifs clairs visant à structurer une coopération scientifique durable. Cinq axes prioritaires ont alors été définis : i) l'élaboration de principes méthodologiques communs pour la production de ressources terminologiques conjointes ; ii) la conduite de recherches partagées et le développement d'outils favorisant l'essor des langues néolatines ; iii) la réalisation de glossaires multilingues dans des domaines d'intérêt commun ; iv) la mutualisation des ressources documentaires, y compris les supports de formation ; v) enfin, la promotion de la formation réciproque par l'échange de formateurs, d'experts et d'étudiants.

Depuis plus de trente ans, REALITER fonctionne comme un réseau vivant, dynamique et productif, dans lequel toute activité terminologique est envisagée comme une contribution aux objectifs du Réseau et comme une réponse aux enjeux de complémentarité entre les langues romanes. Le travail plurilingue constitue ainsi un cadre privilégié pour l'élaboration de solutions terminologiques originales, qui ne pourraient émerger dans un contexte monolingue.

Dès sa fondation, REALITER s'est structuré comme un réseau coopératif dédié à la terminologie et à la néologie. Au-delà de ses objectifs opérationnels, il incarne également un projet éthique fondé sur des valeurs sociales fortes, à savoir la défense de la diversité linguistique, l'égalité entre les langues et la reconnaissance de leur capacité dénomminative. Toutes les langues romanes y sont considérées comme étant égales, indépendamment de leur poids politique, démographique ou économique.

Grâce au soutien institutionnel et à l'engagement de ses membres, REALITER a progressivement renforcé son ancrage scientifique et organisationnel dans l'espace international. La tenue régulière de réunions de travail, d'assemblées générales et de rencontres scientifiques ont contribué à l'institutionnalisation progressive du Réseau. Depuis 1993, ses activités se sont déployées dans de nombreuses villes – entre autres, Paris, Bruxelles, Lisbonne, Barcelone, Rome, Mexico, La Havane, Rio de Janeiro, Faro, Saint-Jacques-de-Compostelle, Nice, Québec, Bucarest, Bologne, Milan – témoignant du rayonnement international du Réseau.

Le rôle institutionnel de REALITER s'est également consolidé grâce à des partenariats structurés avec des organismes engagés dans la politique linguistique nationale.

Parmi ceux-ci figurent notamment la Délégation générale à la langue française et aux langues de France (DGLFLF), le Centre de Terminologie TERMCAT de Barcelone et l'Office québécois de la langue française (OQLF). Ces collaborations ont permis d'harmoniser les choix néologiques entre langues romanes, de renforcer l'implantation des termes face aux équivalents anglais et d'offrir aux instances nationales un panorama élargi de notions émergentes. Elles ont également contribué à la diffusion des terminologies auprès du grand public grâce à plusieurs bases de données telles que le Grand dictionnaire terminologique, FranceTerme, les lexiques du TERMCAT et TERMIUM Plus®. Ce n'est que dans ces cas que l'utilisation de la terminologie peut être réglementée ; dans tous les autres, l'impact se limite au statut de proposition d'usage préférentiel, issu d'un travail terminologique minutieux.

La réflexion scientifique menée au sein de REALITER repose sur une analyse fine des besoins terminologiques contemporains. L'accélération de la circulation des savoirs, la mobilité accrue des locuteurs et des biens, les situations de communication multilingues et les rapports de force entre langues ont mis en évidence l'importance stratégique de la relation entre terminologie et diversité linguistique. Ces évolutions ont conduit le Réseau à interroger en permanence ses méthodes de travail, ainsi que les processus de création terminologique.

Dans cette perspective, REALITER a très tôt intégré la question des politiques linguistiques à ses axes de réflexion, en les concevant comme un ensemble de décisions relatives à la gestion des langues dans des contextes de contact et de pluralité. Le Réseau a ainsi été créé, entre autres, pour affirmer la capacité dénomminative des langues romanes, prévenir la perte de domaines spécialisés et contribuer à la préservation de la diversité linguistique.

3. Les lexiques REALITER : objectifs, méthodologie et portée

Également appelés *Lexiques panlatins*, les lexiques REALITER constituent l'un des instruments centraux de l'activité du Réseau. Leur élaboration repose sur des critères méthodologiques communs et partagés, définis dans les Principes méthodologiques du travail terminologique, adoptés par les membres du Réseau à Lisbonne en 2000. Ces principes, qui structurent l'ensemble des productions terminologiques de REALITER, garantissent la cohérence scientifique, la comparabilité des données et la qualité des ressources produites.

Les lexiques prennent la forme de listes de termes relevant de domaines spécialisés, proposées dans l'ensemble des langues romanes représentées au sein du Réseau et généralement accompagnées d'équivalents anglais à titre référentiel. La forte présence d'anglicismes dans de nombreux secteurs spécialisés met en évidence l'influence croissante de l'anglais et confirme la nécessité d'outils favorisant la circulation de l'information dans le respect de la diversité linguistique. Dans cette perspective, les lexiques REALITER visent à encourager l'usage des dénominations en langues romanes et à contribuer à la régulation de la communication spécialisée.

Sans prétendre transformer immédiatement les usages linguistiques, ces ressources jouent un rôle fondamental en systématisant des terminologies multilingues. Élaborés à partir d'analyses terminologiques monolingues, ils participent à la limitation de l'emprunt direct à l'anglais et facilitent l'identification de solutions endogènes. Leur approche fondée sur la terminologie textuelle permet d'accéder à des discours de différents niveaux de spécialisation et de vérifier l'ancrage contextuel des dénominations, ce qui offre ainsi des propositions terminologiques adaptées aux usages réels.

Les lexiques REALITER constituent également un instrument essentiel pour le maintien du multilinguisme. En mettant en évidence les ressemblances morphosyntaxiques entre les langues romanes, ils contribuent à renforcer la créativité néologique et à consolider des paradigmes communs dans divers domaines spécialisés. Ils valorisent en outre la variation terminologique puisqu'ils ne se veulent pas des outils normatifs, mais des observatoires des usages. Cette dimension est particulièrement évidente dans les langues à large diffusion géographique, telles que le portugais, l'espagnol et le français, où la variation intercontinentale est prise en compte et analysée.

À travers une méthodologie fondée sur l'expertise, REALITER intègre différentes hypothèses de dénomination afin de proposer les solutions les plus pertinentes, en respectant des critères d'harmonisation adaptés à chaque domaine. De ce point de vue, les lexiques constituent de véritables actes de politique linguistique tant par leur conception que par les effets qu'ils visent à produire.

Enfin, tout en s'inscrivant dans la tradition méthodologique fondatrice du Réseau, l'avenir des lexiques REALITER appelle à une réflexion sur l'évolution des pratiques. Les transformations épistémologiques de la terminologie, les nouveaux besoins communicationnels et le développement des technologies numériques rendent nécessaire une actualisation des méthodes de travail, dans la continuité de l'esprit pionnier qui a présidé à leur élaboration.

Les lexiques REALITER s'inscrivent dans une perspective théorique et méthodologique qui vise l'harmonisation de la terminologie dans les langues néolatines, en tenant compte de leur origine commune et de la proximité de leurs mécanismes de formation lexicale. Depuis plus de trente ans, le Réseau œuvre à la mise à disposition de ressources terminologiques plurilingues authentiques, actuelles et fiables, constituant un patrimoine scientifique de grande valeur pour des applications variées, qui vont de la traduction et de l'interprétation à l'analyse terminologique et à la transmission des connaissances spécialisées (Gilardoni, 2011 ; Zanola, 2014b).

Les domaines couverts par ces lexiques sont nombreux et diversifiés – de l'informatique à la médecine, du commerce aux biotechnologies, de la mode aux énergies renouvelables –, ce qui témoigne de leur large portée socioprofessionnelle. Ils s'adressent non seulement aux traducteurs spécialisés, mais également aux experts de domaine, consultants linguistiques, journalistes, rédacteurs, réviseurs et à un public plus large intéressé par une communication plurilingue précise et appropriée (Dankova, Calvi, 2020). Les lexiques REALITER constituent ainsi un répertoire de données essentiel pour les professionnels impliqués dans la production, la gestion et la diffusion des savoirs spécialisés.

Sur le plan méthodologique, leur élaboration repose sur des principes fondamentaux partagés par l'ensemble du Réseau. Ceux-ci incluent : i) le respect de la diversité et de l'égalité linguistiques, chaque langue et variante disposant d'un statut équivalent dans les projets ; ii) l'adoption d'une approche variationniste, inspirée de la socioterminologie, visant à refléter la diversité des usages et des traditions terminologiques ; iii) une approche systémique fondée sur la structuration conceptuelle et la définition rigoureuse des notions. Le travail collaboratif constitue un pilier central de cette démarche, associant à chaque étape des terminologues de différents pays, des spécialistes des domaines concernés et des utilisateurs des ressources. Par ailleurs, les lexiques REALITER répondent à des critères de qualité centrés sur l'utilisateur : accessibilité, actualité et fiabilité terminologique (REALITER, 2000 ; Zanola, 2012). Ces exigences garantissent l'adéquation des ressources aux besoins communicationnels réels, tant dans les contextes professionnels qu'institutionnels.

Les travaux sont systématiquement réalisés en collaboration avec des institutions publiques, des centres terminologiques, des universités et des centres de recherche, ce qui assure leur légitimité scientifique et leur ancrage social. Chaque lexique retrace les étapes de sa construction, répondant à la fois aux besoins des spécialistes – qui voient leur savoir transmis dans plusieurs langues –, du grand public en phase d'initiation à un domaine et des traducteurs à la recherche de contextes d'usage fiables. Ainsi, fondés sur des approches variationniste et systémique, les lexiques REALITER contribuent concrètement au développement harmonieux des langues néolatines, tout en affirmant l'égalité de statut entre toutes les langues et leurs variantes du Réseau.

4. Les Journées scientifiques REALITER (2012-2016) : thématiques et apports

Entre 2012 et 2016, les Journées scientifiques de REALITER ont constitué un espace privilégié de réflexion collective sur le rôle stratégique de la terminologie dans les politiques linguistiques contemporaines, dans un contexte marqué par la mondialisation des échanges, la montée en puissance de l'anglais comme *lingua franca* et la nécessité de préserver la diversité linguistique. Ces rencontres ont permis de mettre en relation des perspectives théoriques, des analyses institutionnelles et des études de terrain, tout en renforçant le positionnement du Réseau comme acteur majeur de la terminologie panlatine (voir les descriptifs des programmes dans la section consacrée aux Journées scientifiques dans le site <www.realiter.net>).

La Journée 2012 *Terminologies et politiques linguistiques* (Università Cattolica del Sacro Cuore, Milan), consacrée aux relations entre terminologie et politiques linguistiques, a posé les bases du débat en examinant les interventions des États, des gouvernements, des organisations internationales et des institutions publiques dans la diffusion et la promotion des terminologies néolatines. Les discussions ont mis en lumière le lien étroit entre langue et développement technique, scientifique, productif et culturel des sociétés, en soulignant la valeur patrimoniale des langues spécialisées. Cette Journée (Zanola, Conceição, Guasco, 2017) a confirmé que la terminologie constitue un levier essen-

tiel de souveraineté linguistique et un instrument stratégique des politiques publiques. Dans le prolongement de ces réflexions, la Journée 2013 (DGLFLF, Paris), intitulée *La terminologie panlatine dans les politiques linguistiques*, a approfondi la question de la diffusion réelle des terminologies néolatines dans l'espace public. Les débats (Depecker, Zanola, 2017) ont porté sur les tensions entre langue spécialisée et usage courant, sur les écarts entre discours formalisé et pratiques médiatiques, ainsi que sur la concurrence exercée par l'anglais comme langue globale. Une attention particulière a été accordée à des secteurs clés tels que l'économie, la finance, la santé, l'éducation, les technologies et les médias, afin d'évaluer l'impact réel des terminologies romanes dans la communication institutionnelle et médiatisée.

La Journée 2014 (Academia de Studii Economice din Bucuresti), placée sous le thème *La terminologie panlatine au service du citoyen*, a marqué un tournant vers une approche résolument communicationnelle et sociale : les travaux ont interrogé la responsabilité des institutions dans la transmission claire et accessible de l'information spécialisée aux citoyens. Les échanges ont notamment porté sur la communication administrative, la diffusion des normes sanitaires et le partage des savoirs techniques. Deux axes majeurs se sont dégagés : d'une part, le rôle de la terminologie dans la promotion du multilinguisme ; d'autre part, l'articulation entre création néologique, normalisation et intelligibilité pour le grand public. Cette Journée a ainsi affirmé la dimension civique de la terminologie.

La Journée 2015 (Académie Royale de Belgique, Bruxelles), consacrée à *Terminologie et multilinguisme*, a renforcé la réflexion sur les enjeux méthodologiques et pratiques liés à la gestion des ressources terminologiques multilingues. Le multilinguisme y a été analysé comme principe fondateur de REALITER et comme objectif central des politiques linguistiques européennes et internationales. Les communications ont abordé la place de la terminologie dans la promotion de la diversité linguistique, les contraintes méthodologiques induites par le travail multilingue, ainsi que les stratégies de diffusion et d'implantation des ressources terminologiques. Cette Journée (Toma, Zanola, 2016) a confirmé la nécessité d'approches interdisciplinaires associant linguistes, terminologues, spécialistes de domaine et décideurs institutionnels.

Enfin, la Journée 2016 (Universidade de Santiago de Compostela), dédiée à *Terminologie et normalisation*, a permis de clarifier les différentes acceptions du concept de normalisation. Celui-ci a été compris dans les deux sens que ce mot possède dans plusieurs langues romanes, à savoir, d'une part, l'extension de l'usage de la langue dans les contextes où cela est jugé nécessaire et, d'autre part, la standardisation des langues spécialisées. Depuis sa création, REALITER a accordé une attention prioritaire à ces deux significations, à la promotion du multilinguisme et de la diversité linguistique, ainsi qu'au développement harmonieux des langues néolatines. Les débats (Nuñez Singala, 2018) ont porté sur les procédures de normalisation, la gestion de la variation terminologique et l'intégration des choix terminologiques dans les politiques linguistiques.

5. Les Journées scientifiques REALITER (2017-2022) : un élargissement progressif des thématiques de recherche

Entre 2017 et 2022, les Journées scientifiques REALITER ont poursuivi et renouvelé la réflexion engagée lors des rencontres précédentes, en approfondissant les relations entre terminologie, politiques linguistiques et transformations sociétales contemporaines. Cette période se caractérise par un élargissement progressif des thématiques, intégrant les enjeux de l'internationalisation, de la médiation linguistique, de l'interculturalité, de la communication sociale et des dynamiques de recherche.

La Journée scientifique, organisée à Barcelone en 2017 – *Terminologie pour la normalisation et terminologie pour l'internationalisation* –, s'est inscrite dans la continuité directe des réflexions menées lors de la Journée de l'année précédente, consacrée à la normalisation linguistique. Elle a examiné les convergences et tensions entre deux dynamiques majeures : la normalisation terminologique, d'une part, et l'internationalisation croissante des institutions académiques et professionnelles, de l'autre. Les débats ont porté sur l'impact de ces processus sur les politiques linguistiques des universités et des organismes où interviennent les membres du Réseau, ainsi que sur les nouvelles exigences en matière de promotion du multilinguisme. Une attention particulière a été accordée à l'évolution des profils professionnels en terminologie, aux compétences émergentes, aux ressources collaboratives en ligne et aux innovations méthodologiques.

La Journée scientifique REALITER, tenue à Paris en 2018 intitulée *Terminologie spontanée et terminologie aménagée*, s'est concentrée sur l'opposition et surtout les interactions entre terminologie spontanée et terminologie aménagée. Les échanges ont interrogé les processus de création terminologique issus des pratiques professionnelles quotidiennes, en les comparant aux démarches institutionnelles de normalisation et d'aménagement linguistique. Les communications (DGLFLF, 2019) ont montré que ces deux modes de production terminologique ne sont ni antagonistes ni indépendants, mais s'inscrivent dans des dynamiques complémentaires. Des études de cas ont illustré leurs convergences et divergences selon les domaines spécialisés et les contextes nationaux, mettant en lumière le rôle des politiques linguistiques dans ces interactions. Une réflexion spécifique a également porté sur les modes de diffusion, les publics ciblés et les usages réels des terminologies produites, contribuant à une compréhension plus fine des mécanismes d'implantation terminologique.

La Journée scientifique REALITER, organisée à l'Universidade do Algarve en 2019 sous le titre *Terminologie et médiation linguistique*, a placé au centre du débat la notion de médiation linguistique. Dans un contexte de mobilité accrue et de circulation intensive des savoirs, la terminologie y a été envisagée comme condition essentielle de toute médiation communicative. Lors de la journée les débats ont exploré les multiples dimensions de la médiation : traduction, interprétation, médiation culturelle, révision éditoriale, enseignement, localisation, sous-titrage et vulgarisation scientifique. La terminologie y apparaît comme un outil permettant de transformer, d'adapter et de diffuser les concepts auprès de publics différenciés. Cette rencontre (Conceição, Zanola, 2020) a donc mis en évidence le rôle fondamental de la rigueur terminologique dans

les processus d'inclusion sociale, en garantissant des conditions d'équité communicative entre les utilisateurs.

La Journée scientifique, organisée à distance par l'Alma Mater Studiorum-Université di Bologna en 2020 sur *Terminologie et interculturalité*, a approfondi la dimension culturelle et interculturelle de la pratique terminologique. Les débats ont porté sur les relations entre terminologie, analyse du discours, traduction interlinguistique et communication interculturelle, dans un contexte marqué par la révolution numérique. Les communications ont souligné les défis liés au croisement des langues et cultures dans la construction discursive des terminologies. Les axes de réflexion ont abordé les résistances à l'altérité linguistique, les phénomènes d'hybridation, les politiques linguistiques face à la diversité culturelle, ainsi que l'impact des technologies de traitement automatique des langues. Cette Journée a permis de poser les bases d'une réflexion théorique plus systématique sur le lien entre terminologie et interculturalité, en s'appuyant notamment sur l'expérience des lexiques REALITER.

La terminologie dans la dimension sociale de la communication a été le sujet approfondi par la Journée scientifique REALITER, organisée en modalité virtuelle en 2021 entre Beyrouth et Porto Alegre. Dans un contexte marqué par la crise sanitaire mondiale, la transformation numérique et l'essor de l'intelligence artificielle, les communications ont interrogé le rôle des termes dans le transfert des connaissances au XXI^e siècle. Les interventions ont porté sur des domaines clés tels que la santé publique, les technologies, la traçabilité des données en ligne, le droit, l'agroalimentaire et la formation. Cette rencontre a mis en évidence l'importance de la précision terminologique dans la communication spécialisée, tout en soulignant la dimension culturelle, juridique et médicale de la médiation linguistique. Elle a confirmé que la terminologie joue un rôle central dans la structuration sociale des discours et dans la gestion des crises contemporaines.

La Journée scientifique REALITER, organisée à Bologne en 2022 en collaboration avec le Colloque annuel de l'Associazione Italiana per la Terminologia-Ass.I.Term, a été consacrée aux dynamiques actuelles de la recherche en terminologie (*Les projets de recherche en terminologie*). Les échanges (Grimaldi, Marzi, Puccini, Zanola, Zollo, 2023) ont porté sur les projets financés par des programmes nationaux et internationaux (ERC, Marie Skłodowska-Curie, appels ministériels, dispositifs Jeunes Chercheurs), ainsi que par des fondations publiques et privées. Cette Journée a offert une vitrine des recherches en cours ou récemment achevées, permettant aux doctorants et chercheurs de tout niveau académique de partager leurs résultats, leurs méthodologies et leurs perspectives. Elle a mis en lumière la vitalité du champ terminologique, sa structuration institutionnelle croissante et son inscription dans les grands programmes scientifiques internationaux.

Sur la période 2017-2022, les Journées scientifiques REALITER témoignent d'un élargissement progressif des champs d'analyse : de la normalisation à l'internationalisation, de la production terminologique à la médiation sociale, de l'interculturalité aux dynamiques de recherche. Elles confirment le positionnement du Réseau comme espace de réflexion scientifique de référence, capable d'anticiper les évolutions méthodologiques, technologiques et sociétales affectant la terminologie. Cette dynamique a contribué à

renforcer le rôle de REALITER comme acteur central de la promotion du multilinguisme et de la diversité linguistique dans l'espace roman.

6. Conclusion : de nouvelles orientations de recherche

Dans la continuité de son histoire scientifique, REALITER inscrit désormais ses travaux dans une perspective résolument interdisciplinaire. Depuis son affirmation comme discipline autonome, la terminologie s'est, en effet, construite à la croisée de plusieurs champs du savoir – linguistique, logique, ontologie, informatique, sciences des objets –, ce qui lui permet d'appréhender les réalités spécialisées tant dans leur état actuel que dans leur évolution diachronique, à la fois conceptuelle et linguistique. Conçue comme une méthode de transmission des connaissances par des moyens linguistiques, la terminologie a favorisé, depuis la seconde moitié du XX^e siècle, un dialogue constant entre linguistes et spécialistes de domaine, dialogue qui demeure aujourd'hui au cœur des projets portés par le Réseau.

Face aux défis posés par les sociétés contemporaines, tels que, entre autres, l'accélération des échanges, les transformations technologiques et la complexification des savoirs, la recherche terminologique actuelle encourage plus que jamais le partage des connaissances entre disciplines afin d'élaborer des réponses méthodologiques et appliquées innovantes. C'est dans cette perspective que s'inscrit la Journée scientifique de REALITER, organisée à Paris en octobre 2023, qui a dressé un état des lieux des dynamiques de recherche en cours et a ouvert un espace de dialogue autour du thème *Terminologie et interdisciplinarité*, au moment symbolique du trentième anniversaire du Réseau.

De cette rencontre de nouvelles pistes de travail se sont dégagées, renforcées par les synergies entre terminologie et autres champs disciplinaires. Plusieurs axes de réflexions ont été proposés pour nourrir le débat au sein du Réseau :

- le renouvellement des approches méthodologiques en recherche terminologique ;
- les relations entre terminologie et traduction spécialisée ;
- la place de la terminologie dans l'enseignement des langues de spécialité ;
- les liens entre terminologie, médiation et communication spécialisée ;
- l'articulation entre terminologies et politiques linguistiques.

Ces orientations témoignent de la volonté de REALITER de consolider son rôle de laboratoire interdisciplinaire, capable d'intégrer des perspectives théoriques, didactiques, technologiques et sociopolitiques. Elles ouvrent des perspectives de recherche fécondes pour les années à venir, confirmant la capacité du Réseau à se renouveler et à anticiper les évolutions du champ terminologique.

Références bibliographiques

Conceição M.C., Zanola M.T. (eds.), *Terminologia e mediação linguística: métodos, práticas e atividades*, Faro, Universidade do Algarve Editora, 2020.

- Dankova K., Calvi S., *Repertori terminologici multilingui fra normatività e uso nella comunicazione istituzionale e professionale*, in C. Marras et al. (a cura di), *La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica. Atti del IX Convegno Annuale AIUCD*, Associazione per l'Informatica Umanistica e la Cultura Digitale, 2020, pp. 98-105. URL : <<https://amsacta.unibo.it/id/eprint/6316/>>, consulté le 19-11-2025.
- Délégation générale à la langue française et aux langues de France (éds.), *Convergences et divergences dans la pratique terminologique. De la terminologie spontanée à la terminologie aménagée*, Paris, Délégation générale à la langue française et aux langues de France, 2019.
- Depecker L., Zanola M.T. (éds.), *Les vingt ans de REALITER. La terminologie panlatine dans les politiques linguistiques*, Milano, EDUCatt, 2017.
- Gilardoni S., *I lessici della Rete Panlatina di Terminologia*, in M.T. Zanola, M.F. Bonadonna (a cura di), *Terminologie specialistiche e prodotti terminologici*, Milano, EDUCatt, 2011, pp. 101-112.
- Grimaldi C., *Terminology Planning and Language Policies across the Romance-speaking Area: the Activities of REALITER*, in R. Resi, F. Steurs (eds.), *Handbook of Terminology. Terminology planning in Europe*, Amsterdam/Philadelphia, John Benjamins, 2025, pp. 32-40.
- Grimaldi C., Marzi E., Puccini P., Zanola M.T., Zollo S.D. (a cura di), *Terminologia e interculturalità. Problematiche e prospettive*, Città di Castello, I Libri di Emil, 2023.
- Grimaldi C., Zanola M.T. (a cura di), *Terminologie e vocabolari. Lessici specialistici e tesauri, glossari e dizionari*, Firenze, Firenze University Press, 2021.
- Núñez Singala Manuel (ed.), *Terminoloxía e normalización. Actas XII Xornada Científica REALITER*, Santiago de Compostela, Servizo de Publicacións e Intercambio Científico da USC, 2018.
- REALITER, *Principi metodologici del lavoro terminologico*, 2000. URL : <<https://www.realiter.net/wp-content/uploads/2013/06/Principi-metodologici-del-lavoro-terminologico.pdf>>, consulté le 19-11-2025.
- Toma A., Zanola M.T. (éds.), *Terminologie et multilinguisme : objectifs, méthodologiques et pratiques*, numéro spécial 13/1 « Diversité et identité culturelle en Europe », Bucarest, Editura Muzeul Literaturii Române, 2016.
- Zanola, M.T., *Costruire un glossario: la terminologia dei sistemi fotovoltaici*, Milano, Vita e Pensiero, 2012.
- Zanola M.T., *Le réseau Realiter, un acteur du plurilinguisme*, in « Plaisance », 32, 2014a, pp. 149-165.
- Zanola M.T., *Les principes méthodologiques des lexiques REALITER*, in « Les Cahiers du dictionnaire », 6, 2014b, pp. 189-202.
- Zanola M.T., Conceição M.C., Guasco P. (a cura di), *Terminologie e politiche linguistiche*, Milano, EDUCatt, 2017.
- Zanola M.T., *Che cos'è la terminologia*, Roma, Carocci, 2018.
- Zanola M.T., Resi R., *From terminological neology to terminology planning for corporate and professional initiatives. The Italian case*, in R. Resi, F. Steurs (eds.), *Handbook of Terminology. Terminology planning in Europe*, Amsterdam/Philadelphia, John Benjamins, 2025, pp. 616-637.

A terminologia da COVID Longa no Português Brasileiro e aspectos de sua variação

IEDA MARIA ALVES, BEATRIZ CURTI-CONTESSOTO, MÁRCIA DE SOUZA LUZ-FREITAS, ANA MARIA RIBEIRO DE JESUS, LUCIMARA ALVES DA CONCEIÇÃO COSTA

1. Introdução

A Organização Mundial da Saúde (OMS) declarou, em 05 de maio de 2023, em Genebra, Suíça, o fim da Emergência de Saúde Pública de Importância Internacional (ESPII) referente à COVID-19. Essa decisão foi tomada pelo diretor-geral da OMS, Tedros Adhanom, atendendo à recomendação do Comitê de Emergência da instituição encarregado de analisar periodicamente a situação dessa pandemia. Essa declaração foi motivada em função de os membros do Comitê destacarem a tendência de queda nas mortes por COVID-19, o declínio nas hospitalizações e internações em unidades de terapia intensiva relacionadas à doença, bem como os altos níveis de imunidade da população ao SARS-CoV-2, coronavírus causador dessa enfermidade.

A COVID-19, no entanto, tem deixado muitas sequelas dentre os que foram atingidos pela doença. Essas sequelas, que perduram por um período bastante diversificado dentre os infectados, têm produzido formas variantes que sugerem diferentes concepções relativas à doença e às suas consequências. Em função disso, este texto objetiva analisar a composição dessas variantes terminológicas, revelando as características que envolvem cada denominação e enfatizando suas diferentes formações e aspectos semânticos, assim como as diferentes perspectivas que cada uma estabelece. Os dados analisados são extraídos do *corpus* de estudo constituído no âmbito do projeto intitulado “Estudo e Divulgação da Terminologia da COVID-19”, que se encontra em desenvolvimento na Universidade de São Paulo (USP). Este trabalho, destinado a um público amplo e não especializado na área médica, está seguindo os princípios de uma corrente, nacional e internacional, de utilização de uma linguagem mais simples e mais facilmente compreendida por usuários não especializados na área de especialidade em foco e denominada Plain Language Association International (PLAIN).

Nas seções seguintes, explicamos inicialmente como os *corpora* foram constituídos e como se observa a produtividade lexical dos termos denominativos do fenômeno da *covid longa* nesses *corpora*. Em seguida, descrevemos a análise da composição dos termos e apresentamos o verbete *covid longa* e nossas considerações finais. Expomos, por fim, as referências que embasaram o estudo.

2. Constituição dos *corpora* e produtividade relativa aos termos em foco

O projeto supramencionado visa à elaboração de um dicionário sobre os termos da COVID-19 em português brasileiro destinado a um público com pouco letramento. Nesse estudo, dois *corpora* foram constituídos: o *Corpus Oficial* (CO), composto de textos institucionais internacionais (comunicados em língua portuguesa da Organização Mundial da Saúde, da Organização Panamericana de Saúde) e brasileiros (Instituto Butantã, Fiocruz, Fapesp, dentre outros); e o *Corpus Jornalístico* (CJ), representado por textos extraídos de jornais de grande circulação no Brasil (Folha de S. Paulo, O Globo, O Estado de S. Paulo) coletados a partir de 2020.

A compilação desses *corpora* seguiu uma metodologia baseada na *web* como *corpus*. Desse modo, foram utilizadas as ferramentas *BootCat Bootstrap Corpora and Terms from the Web* versão 1.21 (Zanchetta *et al.*, 2011) e *AntConc* (Anthony, 2020). A primeira nos auxiliou na busca por materiais sobre o tema em questão que foram disponibilizados nos *sites* das fontes mencionadas. Feita essa compilação, esses documentos, em formato *txt*, foram inseridos no programa *AntConc*. Nesse programa, foram criadas diferentes listas (palavras-chave, *wordlist*, *clusters* e concordância) para cada *corpus* de estudo. A partir da análise dessas listas, foram identificados os termos presentes no CO e no CJ.

Uma vez que os textos especializados são o habitat natural das terminologias, conforme enfatizado por Krieger (2011), neles pode ser observado o comportamento dos termos e, consequentemente, diversos fenômenos linguísticos, como a variação terminológica, definida por Freixa (2002) como um fenômeno em que um mesmo conceito apresenta diferentes denominações. Para a autora, o fenômeno da variação pode ser explicado de maneira abrangente, considerando vários aspectos linguísticos, a exemplo da formação morfológica dos termos:

Freixa (2006) proposes that this phenomenon in Terminology should be called denominative variation because synonymy is usually understood as being between different lexemes, whereas the term denominative variation covers any other type of variation in denominations – from graphical, orthographical, morphological and morphosyntactic variations and reductions to lexical variations – so long as they can be considered denominations, in other words, not paraphrases or occasional occurrences but lexicalised forms with stability and consensus among the users of units in a specialised domain. (Freixa, 2022: 400)

Os nossos *corpora* mostram uma forte presença da variação terminológica no contexto específico da COVID-19, algo que também tem sido notado em outras línguas ao redor do mundo, como salienta a mencionada autora:

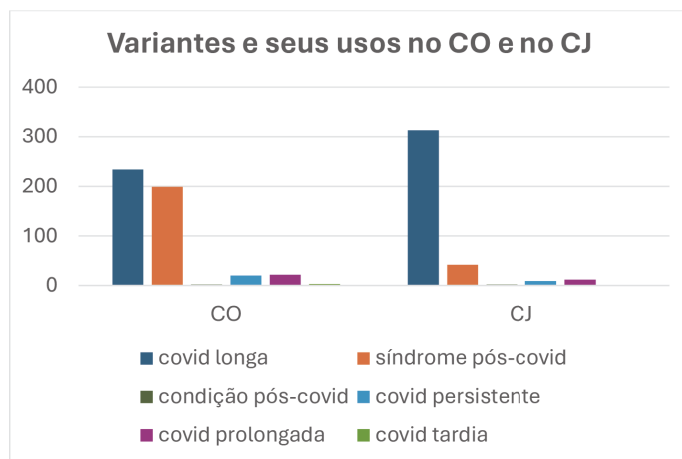
Like all other crises, whether political, socioeconomic, medical or environmental, the global pandemic caused by the SARS-CoV-2 virus has led to the emergence of new terms that speakers use. While some of them already existed in the strictly epidemiological sphere, others have since arisen to denominate new contexts, and many of them have been disseminated in non-specialized communications. All

countries and all languages have had to face up to these new denominative needs in order to find ways of ensuring that terms related to the pandemic can be adapted to the diversity of communication scenarios in today's society. Within this context, terminological variation has played an important role. (Freixa, 2022: 399)

Essa variação também foi constatada em diferentes contextos relacionados à COVID-19 em português (cf. Svobodová, 2021). Esse fato também foi percebido em nossos *corpora*, que têm características diferentes. O CO e o CJ são *corpora* que reúnem textos brasileiros sobre a pandemia em dois tipos de veículos diferentes: o primeiro de caráter mais técnico e oficial e o segundo de divulgação, em linguagem mais acessível. Embora compreendam discursos distintos sobre a COVID-19, tanto o CO quanto o CJ apresentam variantes terminológicas.

Este texto foca na análise da variação dos termos denominativos da doença covid longa que ocorrem no CO e no CJ. O termo *covid longa* refere-se à manifestação da doença caracterizada por um conjunto de sintomas observados em algumas pessoas que, após terem sido infectadas pelo novo coronavírus, não mais apresentam o vírus¹. Essas sequelas têm sido diferentemente nomeadas por termos como *covid persistente*, *covid prolongada*, *covid tardia*, *pós-covid*, *síndrome da covid longa* e *síndrome pós-covid-19*. A ocorrência desses termos difere entre o CO e o CJ, estando distribuída tal como mostra o gráfico apresentado a seguir:

Gráfico 1 - Variantes terminológicas relativas à covid longa e sua ocorrência no CO e no CJ



O Gráfico 1 mostra que o termo *covid longa* apresenta maior frequência nos dois *corpora*. Já o termo *síndrome pós-covid*, embora ocorra no CO e no CJ, é mais empregado no *corpus* oficial. Observa-se ainda baixa ocorrência dos demais termos variantes em ambos os *corpora*.

¹ Conforme a definição do termo *covid longa*, registrado no dicionário em elaboração no âmbito do projeto supramencionado.

3. Análise da composição dos termos

Mostra-se predominante, nos dois *corpora*, a formação sintagmática constituída por substantivo e adjetivo, como se observa nos termos *covid-19 longa*, *covid longa*, *covid prolongada*, *covid persistente* e *covid tardia*.

Essas variantes, embora denominem o mesmo fenômeno, apresentam diferentes nuances conceituais que são expressas por distintos adjetivos. *Longa* e *prolongada* indicam a duração que se estende por um período maior do que o esperado. Já o adjetivo que integra o termo *covid persistente* implica a insistência da permanência dos sintomas no organismo da pessoa infectada. *Covid tardia*, por sua vez, corresponde aos sintomas que surgem após o período de infecção. Essa variação lexical deixa transparecer diferentes nuances a respeito da percepção com relação à durabilidade dos sintomas.

Os adjetivos são comumente classificados como qualificativos e classificadores (cf. Neves, 2000: 184-219). No caso da terminologia apresentada, todos os adjetivos atuam como classificadores, na medida em que especificam o conceito do termo substantival ao qual se ligam no sintagma.

Diferentemente das formações em que o adjetivo exerce o papel de indicar a durabilidade, são também observadas construções em que o recurso morfológico é a derivação prefixal, em que o prefixo *pós-* indica um período posterior à doença, participando da formação das variantes *condição pós-covid-19*, *condição pós-covid*, *síndrome pós-covid-19* e *síndrome pós-covid*. O emprego de *condição* e *síndrome* revela a instabilidade semântica do fenômeno, uma vez que esses termos se referem a conceitos distintos: *condição* indica uma circunstância persistente relativa à saúde de uma pessoa e *síndrome* faz alusão a um conjunto de sinais e sintomas observáveis em vários processos patológicos.

De maneira distinta das variantes em cuja formação transparece o prefixo *pós-*, destaca-se ainda a presença da variante *síndrome da covid longa*, em que o caráter durativo da doença é expresso pelo adjetivo *longa*, o que caracteriza uma variação morfossintática.

Todas as variantes observadas apresentam, em sua formação, a designação da doença por meio das ocorrências *covid-19* ou *covid*. A coexistência dessas duas designações indica uma variação gráfica, em que na segunda forma se nota um processo de redução.

4. Apresentação do verbete *covid longa*

No dicionário em elaboração, os verbetes compreendem, quando pertinente, os seguintes campos: termo, classe e gênero gramatical, equivalente inglês, variante(s), definição, nota, contexto(s) de uso.

Assim, o verbete *covid longa* apresenta a seguinte configuração:

covid longa sf

En: long COVID

Também se diz: *covid-19 longa*, *covid persistente*, *covid-19 persistente*, *covid prolongada*, *covid tardia*, *pós-covid*, *condição pós-covid-19*, *síndrome da covid longa*, *síndrome pós-covid-19*.

Manifestação da doença caracterizada por um conjunto de sintomas observados em algumas pessoas que, após terem sido infectadas pelo novo coronavírus, não mais apresentam o vírus.

Nota: Segundo diferentes estudos realizados, seguem abaixo os sintomas mais prevalentes em pessoas com histórico de COVID-19 e hospitalização, com internação em enfermaria e/ou unidade de terapia intensiva (UTI), condizentes com o que temos observado no Serviço de Reabilitação: Fadiga ou fraqueza muscular (63%). Fadiga (29,9% e 38,1% dos pacientes hospitalizados, sem e com necessidade de ventilação mecânica, respectivamente). Fraqueza Muscular – em pacientes internados em UTI a perda de força muscular foi de até 3,7% ao dia (sendo que a perda muscular usualmente observada em pacientes críticos se situa entre 0,7 a 1,5% ao dia); Dispneia (34,5% e 45,1% dos pacientes hospitalizados, em enfermaria e UTI, respectivamente); (13,1% e 34,2%, dos pacientes hospitalizados, sem e com necessidade de ventilação mecânica, respectivamente); Necessidade de meio auxiliar de locomoção (28,8% e 39,8% dos pacientes hospitalizados, em enfermaria e UTI, respectivamente); Algum grau de dependência para Atividades de Vida Diária (27,3% e 38,9%, dos pacientes hospitalizados, em enfermaria e UTI, respectivamente); Algum grau de dependência para Atividades Instrumentais de Vida Diária (74,5% e 84,6% dos pacientes hospitalizados, em enfermaria e UTI, respectivamente); Dor (27,1% e 33,9% dos pacientes hospitalizados, em enfermaria e UTI, respectivamente); Distúrbios do sono (26%)⁴ e (53,6%); Disfunção Cognitiva (38,4%); Ansiedade (31,4%); Depressão (20,6%); Estresse pós-traumático (14,2%).

A ocorrência destes sintomas persistentes vem sendo denominada de **Síndrome Pós-COVID-19**, “**COVID longa**” e, também, de Síndrome Pós-Cuidados de Terapia Intensiva (Post-intensive Care Syndrome – PICS), nos casos mais graves. HSL, 23-07-2021, URL: <<https://hospitalsiriolibanes.org.br/blog/acontecenosiriolibanes/saiba-mais-sobre-o-programa-de-reabilitacao-pos-covid-19>>

A **covid-19 persistente**, na qual pacientes enfrentam sintomas por meses, pode estar afetando as pessoas de quatro maneiras diferentes de uma só vez. E isso pode explicar porque pacientes com esses sintomas persistentes não estão recebendo tratamento adequado. Além disso, pode haver um enorme impacto psicológico nos pacientes com essa **covid-19 longa**, aponta uma revisão de estudos feita pelo Instituto Nacional de Pesquisa em Saúde do Reino Unido.

UOL, 15-10-2020, URL: <<https://www.uol.com.br/vivabem/noticias/bbc/2020/10/15/covid-19-persistente-se-manifesta-com-sintomas-de-4-sindromes-diferentes.htm>>

Além das grandes repercussões emocionais e traumáticas advindas da doença, suas manifestações clínicas ultrapassam a fase aguda e são chamadas de **Covid tardia** ou **Covid longa**.

FSP, 22-02-2022, URL: <<https://www1.folha.uol.com.br/opiniao/2022/02/covid-longa-e-a-atencao-a-saude-mental.shtml>>

Náusea, tosse, suor, zumbido no ouvido e problemas de sono afetam a vida de Eduarda Norat, de 22 anos. Três meses depois de ter se curado da Covid-19, Eduarda sofre com alguns dos 55 sintomas mais conhecidos [...] de uma doença que vem sendo chamada de “**Síndrome Pós-Covid**”, “**Covid longa**”, “**Covid persistente**” ou “**Covid prolongada**”. (G, 13-02-2021, URL: <<https://g1.globo.com/bemestar/coronavirus/noticia/2021/02/13/sintomas-da-covid-longa-attingem-ate-80percent-dos-infectados-entenda-o-que-e-e-conheca-55-efeitos-de-longo-prazo.ghml>>)

A **condição pós-Covid-19** ocorre em indivíduos com histórico de infecção provável ou confirmada por SARS-CoV-2, geralmente 3 meses após o início da Covid-19, com sintomas que duram pelo menos 2 meses e não podem ser explicados por um diagnóstico alternativo. Os sintomas comuns incluem fadiga, falta de ar, disfunção cognitiva, mas também outros que geralmente têm impacto no funcionamento diário. Os sintomas podem ser um novo início, após a recuperação inicial de um episódio agudo de Covid-19, ou persistir desde a doença inicial. Os sintomas também podem flutuar ou recair ao longo do tempo. Fiocruz, 06-2022, URL: <<https://informe.ensp.fiocruz.br/assets/anexos/6cb-02116ca259e904ba0a7cfffb993f6.PDF>>

O **pós-COVID** afeta várias partes do corpo. Há sequelas sensoriais [alteração de paladar e olfato], musculares [fadiga], circulatórias e inflamatórias, principalmente no sistema respiratório. Há ainda outras sequelas comuns a várias infecções, como zumbido no ouvido e parestesia [dormência] facial, além de lesões na pele para pacientes acamados por longos períodos. FAPESP, 02-05-2022, URL: <<https://agencia.fapesp.br/terapias-baseadas-em-luz-apresentam-bons-resultados-na-recuperacao-de-sequelas-pos-covid/38516>>

5. Considerações finais

Os *corpora* coletados têm-nos revelado que diferentes sequelas podem manifestar-se após o período de infecção pelo vírus, o que tem gerado diferentes formas neológicas variantes referentes a esse fenômeno. Procuramos apresentar, nesta exposição, as características que envolvem cada denominação, analisando suas diferentes formações e aspectos variacionais, assim como as diferentes perspectivas que cada uma estabelece.

Pudemos constatar, como salienta Freixa (2002), que o fenômeno da variação pode ser explicado de maneira abrangente, manifestando-se nos aspectos gráficos, ortográficos, morfológicos, sintáticos, semânticos e lexicais.

Foi observada ainda a relevância exercida pelos adjetivos na constituição das denominações referentes ao prolongamento dos sintomas da doença, o que confirma estudos realizados (Guilbert, 1963: 270) sobre a importância dessa classe gramatical na formação de novos termos.

Referências bibliográficas

- Anthony L., *AntConc* (Version 3.5.9) [Computer Software], Tokyo, Waseda University, 2020.
- Freixa J., *La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'era de mediambient*, Barcelona, Institut Universitari de Lingüística Aplicada/Universitat Pompeu Fabra, 2002.
- Freixa J., *Causes of terminological variation*, in M.-C. L'Homme, P. Faber (eds.), *Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized knowledge*, Amsterdam/Philadelphia, John Benjamins, 2022, pp. 399-420.
- Guilbert L., *La formation du vocabulaire de l'aviation*, Paris, Larousse, 1965.
- Krieger M.G., *Terminologia: uma entrevista com Maria da Graça Krieger*, in « ReVEL », 9-17, 2011. URL: <https://www.revel.inf.br/files/entrevistas/revel_17_entrevista_maria_graca_krieger.pdf>, consultado em 19-11-2025.
- Neves M.H. de M., *Gramática de usos do português*, São Paulo, Editora da UNESP, 2000.
- Svobodová I., *Género gramatical de COVID-xenismos*, in « Études romanes de Brno », 42-1, 2021, pp. 95-122.
- Zanchetta E. et al., *Corpora for the masses: the BootCaT front-end*, in *Proceedings of the Corpus Linguistics Conference*, Birmingham, University of Birmingham 2011.

La terminologie médicale en tant que pratique métalinguistique

DOINA BUTIURCĂ, CRISTINA-ALICE TOMA, ELENA MUSEANU

1. Fondements théoriques : le modèle intégré Jakobson-Bühler

Dans la classification des fonctions du langage, Roman Jakobson est le premier parmi les linguistes européens à théoriser explicitement la fonction métalinguistique, qu'il conçoit comme une fonction distincte, activée par l'émetteur uniquement lorsque le message est orienté vers le code, dans le but d'expliquer les termes du message, d'en vérifier le sens ou d'en corriger l'usage. Dans les termes de Jakobson, la fonction métalinguistique « intervient » lorsque l'expéditeur et le récepteur doivent vérifier s'ils utilisent le même code. Il s'agit de la fonction dont le rôle est d'expliquer le langage et la manière dont celui-ci doit être employé afin de comprendre la réalité, en particulier telle qu'elle est comprise par l'émetteur du message (Jakobson, 1960). Dans son rapport à la langue, la fonction métalinguistique a pour rôle de décrire, d'apporter des explications pertinentes ou d'analyser le code linguistique proprement dit – en d'autres termes d'explicitier le sens de certains mots, les règles grammaticales, avant tout, plutôt que de formuler et de transmettre des informations concernant la réalité extralinguistique. Elle est fondamentale, dans la conception de Jakobson, dans les situations de définition, de reformulation, de correction sémantique ou de délimitation conceptuelle, et elle se distingue de la fonction référentielle. La fonction métalinguistique a pour objet le langage même. À ce titre, la terminologie constitue l'une des réalisations les plus édifiantes de la métalinguistique appliquée, puisqu'elle opère de manière systématique avec des définitions, des classifications et des relations conceptuelles. La fonction métalinguistique assure la compréhension du message issu des sciences du langage, par exemple à travers les dictionnaires explicatifs, étymologiques, les dictionnaires de néologismes, les définitions grammaticales etc. Dans tous ces modèles de connaissance interdisciplinaire relevant des sciences du langage, le message est orienté vers le code lui-même, étant utilisé pour faciliter la compréhension, pour expliquer, pour définir une classe grammaticale, une unité syntaxique, un type de relation linguistique, ou une modalité de vérification du sens des unités linguistiques.

La dichotomie relative aux sphères de fonctionnement d'une langue – la langue comme phénomène et la langue comme objet – a été formulée de manière indirecte par Karl Bühler (1934), puis développée par Ferdinand de Saussure (1998) et Roman Jakobson (1963), entre autres. Étudier la structure d'une langue implique un processus d'abstraction et de conceptualisation, suppose des taxonomies et la structuration des notions

à un niveau cognitif supérieur, à la différence de parler, contexte dans lequel la langue est utilisée comme produit. Dans le modèle organon de Karl Bühler, le langage remplit trois fonctions fondamentales : la fonction de représentation (*Darstellung*), la fonction expressive (*Ausdruck*) et la fonction d'appel (*Appell*). Bien que Bühler ne mette pas explicitement en évidence une « fonction métalinguistique », la réflexivité du langage est implicite dans la fonction de représentation : le langage peut représenter toute chose, y compris son propre code. Il s'agit d'une caractéristique exploitée de manière systématique en grammaire et en terminologie. Dans la terminologie médicale classique, par exemple, la fonction de représentation se replie sur le langage lui-même, préparant ainsi le terrain à une conceptualisation métalinguistique explicite. Comme nous l'avons mentionné à une autre occasion, « l'auteur associe le concept d'« acte de conscience », théorisé dans la phénoménologie de Husserl, à celui d'« acte intentionnel », dans le processus de nomination des objets de la réalité environnante. Nommer les objets, les choses et les concepts par l'intermédiaire du langage (fonction de représentation) constitue, dans la vision de Bühler, une question de conscience et d'intention, deux éléments par lesquels l'être humain se rapporte au référent et, implicitement, au monde » (Butiurca, 2025 : 34).

La fonction métalinguistique n'est donc pas un phénomène marginal, mais l'axe central qui permet le passage du langage commun aux langages spécialisés. Le modèle intégré Jakobson-Bühler met en évidence le fait que la terminologie représente la forme de réflexivité du langage.

2. La terminologie médicale en tant que pratique métalinguistique

Du point de vue des structures, la terminologie médicale fonctionne dans un régime de métalangage. Dans la pratique des nomenclatures, la terminologie ne recourt pas à la description directe de la réalité anatomique, mais établit et coordonne/contrôle le code linguistique par lequel celle-ci est décrite. En pensant dans la terminologie de Roman Jakobson, l'objet de la communication médicale est le concept (clinique, anatomique, neurologique, génétique etc.). L'activité terminologique vise la dénomination et la définition du concept, en d'autres termes le code. Dans ce contexte, les définitions du concept médical, les glossaires, les nomenclatures (*ICD*, *SNOMED CT*, *Nomina Anatomica*), mises en évidence par ISO 704 : 2022, constituent des actes de nature métalinguistique (ISO 704 : 2022, *Terminology work – Principles and methods*).

La définition médicale est, au plus haut degré, l'expression de la fonction métalinguistique. Aucune définition médicale ne peut être considérée comme un simple énoncé descriptif, ne serait-ce qu'à un examen superficiel de son langage hermétique. La définition est un énoncé métalinguistique à valeur normative, qui fixe les traits intrinsèques d'un concept. Par exemple, la définition du terme *amnésie antérograde* dans les dictionnaires médicaux ou dans les traités de médecine interne [(« ... incapacité à former et à fixer de nouveaux souvenirs après l'apparition d'une lésion ou d'un dysfonctionnement cérébral », notre traduction (*Adams and Victor's Principles of Neurology*, 2014))] n'illustre pas une situation clinique particulière ; le langage ne renvoie pas à une description ou à

un récit, mais définit le sens du terme, en fixant ses conditions d'emploi dans le discours médical. *L'amnésie antérograde* est définie par rapport à un critère précis (l'incapacité à former et à fixer de nouveaux souvenirs), et non par l'énumération de symptômes. Cette définition délimite le concept, le différencie des notions voisines et permet une communication intersubjective standardisée. Ainsi, la fonction référentielle se subordonne à la fonction métalinguistique, étant donné qu'avant de parler d'un symptôme cognitif neuropsychique, il est nécessaire de définir le code scientifique.

Ce n'est pas un hasard que la fonction métalinguistique est prédominante en médecine scientifique ou dans les sciences techniques. La première raison est que la médecine constitue un domaine dans lequel les erreurs sémantiques ont des conséquences cliniques réelles et irréversibles. Les termes doivent être univoques, reproductibles et traduisibles d'une langue à l'autre. En conséquence, la fonction métalinguistique est devenue une condition *sine qua non* de la communication médicale, en l'absence de laquelle celle-ci glisserait vers la polysémie, vers des métaphores étrangères au concept et chargées de subjectivité, vers l'ambiguïté pragmatique. C'est la raison pour laquelle la médecine développe des terminologies standardisées, le langage médical étant, de manière structurée, autoréflexif (Cabré, 1999). La fonction métalinguistique n'est pas neutre dans la terminologie médicale, celle-ci instituant un type d'autorité épistémique (Temmerman, 2000 ; Foucault, 1963). En vertu de cette instance de la science, il est décidé quels termes sont devenus improductifs, quelles formes sont archaïques ou dialectales, quelles définitions sont privilégiées et quels modèles linguistiques doivent être évités. Ainsi, par exemple, en sémiologie médicale, la différenciation stricte entre les concepts de signe et de symptôme est soutenue sur le plan métalinguistique, de même que la redéfinition des entités nosologiques. Ce ne sont là que quelques aspects qui illustrent le fait que la fonction métalinguistique est devenue un instrument de gouvernance des concepts, des définitions et des relations de nature conceptuelle. Par conséquent, si la terminologie médicale comprend le terme médical, la définition clinique, l'univocité conceptuelle, en ayant un statut épistémique et normatif, la linguistique assure, par la fonction métalinguistique, le code, la définition et l'explicitation des sens, évitant ainsi l'ambiguïté et, au niveau de la pratique médicale, l'erreur clinique.

La standardisation du code représente un autre aspect d'une importance majeure de la pratique métalinguistique en médecine. Un système tel que *Nomina Anatomica* ne décrit pas la réalité anatomique, mais harmonise le langage par lequel l'anatomie humaine est décrite. Il s'agit d'une démarche qui, conformément à ISO 704, suppose la détermination des concepts, les relations entre ceux-ci et leur désignation au moyen de dénominations univoques. *SNOMED CT*, la terminologie clinique multilingue la plus complète au monde, l'*ICD*, utilisée pour la classification des maladies, l'*ATC*, le système de classification chimique, thérapeutique et anatomique des médicaments, etc. constituent d'autres exemples où la standardisation permet de réduire les erreurs médicales causées par une interprétation incorrecte des termes. Tous ces systèmes terminologiques sont, par excellence, de nature métalinguistique. Il s'agit de systèmes dont le rôle est de

créer un méta-niveau du discours médical, rendant possible la communication clinique, la recherche et l'enseignement médical.

3. Terminologique et linguistique en neurologie

Tous les sous-domaines médicaux constituent des sphères présentant une dépendance structurelle à l'égard du langage. La neurologie est un domaine médical dans lequel le langage n'est pas seulement un instrument de description, mais une condition épistémique de la connaissance. Les signes et les syndromes neurologiques ne se sont pas imposés comme de simples sens et ne sont pas des termes sélectionnés de manière aléatoire par les spécialistes. Les termes relevant de la neurologie se sont imposés avec le statut de constructions conceptuelles stabilisées par des définitions. Cette situation fait de la neurologie un domaine privilégié pour l'étude de la fonction métalinguistique, au sens de la théorie de Roman Jakobson, ainsi que pour l'articulation de la relation interdisciplinaire entre terminologie et linguistique. Dans la tradition classique, Jean-Martin Charcot et Joseph Jules Déjerine considéraient que la neurologie se construit comme une science des délimitations subtiles : entre fonctions préservées et fonctions perdues, entre troubles primaires et secondaires, entre signe et symptôme. Ces délimitations sont réalisées par le langage, plus précisément par la terminologie. Si le fondateur de la neurologie clinique moderne, Jean-Martin Charcot (1872-1887), était orienté vers la corrélation entre symptôme et lésion, Joseph Jules Déjerine (1914), opérant la transition vers une vision plus fonctionnelle, admettait une ouverture plus large vers la linguistique, en ce qui concerne le langage de la neurologie. La question que nous nous posons est de savoir ce qui relève de la terminologie et ce qui relève de la linguistique dans le domaine de la neurologie. La désignation des concepts neurologiques (aphasie, apraxie, agnosie etc.), les relations conceptuelles (partie-tout, cause-effet, syndrome-signes), la standardisation des nomenclatures (dictionnaires, traités, classifications) relèvent de la terminologie. La structure des termes, la relation entre le langage et la fonction cérébrale, les mécanismes sémiotiques impliqués dans la production et la compréhension du langage relèvent de la linguistique, ainsi que la différence entre déficit linguistique et déficit cognitif général. Dans la neurologie traditionnelle, l'aphasie (*Adams and Victor's Principles of Neurology*, 2014) est définie comme un trouble du langage dû à une lésion cérébrale, en l'absence de paralysie, de déficit sensoriel primaire ou de troubles psychiques globaux. La définition constitue en elle-même un acte métalinguistique, qui désigne et fixe le sens du terme ainsi que ses limites d'application. L'objet d'analyse de la linguistique est, dans ce cas, la structure du langage affecté (au niveau phonologique, morphologique, syntaxique et sémantique), ainsi que les types d'aphasie (Broca, Wernicke, globale), corrélés aux profils linguistiques. Il s'agit du contexte dans lequel le langage devient objet d'analyse scientifique, et non seulement instrument de diagnostic. Un autre terme est l'*apraxie*, définie comme l'incapacité d'exécuter des actes moteurs volontaires en l'absence de déficit moteur primaire. Il s'agit d'une définition négative pratiquée régulièrement en neurologie, qui impose une attention particulière en ce qui concerne la fonction métalinguistique

(*Sémiologie médicale*). La linguistique est étroitement liée à la cognition. La linguistique cognitive et la neurolinguistique se concentrent sur la relation entre langage, perception et catégorisation, ainsi que sur le rôle de la dénomination dans la reconnaissance des objets. Un exemple en ce sens est le concept d'agnosie.

La linguistique et les sciences du langage contribuent à l'analyse de la relation entre geste, symbole et langage, ainsi qu'à celle de la différence entre acte moteur et acte de signification et à l'analyse de la structure du terme (grec, préfixe privatif *a-* + gr. *praxis*). La terminologie de la neurologie présente de nombreuses particularités : elle utilise le langage comme objet, délimite les concepts par négation (aphasie, apraxie), utilise le langage comme instrument de délimitation cognitive (agnosie), etc.

4. Principes de la terminologie neurologique classique

La langue grecque a joué un rôle décisif dans la terminologie neurologique classique, étant la langue dans laquelle ont été conceptualisées les fonctions cognitives (gr. *φάσις*, gr. *μνήμη*, gr. *γνώσις*), les relations entre action, perception et langage, ainsi que les processus sensoriels (gr. *ὄψις*, gr. *ἀκοή*, gr. *αἴσθησις*, gr. *ἀφή*, gr. *γεῦσις*, gr. *ὀσμή*). Ainsi, par exemple, au niveau des nomenclatures, *ὄψις* (« vue ») a servi de base pour *optique*, *opsie*, *hémianopsie*, tandis que *ἀκοή* (« ouïe ») constitue la base pour *acoustique* et *acouphènes*. La forme grecque *αἴσθησις* (« sensation », « perception ») a fourni le fondement nécessaire pour *anesthésie* et *hyperesthésie* ; *γεῦσις* (« goût ») est resté la base pour *agueusie* et *dysgueusie*, et *ὀσμή* (« odorat ») pour *anosmie*.

En tant qu'instruments de définition métalinguistique, les préfixes privatifs grecs peuvent désigner la perte d'une fonction, son absence complète. Il convient de retenir quelques exemples : *ἀ-ἀν-* (*a-lan-*), préfixes privatifs (*aphasie*, *apraxie*, *analgésie*, *anosmie*). Le préfixe *δυσ-* (*dys-*) n'indique pas une absence totale, mais un fonctionnement anormal, comme dans *dysphasie* ou *dysarthrie*. Considéré comme un préfixe quasi privatif, *ὑπο-* (*hypo-*, « sous ») marque le sens de déficit partiel dans des termes tels : *hypoesthésie*, *hypotonie*, etc. Les différences au sein de la classe des préfixes négatifs sont de nature conceptuelle et diagnostique, plutôt que d'ordre stylistique. L'opposition conceptuelle à valeur clinique se réalise également au niveau des préfixes (*aphasie* par opposition à *dysphasie*). Pour la terminologie médicale européenne, le latin médical constitue une langue de stabilisation, de normalisation et de circulation scientifique des concepts issus de la langue grecque, langue de création conceptuelle *ab initio*. Le latin médical fixe la forme des termes d'origine grecque, assure la cohérence sémantique et la circulation interlinguistique des connaissances neurologiques.

5. La définition médicale en tant qu'acte métalinguistique (Cabré)

Les sens spécialisés se sont constitués à partir du grec médical et ont été poursuivis dans le latin médical non seulement par les termes, mais aussi par les définitions. La tradition de la définition médicale comme acte métalinguistique (conception qui, certes, n'a pas été

formulée théoriquement dans l'Antiquité – ce n'est qu'à la fin du XX^e siècle que Maria Teresa Cabré propose le concept de la définition en tant qu'acte métalinguistique) prend naissance dans la médecine grecque et se poursuit dans la médecine latine. Les fondements logiques et philosophiques du contenu de la définition ont été identifiés dans les écrits d'Aristote (1965, chapitre 10). Le modèle aristotélicien constitue la « matrice logique » de toutes les définitions médicales grecques, poursuivi par la définition pratiquée dans le latin médical et par la médecine moderne. En vertu de ce modèle, on précise : a. le terme défini (le sujet) ; b) la nature du concept médical, information relevant du genre prochain ; c) quels sont les éléments différenciateurs, à savoir la différence spécifique. La définition dans le domaine de la neurologie (et pas seulement) constitue un acte épistémique, et non un acte descriptif. Parmi les premières définitions médicales fonctionnelles et descriptives pratiquées en Grèce figurent celles élaborées par Hippocrate, qui recourt à des substantifs abstraits pour désigner les concepts et définit la maladie par la fonction atteinte, et non comme un symptôme isolé. Il s'agit de ce que la littérature spécialisée désigne comme une définition clinique fonctionnelle, admise par la neurologie moderne. Galien a fixé la matrice définitoire de la définition médicale classique, reprise par le latin médical. Le médecin définit les maladies par la cause, le siège et la fonction, l'apport essentiel résidant dans l'introduction explicite de la définition négative. La typologie de la définition médicale, au sens large, est extrêmement variée et diffère du point de vue de l'aspect énonciatif et de la finalité scientifique et médicale, de nature conceptuelle. À partir de la Grèce antique, la définition essentielle, de forme classique, a pour rôle de fixer l'identité conceptuelle de l'entité (signe, symptôme, etc.) et se développe sur la base de la matrice aristotélicienne du type « A est un type de B caractérisé par C ». Il s'agit de la définition standard dont dérivent tous les autres types de définitions ou sur laquelle ils prennent appui.

La définition négative, formée par exclusion (Galen, *De symptomatum causis*, in *Opera omnia*, vol. VII), a pour rôle de réaliser une délimitation différentielle fine (fondamentale pour le diagnostic), fondée sur ce qui n'est pas présent (« A est B sans C, en l'absence de, sans C »). Ce type de définition possède une valeur métalinguistique maximale. La définition différentielle-négative s'est développée à partir des préfixes négatifs. En grec, prédominaient les préfixes négatifs *ἀ-*, *ἀνευ* (« sans »), *ἀ-/ἀν-* (*a-/an-*), tandis qu'en latin étaient utilisées les formes *sine* (« sans »), *absque* (« en l'absence de »), de manière constante (ex., *Apraxia est defectus actionis voluntariae sine paralyisi*) et ayant les mêmes valeurs négatives. Au niveau de la définition en tant qu'acte métalinguistique, la négation constitue le mode privilégié d'expression des raisonnements cliniques dans le discours scientifique contemporain.

La définition fonctionnelle (Hippocrate, *De natura hominis*, 1849) définit l'entité par la fonction atteinte et propose la corrélation symptôme-fonction cérébrale (par exemple, *agnosie*, trouble de la fonction de reconnaissance, où « A est une altération de la fonction Z »).

La définition causale ou étiologique (Galien, *De symptomatum causis*) définit le concept par la cause et par le mécanisme. Galien a également défini ses concepts par la corrélation symptôme-siège (*De locis affectis*), au moyen de définitions de nature anatomoclinique.

La nouveauté apportée par Isidore de Séville (Isidor, *Etymologiae*, *De medicina IV*, 1935) consiste à introduire la relation étymologie-définition, renforçant la stabilité des sens médicaux dans le latin médiéval et consolidant la tradition gréco-latine, définitoire pour le Moyen Âge.

Dans la médecine grecque, la définition est descriptive. Dans le latin médical, la définition devient norme et doit exprimer le sens stable et utile du terme, et non le statut métaphysique de la maladie. Le schéma grec est repris, mais Aulus Cornelius Celsus (Celsus, *De Medicina I-II*, 1935) propose une simplification de la définition, au niveau syntaxique et stylistique, en fixant un modèle latin standard du type : « A est un virus/ signe/ symptôme/ maladie ». Par comparaison avec Galien, Celsus utilise assez rarement la définition négative (*sine*, *absque*), soit pour marquer la valeur différentielle, soit pour éviter des confusions cliniques majeures.

Charcot (1872-1887) et Joseph Déjerine (1914) observent que cette réserve de Celsus justifie le fait que la définition négative ne se développe pleinement que dans la neurologie classique. Nous nous proposons d'illustrer la conception de Celsus en analysant l'une de ses définitions : *Tinnitus est affectio auditus, qua sonus percipitur sine causa externa* (« L'acouphène est un trouble de l'audition dans laquelle un son est perçu en l'absence de cause externe »). L'analyse structurale de la définition met en évidence les éléments suivants : i) *tinnitus*, terme défini et objet de la définition au niveau du raisonnement ; ii) *est*, copule à rôle de stabilisation normative ; iii) *affectio auditus*, genre prochain à fonction d'encadrement fonctionnel ; iv) *sonus percipitur*, différence spécifique ayant la fonction de phénomène définitoire ; v) *sine causa externa*, délimitation négative à fonction d'exclusion essentielle. Il s'agit d'une définition complète, dépourvue de toute ambiguïté.

Nous proposons également l'analyse d'une variante à accent causal, dans l'esprit des définitions proposées par Celsus : *Tinnitus est vitium auditus, ob internam auris perturbationem ortum* (« L'acouphène est un défaut de l'audition, apparu en raison d'une perturbation interne de l'oreille »), où *vitium*, substantif abstrait, désigne le concept de défaut stable. Dans cette définition, *ob* exprime le rapport de causalité et *perturbatio* est, dans ce contexte spécialisé, un terme abstrait à valeur normative.

Le modèle de Celsus appliqué à l'*acouphène* montre comment la définition médicale latine normative transforme un phénomène perceptif subjectif en un concept clinique stable, utilisable de manière intersubjective. Ce type de définition constitue la base formelle de la terminologie neurologique et otologique moderne. Dans ses définitions, Celsus a refusé l'énumération d'étiologies multiples, a évité de décrire des vécus subjectifs, d'introduire des classifications ou de recourir à un langage métaphorique. Toutes ces caractéristiques se retrouvent dans les définitions de la neurologie et de l'otologie modernes.

Dans un autre ordre d'idées, Aulus Cornelius Celsus (*De Medicina I-II*) adapte la définition aux exigences de la langue latine, en insistant sur la clarté et le respect de la norme. Dans *De Medicina*, il précise que les termes doivent être clairs pour le praticien, permettre l'application correcte de « l'art médical » et réduire l'ambiguïté terminologique. Il introduit le verbe copulatif et le complément de cause (*propter*, *ob*) dans la structure de la définition médicale (du type : « L'aphasie est... »). Par l'introduction

d'une cause explicite et claire, la définition devient opérationnelle, indiquant non seulement ce qu'est la maladie, mais aussi pourquoi elle apparaît, dans les limites des connaissances de l'époque de Celsus.

Le maintien et l'enrichissement du langage par des substantifs abstraits constituent un procédé essentiel pour le développement de la terminologie neurologique. Les définitions de Celsus utilisent fréquemment des substantifs tels que *morbus* (« maladie »), *affectio* (« affection »), *vitium* (« défaut »), *perturbatio* (« trouble »), renforçant le statut métalinguistique de la définition. Il s'agit d'une substantivation par laquelle le phénomène clinique est réifié sur le plan conceptuel, détaché du cas individuel et intégré dans une classification stable. Bien qu'il ne théorise pas les aspects du langage, Celsus pratique une métalinguistique implicite, par la normalisation du sens des termes et par la création d'un code médical latin stable, son traité devenant ainsi un modèle de normalisation terminologique, et non seulement un traité médical.

L'impact du modèle de définition proposé par Celsus a été majeur, en ce qu'il a imposé le latin comme langue de la définition normative, a stabilisé les formules définitoires et transmis la terminologie grecque sous une forme standardisée et durable. La neurologie moderne (à travers Charcot, Déjerine, Adams & Victor) poursuit, sur le plan structurel, cette tradition : des définitions claires, concises et normées.

Parmi les théoriciens de la période moderne, Eugen Wüster (*Einführung in die Allgemeine Terminologielehre*) propose la définition dite terminologique, de nature conceptuelle, impliquant la séparation concept-terme-définition, tandis que Maria Teresa Cabré (1999) théorise le concept de définition interdisciplinaire, apte à fonder l'idée de la définition comme acte métalinguistique. La définition en tant qu'acte métalinguistique et le rôle majeur des disciplines classiques dans la terminologie scientifique sont soutenus dans la majorité des études de terminologie scientifique par Maria Teresa Cabré.

6. Conclusions

Une première conclusion qui se dégage de notre recherche est le caractère interdisciplinaire de la terminologie en général, et de la terminologie médicale en particulier. La terminologie et la linguistique concourent à la création des nomenclatures, à la transparence des définitions, protègent les termes contre des réinterprétations métaphoriques incontrôlées et maintiennent la rigueur scientifique du langage médical en général, et du langage de la neurologie en particulier. Une autre conclusion qui s'impose est l'importance terminologique majeure du modèle de Celsus pour la neurologie moderne, modèle qui non seulement fonde, mais harmonise depuis plus de deux millénaires la définition médicale. Même si l'on ne considère qu'une seule fonction de la langue – la fonction métalinguistique – que nous avons appliquée, dans notre recherche, à différents modèles de définitions issus de la médecine grecque et du latin médical, le rôle de la linguistique et des sciences du langage est majeur en terminologie. La définition médicale est métalinguistique ; elle stabilise le code. Outre la clarté assurée par la fixation correcte du sens des termes, la fonction métalinguistique a pour rôle d'établir l'usage correct et de créer un code médical latin stable.

Références bibliographiques

- Cabré M.T., *Terminology: Theory, Methods and Applications*, Amsterdam/Philadelphia, John Benjamins, 1999.
- Bühler K., *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Jena, Fischer, 1934.
- Butiurcă D., *The functions of language and the organon concept in Karl Ludwig Bühler's linguistics*, in « Acta Marisiensis Philologia », 7, 2025, pp. 32-35.
- Jakobson R., *Linguistics and Poetics*, in A. Thomas (ed.), *Style in Language*, Cambridge, MIT Press, 1960, pp. 350-377.
- Jakobson R., *Essais de linguistique générale*, Paris, Les Éditions de Minuit, 1963.
- Saussure (de) F., *Cours de linguistique générale*, traduction et préface d'Irina Izverna Tarabac, Iași, Polirom, 1998.
- Temmerman R., *Towards New Ways of Terminology Description: the Sociocognitive Approach*, Amsterdam/Philadelphia, John Benjamins, 2000.

Sources consultées

- Adams R.D., Victor M., Ropper A.H., *Adams and Victor's Principles of Neurology*, Columbus, McGraw-Hill, 10th edition, 2014.
- Aristotel, *Analiticele posterioare*, traduction, étude introductive et notes par Mircea Florian, București, Editura Științifică, 1965.
- Cambier J., Masson M., Dehen P., *Sémiologie médicale*, Paris, Masson et C^{ie}, 1979.
- Celsus, *De medicina*, trans. W.G. Spencer, Cambridge, Loeb Classical Library, 1935.
- Charcot J.-M., *Leçons sur les maladies du système nerveux*, Paris, Adrien Delahaye et Émile Lecrosnier, 1872-1887.
- Déjerine J.J., *Anatomie des centres nerveux*, Paris, Rueff et C^{ie}, 1895-1901.
- Déjerine J.J., Déjerine-Klumpke A., *Sémiologie des affections du système nerveux*, Paris, Masson et C^{ie}, 1914.
- Foucault M., *Naissance de la clinique. Une archéologie du regard médical*, Paris, PUF, 1963.
- Galen, *De symptomatum causis*, Liber I, cap. 3, in *Opera omnia*, vol. VII, ed. Kühn, URL : <<https://scaife.perseus.org/reader/urn:cts:greekLit:tlg0057.tlg044.1st1K-grc1:1.6>> (consulté le 24-11-2025).
- Hippocrate, *De la nature de l'homme*, in *Œuvres complètes d'Hippocrate*, texte grec et traduction française par Émile Littré, vol. VI, Paris, J.-B. Baillière, 1849, URL : <<https://gallica.bnf.fr/ark:/12148/bpt6k6269561j/f346.vertical>> (consulté le 24-11-2025).
- Isidor din Sevilla, *Etymologiarum sive Originum libri XX* (ed. W.M. Lindsay), Oxford, Clarendon Press, 1911, Liber IV ("De medicina"), URL : <<https://www.thelatinlibrary.com/isidore.html>> (consulté le 24-11-2025).
- ISO 704 : 2022, *Terminology work – Principles and methods*, International Organization for Standardization (ISO), 2022.
- Liddell H.G., Scott R., Jones H.S., *Greek-English Lexicon*, 9^a rev., Oxford, Oxford University Press, 1996.

Traduction spécialisée et post-édition : pour une modélisation du processus de gestion des textes

SILVIA CALVI, KLARA DANKOVA¹

1. Introduction

L'essor de l'intelligence artificielle (IA) et de ses applications dans le domaine de l'industrie de la langue capables, entre autres, de rédiger ou de traduire des textes en réponse aux stimuli humains, a révolutionné le secteur de la traduction. Loin d'être une menace, l'emploi de nouvelles technologies représente une opportunité en termes d'efficacité : en déléguant aux logiciels l'élaboration d'une première traduction, de qualité de plus en plus convaincante (Popel *et al.*, 2020), les spécialistes des langues peuvent se focaliser sur les aspects les plus problématiques (Flöter-Durr, 2022). En effet, il n'est pas possible d'accepter les propositions des outils d'IA sans une évaluation critique : pour ce faire, une sensibilisation des professionnels aux enjeux et aux opportunités du traitement automatique des langues s'impose.

Dans ce contexte, des études récentes (Villa *et al.*, 2022 ; Calvi, Dankova, 2022 ; Ali *et al.*, 2023) soulignent les limites des outils de traduction automatique (TA) par rapport au traitement terminologique, notamment en ce qui concerne la variation. Dans la plupart des cas, les systèmes des traducteurs automatiques ne permettent pas de traduire la terminologie en fonction des besoins de communication propres à chaque contexte et à chaque typologie textuelle (Zanola, 2018). Réfléchissant sur les étapes au cours desquelles l'emploi de la traduction automatique, neuronale ou générative, peut s'avérer efficace, nous proposons une modélisation du processus de la post-édition – « activité qui consiste à compléter, modifier, corriger, remanier, réviser et relire un texte brut prétraduit automatiquement pour le rendre humainement intelligible » (Robert, 2010 : 137-138) – pour les textes spécialisés avec une attention spéciale aux aspects terminologiques.

Tout d'abord, les enjeux majeurs de la traduction automatique seront décrits : des exemples de solutions de traduction imprécises ou erronées seront illustrés pour présenter une typologie d'erreurs effectuées par deux outils, le traducteur automatique DeepL et l'agent conversationnel ChatGPT. Ensuite, en nous s'inspirant de la position de

¹ Silvia Calvi a rédigé les paragraphes 1, 2, 3.2.1, 4 ; Klara Dankova les paragraphes 3, 3.1, 3.2., 3.2.1.1.

Conceição² selon lequel le traitement de la terminologie de la part des systèmes d'IA est souvent limité à la dimension linguistique des termes, nous avons choisi de centrer notre attention sur la performance de ChatGPT en ce qui concerne la dimension conceptuelle de la terminologie³. Ces analyses nous permettront de proposer une modélisation du processus de post-édition des textes spécialisés (Schumacher, 2019) dont les étapes clés sont indiquées dans la Fig. 2.

2. L'efficacité de la TA : aspects méthodologies

Notre modélisation du processus de post-édition des traductions spécialisées propose de fournir aux intervenants humains une feuille de route pratique pour l'utilisation la plus efficace possible des outils de traduction automatique neuronale (TAN) et des agents conversationnels. En exploitant les versions de DeepL⁴ et de ChatGPT⁵ en libre accès, nous allons analyser leur performance relativement à la qualité du texte prétraduit généré automatiquement dans le but d'identifier les tâches qui peuvent être facilitées par l'emploi de ces deux types d'outils, tout en gardant un esprit critique et d'analyse⁶. ChatGPT sera testé également pour ce qui est de sa capacité à contribuer à la collecte des termes et à leur analyse pour les besoins de traduction.

Le modèle proposé est conçu pour être appliqué à différents contextes de traduction des textes de spécialité, même s'il est possible de supposer que les performances de ces outils varient considérablement en fonction de plusieurs variables, dont :

- la combinaison linguistique ;
- le domaine de spécialité ;
- la typologie textuelle ;
- le degré de spécialisation des textes.

Celles-ci influencent la disponibilité et la représentativité des données linguistiques utilisées pour l'entraînement des outils de TA. Le critère de la combinaison linguistique considère la proximité des langues et leur présence sur internet : la TA sera probablement plus performante entre des langues voisines présentant des similarités aux niveaux syntaxique et lexical, aussi bien qu'entre des langues qui sont bien représentées sur internet. Le domaine de spécialité joue un rôle important, notamment pour ce qui est de la tradition de la rédaction de textes : en principe, les experts en sciences naturelles pro-

² Nous renvoyons à l'intervention du Professeur Manuel Célio Conceição du 29 mai 2024 lors des Journées d'études du Réseau LTT-Lexicologie, Terminologie, Traduction, intitulée « Terminologie et intelligence artificielle (IA) : impacts et questionnements épistémologiques et méthodologiques ».

³ Les données présentées sont issues d'une analyse effectuée en juin 2024.

⁴ URL : <<https://www.deepl.com/it/translator>> (consulté le 24-11-2025).

⁵ URL : <<https://chat.chatgptdemo.net/?via=topaitools>> (consulté le 24-11-2025).

⁶ Nous sommes bien conscientes qu'il existe des versions plus performantes de DeepL aussi bien que de ChatGPT : notre objectif n'étant pas d'évaluer le degré de performance des deux logiciels, mais plutôt celui de déterminer les aspects de la traduction spécialisée pour lesquels ceux-ci peuvent être utilisés – en fournissant des résultats de qualité différente –, l'emploi des versions citées s'avère être assez suffisant.

duisent des textes descriptifs et donc, vraisemblablement, plus facilement accessibles à la traduction automatique, tandis que ceux des sciences humaines et sociales élaborent des textes de type narratif, dont les résultats donnés par la traduction automatique s'avèrent plus déficitaires. La typologie textuelle et le degré de spécialisation qui en dérive sont déterminants, surtout pour les choix terminologiques : à titre d'exemple, un manuel d'arithmétique pour les élèves de l'école primaire utilisera probablement des termes de base, largement diffusés sur le web ; en revanche, un manuel d'algèbre destiné aux étudiants universitaires contiendra des termes d'un haut degré de spécialisation, qui peuvent être moins présents en ligne.

Notre expérimentation porte sur les traductions automatiques du français vers l'italien, deux langues romanes relativement bien représentées sur internet. Selon l'étude présentée dans le rapport *La langue française dans le monde* (OIF, 2022 : 314-317), le français et l'italien appartiennent aux quinze langues les plus puissantes de l'internet en 2021, leur présence correspondant respectivement à 3,5% et à 1,4 %, ce qui fait du français et de l'italien la quatrième et la douzième langue de la toile⁷. Quant au domaine de spécialité, nous avons considéré la mode, un domaine très vaste, qui peut être exploité en adoptant plusieurs points de vue : celui des sciences naturelles ayant des applications dans le domaine de l'industrie, aussi bien que celui des sciences humaines et sociales.

Les typologies textuelles sélectionnées pour l'évaluation de la TA et de l'extraction terminologique incluent une étude scientifique et un article de revue ; plus spécifiquement, il s'agit de :

- un extrait du chapitre *Arts et métiers du textile* de l'étude terminologique de Zanola (2014 : 174-176) : 588 occurrences ;
- un article de la revue de mode « Vogue » – *Chic et pratique, ce sac signé Loewe est la valeur sûre de l'automne*⁸ (06/10/2023) : 336 occurrences.

L'étude de Zanola (2014) a été prise en compte pour analyser la performance de la TA par rapport à un texte narratif de structure complexe, contenant des termes très spécialisés. Le choix d'un article de revue de mode s'explique par la volonté d'étudier un texte du même domaine, présentant une structure plus simple et des unités terminologiques d'un degré de spécialisation moins élevé.

Les aspects concernant l'analyse de la dimension conceptuelle de la terminologie (§3.2.1.1) ont été explorés en utilisant des extraits provenant des textes appartenant aux typologies suivantes :

- le manuel technique :
 - J. Anquetil, *Le tissage. Encyclopédie contemporaine des métiers d'art*, Paris, Dessain et Tolra, 1977 ;
 - P. Luc, *Le tissage de la soie artificielle*, Paris, L'édition textile, 1929 ;
 - D. Weidmann, *Aide-mémoire Textiles techniques*, Paris, Dunod, 2010 ;

⁷ Les trois premières places reviennent à l'anglais (25 %), au chinois (15 %) et à l'espagnol (7 %) (OIF 2022 : 317).

⁸ « Vogue », *Chic et pratique, ce sac signé Loewe est la valeur sûre de l'automne*. Disponible sur : <<https://www.vogue.fr/article/loewe-tendance-sac-squeeze-mode>> (consulté le 24-11-2025).

- l’ouvrage de vulgarisation :
 - C. Fauque, S. Bramel, *Une seconde peau : fibres et textiles d’aujourd’hui*, Paris, Éditions Alternatives, 1999 ;
- l’article de revue :
 - *Tissu en velours : notre sélection haute en douceur*⁹ (« Elle », 16/08/2018).

Leur choix a été motivé par la nature des termes employés, se prêtant bien à l’observation de la performance de l’agent conversationnel ChatGPT relativement à différents aspects de l’analyse terminologique, dont l’évolution conceptuelle.

3. La post-édition des traductions spécialisées : pour une proposition de modélisation du processus de gestion des textes

Les outils ne fournissant qu’un texte prétraduit généré par l’analyse automatique des données linguistiques disponibles, il est nécessaire d’effectuer une série d’interventions permettant de le rendre approprié tant d’un point de vue formel que fonctionnel. Pour ce faire, plusieurs aspects relevant de la macrostructure et de la microstructure du texte devraient être examinés dans la phase de la post-édition.

3.1 La macrostructure

Au niveau de la macrostructure, le contrôle se concentre sur des aspects d’ordre textuel et formel, notamment :

- l’organisation générale du texte (son plan) ;
- les conventions typographiques, telles que l’utilisation de caractères gras, italiques ou de soulignement ;
- la présence des figures, dont des graphiques, des tableaux ou des illustrations.

Notre analyse a relevé que les deux outils considérés fournissent des traductions automatiques présentant un plan identique à celui du texte de départ. Quant à la reproduction des aspects typographiques, si le texte à traduire est inséré directement dans l’espace dédié, DeepL génère un texte brut indépendamment du formatage du texte source, tandis que ChatGPT produit un texte qui peut contenir des marques typographiques, même si celles-ci ne correspondent pas toujours à celles du texte original. La traduction au moyen d’un fichier chargé dans le système a donné les mêmes résultats pour ChatGPT ; DeepL, en revanche, a fourni une traduction automatique bien formatée.

Relativement aux figures, nous avons testé les deux outils en ce qui concerne le traitement des tableaux, des graphiques et des illustrations¹⁰. Nous avons observé que

⁹ « Elle », *Tissu en velours : notre sélection haute en douceur !*, URL : <<https://www.elle.fr/Deco/News-tendances/Tissu-en-velours>> (consulté le 24-11-2025).

¹⁰ Les textes considérés ne présentant pas de graphiques ou de tableaux, nous avons effectué cette partie de l’analyse en utilisant un graphique et un tableau créés *ad hoc* dans un document Word : cela nous a permis, entre autres, de fournir aux outils choisis des figures intelligibles. La reproduction des illustrations a été testée en utilisant celles de l’article de « Vogue ».

ChatGPT est capable de fournir une traduction des tableaux – même si dans un format différent –, alors qu’il semble incapable de générer des graphiques : lors de notre essai, il les a complètement omis ou bien il n’a proposé que la traduction de leur légende. Il n’a même pas reproduit les illustrations en format d’image. Le traducteur automatique de DeepL peut traiter les tableaux et les illustrations présents dans un fichier chargé dans le système : les illustrations semblent être placées à un endroit convenable dans le document et le formatage des tableaux reflète les choix du texte de départ, mais nous avons observé que leur traduction peut être incomplète, certains champs pouvant rester dans la langue de départ. Quant aux graphiques, ce logiciel les a reproduits sans aucune traduction de leur légende. Quoique les résultats de notre analyse concernant les conventions typographiques et la traduction ou la reproduction des figures doivent être considérés avec prudence – l’analyse étant effectuée sur un échantillon de taille assez limitée –, nous pouvons en déduire qu’il s’agit d’aspects qui devraient être bien contrôlés dans la post-édition.

3.2 La microstructure

Les éléments relatifs à la microstructure du texte qui font l’objet d’examen lors de la post-édition peuvent être reconduits aux catégories suivantes :

- l’orthographe lexicale ;
- les aspects grammaticaux ;
- les choix lexicaux ;
- les choix terminologiques.

Relativement à l’orthographe lexicale, nous avons vérifié la correspondance entre graphèmes et phonèmes des mots pris isolément, c’est-à-dire sans considérer leurs morphèmes grammaticaux. À cet égard, DeepL aussi bien que ChatGPT s’avèrent être des outils très performants, permettant d’éviter des fautes de frappe ou d’autres erreurs relevant de la méconnaissance de l’orthographe des mots.

La pertinence des aspects grammaticaux considérés dans un sens large – la morphologie des noms, des adjectifs, des verbes, les reprises pronominales, l’ordre des mots, etc. – mérite d’être contrôlée avec attention : même si ces outils semblent devenir de plus en plus efficaces à ce niveau, cela n’empêche pas que les traductions automatiques puissent comporter des erreurs de ce type. Observons, par exemple, celles qui sont liées à la forme des adjectifs [1] et des pronoms [2]¹¹ :

- [1] Chic et pratique, ce sac signé Loewe [...] (article « Vogue »).
Elegante e pratico (*pratica*), questa borsa firmata Loewe [...] (TA, ChatGPT).

¹¹ Les erreurs de la TA qui font l’objet d’analyse aussi bien que la partie correspondante dans le texte de départ sont soulignées. Pour plus d’intelligibilité, nous n’avons pas souligné d’autres types d’erreurs. Des solutions de traduction possibles sont proposées en italique entre parenthèses (pour toutes les erreurs détectées). L’ajout d’un élément est marqué par le signe +.

- [2] Pour notre curiosité terminologique, il suffira de nous arrêter sur l'exemple des manufactures d'Aubusson, pour nous rendre compte que cela mériterait une étude beaucoup plus approfondie (Zanola, 2014 : 175).
Per ~~la nostra~~ (+ *rispondere alle*) curiosità terminologica (*terminologica*), ci basterà soffermarci (*soffermarsì*) sull'esempio delle manifatture di Aubusson, per renderci (*rendersì*) conto che questo (+ *argomento*) merita uno studio molto più approfondito (TA, ChatGPT).

Dans l'extrait [1], ChatGPT traduit l'adjectif *pratique* avec *pratico* : en se référant à *borsa*, un nom de genre féminin, la forme correcte de l'adjectif devrait être *pratica*. Cette erreur est due probablement à la difficulté de lecture des relations syntaxiques dérivant de la position de l'adjectif en apposition avant le nom ; en effet, les formes appropriées du déterminant démonstratif (*questa*) et du participe passé (*firmata*) montrent que l'outil a identifié correctement le genre féminin de *borsa*. L'erreur dans la TA effectuée par ChatGPT dans l'extrait [2] concerne l'emploi des pronoms : en conformité avec le mode impersonnel privilégié dans les écrits scientifiques en italien, les pronoms de la première personne du pluriel du texte en français devraient être omis ou, éventuellement, rendus avec la forme impersonnelle (*sì*) en italien.

Les imprécisions liées aux choix lexicaux, incluant aussi des collocations ou des expressions idiomatiques, sont en principe beaucoup plus nombreuses que les erreurs de nature grammaticale (voir Villa *et al.*, 2022). Dans les textes analysés, nous avons relevé notamment des erreurs relatives au choix des noms [3], des collocations de nature adjectivale [4], des verbes [5], des prépositions [6] et des expressions idiomatiques [7] :

- [3] Nous observerons la typologie de deux ouvrages sur le sujet au cours du XVIII^e siècle [...] (Zanola, 2014 : 176).
Osserveremo (*Si osserverà*) la tipologia di due opere sul soggetto (*sull'argomento*) nel corso del XVIII secolo [...] (TA, ChatGPT).
- [4] La créativité sans faille [...] (article « Vogue »).
L'immancabile (*impeccabile*) creatività [...] (TA, DeepL).
- [5] Il faut attendre que les historiens aient fait l'histoire de ces transformations et espérer qu'ils ne négligeront pas l'effet qu'elles ont eu sur les mots (Zanola 2014 : 176).
Dobbiamo (*Bisogna*) aspettare che gli storici scrivano (*descrivano*) la storia di queste trasformazioni e sperare che non trascurino l'effetto che ebbero sulle parole (TA, DeepL).
- [6] Emily Ratajkowski a été aperçue à plusieurs reprises avec le Squeeze de Loewe à l'épaule (article « Vogue »).
Emily Ratajkowski è stata avvistata più volte con la Loewe Squeeze sulla (*in*) spalla (TA, ChatGPT).
- [7] Le Fashion Month a battu son plein en septembre, avec des défilés qui se sont déroulés entre New York, Londres, Milan et Paris (article « Vogue »).
Il mese della moda era in pieno svolgimento (*è stato al culmine*) a settembre, con sfilate a New York, Londra, Milano e Parigi (TA, DeepL).

Même si la TA paraît de plus en plus performante en ce qui concerne les choix lexicaux, ceux-ci devraient être évalués de manière rigoureuse et systématique dans la phase de la post-édition pour proposer une traduction de qualité.

3.2.1 *Les choix terminologiques*

Lors de la post-édition d'une traduction automatique d'un texte de spécialité, il est nécessaire de vérifier avec attention les termes proposés par la TA, en effectuant une recherche sur les équivalents terminologiques pertinents pour le texte et le contexte donnés. Leur identification requiert la considération de plusieurs critères. Avant de procéder à l'évaluation de la terminologie, il est fondamental d'analyser le texte pour ce qui est des besoins des destinataires envisagés, de sa typologie et de son degré de spécialisation. La réflexion sur ces aspects permet de sélectionner des termes compréhensibles pour le public visé et, le cas échéant, de compléter le texte traduit par des notes explicatives du traducteur.

Tableau 1 - *Analyse des besoins des destinataires et de la typologie textuelle*

Évaluation	Article « Vogue »	Zanola 2014
Besoins des destinataires	Connaître les dernières tendances de la mode	Étudier les arts et métiers du textile du XVIII ^e siècle et leur terminologie
Typologie textuelle	Presse (revue de mode)	Texte académique (monographie)
Degré de spécialisation	Bas	Très haut

S'adressant au grand public intéressé par les tendances de la mode, l'article de la revue « Vogue » présente principalement des termes fondamentaux du domaine. Dans le cas des termes désignant des concepts plus spécialisés – tels que *nappa*, « peau souple de pleine fleur pour gants et vêtements, préparée à partir d'une peau de mouton d'agneau ou de chevreau, non refendue »¹² – le post-éditeur ou la post-éditrice peut envisager de faciliter la compréhension de l'équivalent en italien (*nappa*) par l'ajout d'une paraphrase. Dans les TA analysées, nous avons repéré l'équivalent en italien sans aucune explication (DeepL) et l'emploi de l'hyperonyme *pelle* (ChatGPT) : la solution de DeepL est correcte, mais probablement peu appropriée aux destinataires, alors que celle de ChatGPT est accessible aux lecteurs, mais ne transmet pas toutes les informations nécessaires à l'identification du concept en question. L'analyse de l'extrait du chapitre *Arts et métiers du textile* (Zanola 2014) contenant un grand nombre de termes d'un haut degré de spécialisation a porté à des conclusions différentes : destiné aux experts du domaine intéressés à l'histoire qui maîtrisent bien les termes ou aux terminologues qui connaissent les

¹² *Grand dictionnaire terminologique*, s.v. *nappa*, URL : <<https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8476244/nappa>> (consulté le 24-11-2025).

sources permettant l'accès à cette terminologie, la traduction du texte en italien peut être réalisée en employant des termes techniques sans aucun commentaire supplémentaire.

Pour valider les choix terminologiques, il est nécessaire de vérifier non seulement si le terme désigne effectivement le concept donné, mais également s'il est approprié à la situation de communication, en tenant compte de la variation terminologique aux niveaux diatopique, diaphasique et diachronique. Le texte considéré présente un grand nombre de termes techniques, dont plusieurs n'ont pas été rendus correctement dans les TA :

- [8] Vaucanson invente le tour pour une meilleure croisure et les moulins à organsiner pour adapter la vitesse des bobines. Pour la fabrication du textile de soie (le tissage, le moirage, l'écrasage), il invente le métier à tisser automatique, doté d'un automatisme intégral (Zanola, 2014 : 175). Vaucanson inventò il tornio per una migliore reticolazione (*torsione*) e i mulini organizzatori (*mulini da organzino*) per adattare la velocità delle bobine. Per la produzione di tessuti di seta (tessitura, moiré (*marezzatura*) e schiacciamento), inventò il telaio completamente automatico (TA, DeepL).

Dans cet extrait, les termes problématiques désignent des procédés – la *croisure*, le *moirage* – et des instruments – les *moulins à organsiner* – utilisés dans l'industrie textile ; la non-identification de leurs équivalents en italien dérive vraisemblablement de leur degré de spécialisation et, dans une certaine mesure, aussi de leur emploi : dans l'extrait, ils n'apparaissent qu'une seule fois, sans un contexte plus large. Employés plusieurs fois et dans des contextes permettant de cerner le concept en question, la TA pourrait donner des résultats plus satisfaisants. Il est intéressant de noter que, dans le cas de *moulins à organsiner*, le logiciel analyse le verbe *organsiner* le confondant avec *organiser*.

Les outils de TA ont échoué fréquemment aussi en ce qui concerne leur proposition d'équivalents des termes désignant des concepts anciens, présents dans l'extrait de Zanola 2014 :

- [9] La France manquait de manufactures capables de produire des bobines de soie régulières (Zanola, 2014 : 174). In Francia mancavano fabbriche (*manifatture*) in grado di produrre bobine di seta regolari (TA, DeepL).
- [10] une Ordonnance, adressée aux marchands drapiers, maîtres drapans, sergers, ouvriers et façonniers des villes, bourgs, et villages du royaume fixe les largeurs des pièces (Zanola, 2014 : 175). un'ordinanza, indirizzata (*rivolta*) ai mercanti di panni, ai mastri pannieri, ai serratori (ai *mercanti e produttori di saia*), agli operai e ai fabbricanti (ai *façonniers*) delle città, dei paesi (*borghi*) e dei villaggi del regno, stabiliva le larghezze (*la larghezza*) delle pezze di stoffa (TA, DeepL).

La difficulté réside probablement dans l'analyse des formes linguistiques qui ne considère pas le critère temporel : pour *manufacture*, l'équivalent en italien proposé par DeepL est *fabbrica*, un terme inadéquat en tant qu'équivalent au XVIII^e siècle. Pour désigner les établissements des premières productions industrielles de cette période, les termes adéquats

en italien seraient *manifattura*¹³ ou *opificio*¹⁴. D'autres termes qui n'ont pas été traduits correctement – *sergers* et *façonniers* – désignent des professions qui n'existent plus. À la différence de *marchands drapiers* et *maîtres drapans*, où la TA reconnaît probablement la base *drap* (it. *panno*) facilitant l'identification des équivalents, dans *serger* défini comme « un ouvrier, un marchand qui fabrique ou qui vend des serges » (*Encyclopédie, ad vocem*), la TA ne reconnaît pas la base *serge* (it. *saia*) : son identification, en combinaison avec la recherche concernant les activités de cette figure professionnelle, amène le post-éditeur ou la post-éditrice à la traduction *mercanti e produttori di saia*. Le terme *façonnier* désigne une catégorie d'artisans produisant des étoffes d'or, d'argent et de soie qui travaillaient pour le compte des marchands fabricants, tout en possédant une boutique et des instruments (Garden, 2008). Considérant qu'il s'agit d'un terme qui désigne une réalité du système de manufactures françaises, le post-éditeur ou la post-éditrice peut envisager d'employer le terme en français, éventuellement accompagné d'une note explicative.

Ce dernier exemple montre clairement le rôle essentiel d'une recherche approfondie concernant la terminologie propre d'un espace culturel. Notre analyse a montré que l'apport de la post-édition consiste notamment dans l'identification du concept désigné et, éventuellement, dans son explication à l'aide d'une paraphrase :

- [11] Brunot le disait déjà lorsqu'il rappelait la foule de noms des draps et étoffes de laine et de fil [...] (Zanola, 2014 : 175).

Brunot lo diceva (*affermava*) già quando ricordava (*nel ricordare*) la miriade di (*i numerosi*) nomi di panni e tessuti (*stoffè*) di lana e di filo (+ *di lino o canapa*) (TA, DeepL).

Brunot lo diceva (*affermava*) già quando ricordava (*nel ricordare*) la moltitudine di (*i numerosi*) nomi di panni e tessuti di lana e filati (*di filo di lino o canapa*) (TA, ChatGPT).

Un examen portant sur les concepts désignés s'avère indispensable notamment dans des cas où plusieurs termes du domaine présentent la même forme linguistique ; par exemple, *fil* dans le domaine textile : cherchant l'équivalent de *fil* dans l'extrait [11], le post-éditeur ou la post-éditrice – conscient.e du fait que le terme est utilisé dans le contexte français du XVIII^e siècle – peut identifier le concept en question *fil de lin ou de chanvre*¹⁵. Les équivalents de *fil* proposés par les outils de TA – *filo* et *filato* – pourraient être pertinents dans d'autres contextes ; dans ce texte, il est nécessaire d'employer le terme *filo* en précisant les matières utilisées pour sa fabrication : *filo di lino o canapa*.

Une fois les équivalents terminologiques validés ou corrigés, le post-éditeur ou la post-éditrice devrait s'assurer que leur emploi dans le texte soit cohérent. Observons, à titre d'exemple, les équivalents de *manufacture* proposés par DeepL :

¹³ Treccani, s.v. *manifattura*, URL : <<https://www.treccani.it/enciclopedia/manifattura/>> (consulté le 24-11-2025).

¹⁴ Treccani, s.v. *opificio*, URL : <[https://www.treccani.it/vocabolario/opificio_\(Sinonimi-e-Contrari\)/](https://www.treccani.it/vocabolario/opificio_(Sinonimi-e-Contrari)/)> (consulté le 24-11-2025).

¹⁵ Les autres concepts désignés dans le domaine textile par *fil* sont : « semi-produit utilisé dans la fabrication de textiles », « élément textile de très grande longueur », « fil de chaîne ».

- [12] La France manquait de manufactures capables de produire des bobines de soie régulières (Zanola, 2014 : 174).
Francia mancavano fabbriche (*manifatture*) in grado di produrre bobine di seta regolari (TA, DeepL).
- [13] Il imagine un moulin à faire l'organsin, qui est mis en service en 1751 et réalise la liaison entre la manufacture et le savoir du spécialiste (Zanola, 2014 : 175).
Progettò un mulino per la produzione di organza (*organzino*), che fu messo in funzione (*fu utilizzato per la prima volta*) nel 1751 e costituì un collegamento (*creò un legame*) tra la fabbrica (*manifattura*) e le conoscenze degli specialisti (TA, DeepL).
- [14] Le travail du textile trouve son essor aussi dans la tapisserie, et il serait intéressant de parcourir l'histoire de ces manufactures royales (Zanola, 2014 : 175).
L'arazzo (*La tappezzeria*) fu un altro settore in cui fiorì l'attività tessile e sarebbe interessante approfondire la storia di queste manifatture reali (TA, DeepL).
- [15] Pour notre curiosité terminologique, il suffira de nous arrêter sur l'exemple des manufactures d'Aubusson [...] (Zanola, 2014 : 175).
Per ~~la nostra~~ (+ *rispondere alle*) curiosità terminologica (*terminologiche*), basterà guardare (*osservare*) l'esempio delle manifatture di Aubusson [...] (TA, DeepL).

Les deux premières occurrences de *manufacture* sont traduites en italien par le terme *fabbrica*, qui n'est pas approprié (voir le commentaire de l'exemple [9]), tandis que pour les deux dernières occurrences, le traducteur automatique propose l'équivalent plus pertinent, *manifattura*. Nous supposons que ce changement est dû au contexte d'emploi de *manufacture* : utilisé de manière isolée, l'équivalent fournit est *fabbrica*, employé dans des syntagmes nominaux plus spécifiques renvoyant à des concepts anciens – *manufactures royales* et *manufactures d'Aubusson* – le logiciel propose le terme *manifattura*.

3.2.1.1 La création d'une base de données terminologique et l'analyse des termes du domaine

Les outils d'IA de type agent conversationnel peuvent être employés pour certaines tâches liées à la collecte des termes du domaine en vue de créer une base de données sur mesure pour un projet. Dans un premier temps, les agents conversationnels peuvent faciliter la construction d'un corpus : les critères de sélection des documents/des sites de la toile adoptés par l'IA n'étant pas connus, il est fondamental d'évaluer attentivement les sources proposées pour ne retenir que celles qui s'avèrent fiables. Ce premier ensemble de documents recueillis devrait ensuite être complété par d'autres textes sélectionnés *ad hoc*, afin de construire un corpus représentatif et équilibré.

Dans l'étape suivante, les agents conversationnels peuvent être exploités pour l'extraction terminologique à partir du corpus ainsi créé. Pour rendre compte de la performance de ChatGPT en matière, nous l'avons comparée avec celle du TermoStat Web 3.0, un logiciel conçu spécialement pour l'extraction terminologique (Drouin, 2003). Nous résumons, dans le tableau qui suit, les résultats de l'extraction terminologique ayant

pour objet les termes du domaine textile, effectuée par ces outils à partir de deux textes considérés : le nombre de candidats-termes générés automatiquement ainsi que celui de termes retenus après leur filtrage manuel. Rappelons qu’à la différence de ChatGPT, TermoStat Web 3.0 n’est pas en mesure d’effectuer une extraction terminologique basée sur une demande spécifique (ex. *Parmi les termes du domaine textile, est-ce que tu peux isoler ceux désignant les fibres textiles ?*) : les candidats-termes qu’il propose concernent toutes les unités pouvant avoir le statut de terme en fonction des critères statistiques et linguistiques préétablis¹⁶.

Tableau 2 - Résultats de l'extraction terminologique menée par ChatGPT et TermoStat Web 3.0

Extraction des candidats-termes	Article de « Vogue »	Zanola 2014
ChatGPT		
– Extraction des candidats-termes	35	32
– Termes retenus	22 (62%)	29 (90%)
TermoStat Web 3.0		
– Extraction des candidats-termes	25	72
– Termes retenus	13 (52%)	50 (69%)

Relativement à l’article de « Vogue », ChatGPT fournit un nombre plus élevé de candidats-termes que TermoStat Web 3.0, respectivement 35 et 25, tandis que le pourcentage des termes retenus est relativement proche : 62% pour ChatGPT et 52% pour TermoStat Web 3.0. La différence de performance entre ces logiciels est plus marquée dans le cas de l’extraction menée à partir de l’extrait de Zanola (2014), un texte caractérisé par un haut degré de spécialisation. TermoStat Web 3.0 fournit un nombre plus élevé de candidats-termes (72), parmi lesquels environ 70% sont des termes du domaine textile. Les candidats-termes proposés par ChatGPT – en se concentrant sur l’extraction des termes du domaine textile – sont en nombre moins important (32), mais presque tous (90%) sont validés comme termes du domaine. Même si l’agent conversationnel est capable d’effectuer des extractions portant sur un domaine, à l’heure actuelle, TermoStat Web 3.0 fournit des données plus précises, d’un point de vue quantitatif aussi bien que qualitatif.

Après l’identification des termes, le post-éditeur ou la post-éditrice devrait procéder à leur analyse en vue de comprendre la nature des concepts désignés, une condition *sine qua non* pour une évaluation correcte des termes proposés par la TA et une éventuelle recherche d’équivalents. En nous interrogeant sur les possibilités offertes dans ce sens par l’IA, nous avons attribué à ChatGPT une série de tâches portant sur le contenu conceptuel des termes du domaine textile. Les questions ont été formulées de manière à faciliter au maximum les tâches demandées. Nous confirmons aussi que les extraits contenant des termes analysés sont suffisamment clairs et larges ; ils permettent à un expert humain

¹⁶ Pour plus d’informations, nous renvoyons au *Guide de l'utilisateur*, URL : <https://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html> (consulté le 24-11-2025).

d'effectuer la tâche en question sans aucune difficulté¹⁷. En général, ChatGPT fournit plusieurs informations, dépassant largement le cadre de la tâche assignée ; nous précisons que nos analyses portent exclusivement sur la partie de la réponse qui concerne directement notre question.

Premièrement, nous avons demandé à ChatGPT¹⁸ de définir certains termes du domaine textile qui sont souvent utilisés de manière interchangeable dans le langage courant. L'objectif de cette expérimentation a été de vérifier si l'outil d'IA est capable de distinguer les usages dans la langue générale de ceux qui sont propres à un domaine de spécialité :

Question 1 : Est-ce que tu peux définir les concepts désignés dans le domaine textile par les termes *étoffe* et *tissu* ?

Réponse ChatGPT : *L'étoffe est un terme général qui désigne tout matériau textile flexible fabriqué à partir de fibres. Ce matériau peut être tissé, tricoté, feutré ou fabriqué par tout autre procédé qui permet de créer une structure textile.*

Le tissu est un sous-ensemble d'étoffe qui se réfère spécifiquement à un textile produit par un processus de tissage. Le tissage implique l'entrelacement de fils ou de fibres selon un motif régulier pour créer une surface plane.

Comparaison entre Étoffe et Tissu : l'étoffe est un terme plus général qui inclut le tissu ainsi que d'autres formes de textiles comme les tricots ou les feutres.

Question 2 : Quelle est la différence entre les fibres artificielles et les fibres synthétiques ?

Réponse ChatGPT : *Les termes fibres artificielles et fibres synthétiques désignent deux catégories distinctes de fibres utilisées dans l'industrie textile. Bien qu'ils soient parfois confondus dans le langage courant, ils se distinguent principalement par leur origine et leur méthode de fabrication.*

Les fibres artificielles sont dérivées de matières premières naturelles. La matière de base est généralement la cellulose, extraite de plantes telles que le bois ou le coton.

Les fibres synthétiques sont fabriquées à partir de polymères entièrement synthétiques. Ces polymères sont produits à partir de produits chimiques dérivés du pétrole ou du gaz naturel.

Nous avons observé que ChatGPT décrit les concepts désignés par les termes *étoffe*, *tissu*, *fibre artificielle* et *fibre synthétique* en conformité avec leur définition dans le domaine textile. Même si les descriptions proposées ne sont pas tout à fait irréprochables – au niveau de contenu, par exemple, les fibres ne sont jamais utilisées directement dans le tissage, le bois n'est pas une plante et, enfin, les fibres synthétiques ne se produisent pas à partir du gaz naturel –, la différence principale entre les deux concepts véhiculés est, dans les deux cas, bien indiquée.

Dans l'étape suivante, nous avons testé la capacité de ChatGPT de traiter la terminologie utilisée dans de courts extraits des textes de spécialité soumis à l'analyse. La première tâche a porté sur l'identification de la structure conceptuelle :

¹⁷ L'analyse des concepts spécialisés du domaine textile se base sur les données présentées dans Dankova (2023).

¹⁸ Nous précisons que les réponses ont été générées par ChatGPT en juin 2024.

Question 3 : Est-ce que tu peux générer la représentation des relations entre les concepts du domaine textile désignés par les termes présents dans cet extrait¹⁹ ?

Extrait :

Des tissus en velours et en relief

1/ “Bella Coola” - en coton et polyester - 3 coloris - environ 203 € le mètre en 132 cm de large - PIERRE FREY

2/ “Duomo”, en coton - viscose et polyester - 9 coloris - 181 € le mètre en 146 cm de large - NOBILIS

3/ “Luchino” - en coton - polyester et viscose - 8 coloris - 190 € le mètre en 135 cm de large - RUBELLI

4/ “Vulcain” - en trevira - 41 coloris - 111,10 € le mètre en 141 cm de large - LELIÈVRE

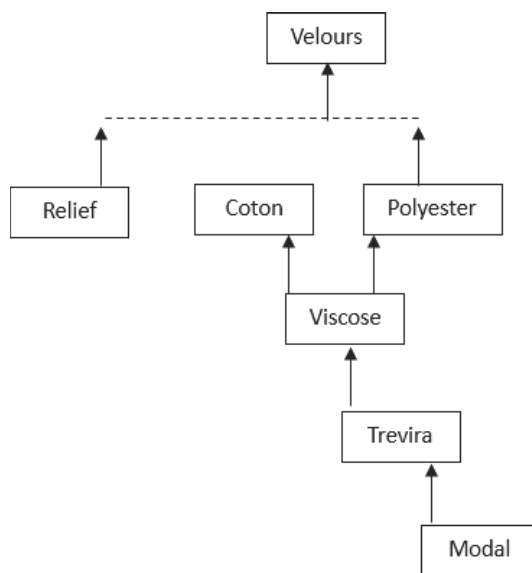
5/ “Aïme” - en coton et modal - 33 coloris - 114 € le mètre en 140 cm de large - MISIA

6/ “Felis” - en viscose - polyester et coton - 7 coloris - 84 € le m en 140 cm de large - CASAMANCE

Le tissu “Felis” (6) tapisse les chaises “Stella” - version contemporaine de l’assise XVIIIe - en bois noir - GUILLAUME ALAN

Réponse ChatGPT :

Fig. 1 - Représentation des relations conceptuelles générée par ChatGPT



L'extrait sélectionné contient des termes désignant un tissu (*velours*) et des fibres textiles d'origine naturelle (*coton*) ou chimique – des fibres artificielles (*viscose*, *modal*) aussi bien que synthétiques (*polyester*, *trevira*). L'organisation des concepts proposée par ChatGPT (Fig. 1) est tout à fait erronée : le concept de *trevira*, une fibre synthétique de polyester,

¹⁹ L'extrait provient de l'article *Tissu en velours : notre sélection haute en douceur !* de la revue « Elle », 16/08/2018. URL : <<https://www.elle.fr/Deco/News-tendances/Tissu-en-velours>> (consulté le 24-11-2025).

est lié aux concepts de *viscose* et de *modal*, deux fibres artificielles ; le concept de *viscose*, une fibre artificielle, est lié aux concepts de *coton* et de *polyester*, des fibres respectivement, naturelle et synthétique.

La tâche suivante visait à vérifier si ChatGPT est en mesure d'identifier des concepts véhiculés par des termes ayant la même forme linguistique au sein d'un domaine en synchronie :

Question 4 : Est-ce que tu peux définir les concepts désignés par les termes *fibre* présents dans ces deux extraits ?

Extrait 1 : Les fibres_1²⁰ proviennent des trois grands règnes naturels : végétal, animal et minéral, et du règne de l'homme : chimique. Chaque matière se présente sous des formes, des longueurs, des couleurs, des qualités différentes :

- fibre_2 longue de la tige du lin ;
- bourre de la fleur de coton ;
- poils des toisons des moutons ;
- fils du dévidage du cocon du ver à soie ;
- fils continus des textiles chimiques²¹.

Extrait 2 : Par exemple, on peut obtenir une face jersey en fil synthétique continu de type « coolmax », qui favorise l'évacuation de la sueur et une face bouclette en filés de fibres_3 au toucher doux type coton ou viscose. Les combinaisons sont multiples. On trouve sur le marché un tricot bouclette constitué d'un fil à âme de céramique guipé de fibres_1 de polyester qui permet de maintenir la température naturelle du corps par renvoi, d'un côté, de sa propre chaleur vers le corps et, de l'autre côté, par réflexion vers l'extérieur des rayons UV²².

Réponse ChatGPT :

Extrait 1 :

Dans ce premier extrait, le terme "fibre" désigne un élément fin et flexible provenant de différents règnes naturels (végétal, animal, minéral) ou artificiels (chimique).

Extrait 2 :

Dans ce deuxième extrait, le terme "fibre" est utilisé de manière plus spécifique. Il désigne un fil ou un brin de matière textile qui peut être constitué d'une seule fibre continue ou d'un assemblage de plusieurs fibres.

ChatGPT suppose que, étant donné qu'il s'agit de deux extraits différents, *fibre* désigne un concept dans le premier extrait et un autre dans le deuxième. En réalité, les deux extraits présentent *fibre_1* « matière textile » ; le premier aussi *fibre_2* « long filament textile d'origine végétale » et le deuxième *fibre_3* « élément textile discontinu de courte longueur ». Par rapport au premier extrait, ChatGPT cherche à définir *fibre_1*, en utilisant le terme *artificiel* de manière erronée en tant que synonyme de *chimique*, même si notre question 2 a montré que l'outil « connaît » bien la définition de *fibre artificielle*. La définition donnée pour l'extrait 2 ne porte sur aucun concept désigné par *fibre* :

²⁰ Pour une meilleure compréhension, nous avons marqué les différents termes par des numéros (*fibre_1*, *fibre_2*, *fibre_3*). Ces numéros étaient absents de l'extrait soumis à l'analyse par ChatGPT.

²¹ Anquetil (1977 : 94).

²² Weidmann (2010 : 163).

fibre ne désigne jamais un fil et il ne peut pas être déterminé par l'adjectif *continue* ; au contraire, le concept désigné par *fibre_3*, présent dans cet extrait, se caractérise par une forme discontinue.

Nous avons également testé la capacité de ChatGPT à analyser les concepts en tenant compte du critère temporel. Pour ce faire, deux extraits décrivant des concepts différents ayant la même forme linguistique en diachronie ont été analysés. Pour mettre en évidence la nécessité de considérer la période dans laquelle ces termes ont été employés, nous avons indiqué de manière explicite l'année de publication des ouvrages consultés :

Question 5 : Est-ce que tu peux définir les concepts désignés par les termes *fibre artificielle* présents dans ces deux extraits ?

Extrait 1, texte publié en 1929 : Ces fibres artificielles peuvent être fabriquées pour ressembler à celles du coton ou de la laine ou avoir un caractère propre ; elles peuvent être employées seules ou mélangées avant filature avec d'autres fibres textiles.

Depuis longtemps déjà, on a cherché à obtenir avec des fibres artificielles du fil ressemblant au fil de coton ou au fil de laine.

Déjà, en 1908, Beltzer obtenait un fil ressemblant à la laine en partant des déchets d'une bourre de soie artificielle²³.

Extrait 2, texte publié en 1999 : Dans cette famille, deux grands producteurs : Lenzing l'autrichien, toujours lui, qui produit le Lyocell Lenzing, tout simplement, et Courtaulds Fibres avec son fameux Tencel. Différents traitements lui confèrent un toucher « peau de pêche », « peau de chamois » tout comme on l'obtient sur un jean de coton bien « maltraité » pour le rendre délicieusement doux. Le Tencel a bien entendu les qualités coloristiques inhérentes à sa famille de fibres artificielles : il peut avoir des tons très intenses mais rend aussi très bien les pastels et les reflets²⁴.

Réponse ChatGPT :

Extrait 1 : Fibres artificielles : Ces fibres sont produites par l'homme à partir de matières premières naturelles, telles que la cellulose du bois ou du coton.

Extrait 2 : Fibres artificielles : Comme dans l'extrait 1, le terme "fibres artificielles" désigne des fibres synthétiques produites à partir de matières premières naturelles.

Ces extraits ont été sélectionnés pour analyser des contextes contenant le terme *fibre artificielle*, soit avant (extrait 1, 1929), soit après (extrait 2, 1999) le développement des fibres synthétiques dans les années 1930. Il s'agit de deux périodes temporelles dans lesquelles il est possible d'observer le changement de la définition de *fibre artificielle* : dans la première période, les fibres artificielles correspondaient à la catégorie des fibres manufacturées ou chimiques, car il n'y avait pas d'autre catégorie ; après l'introduction de fibres synthétiques, les fibres artificielles deviennent une catégorie de fibres manufacturées, celles qui sont fabriquées par transformation chimique de polymères naturels (Baum, Boyeldieu, 2018 : 257). Nous avons remarqué que la définition proposée par

²³ Luc (1929 : 539).

²⁴ Fauque, Bramel (1999 : 36).

ChatGPT pour *fibre artificielle* en 1929 est correcte, cependant, les informations fournies pour *fibre artificielle* dans l'extrait 2 (1999) ne sont pas pertinentes : le terme est analysé de manière erronée comme synonyme de *fibre synthétique*, en associant celui-ci aussi à *fibre artificielle* de l'extrait 1 (1929). La réponse de ChatGPT montre que l'IA n'a reconnu ni la pluralité de concepts ni la différence entre eux, même si notre question l'avait suggéré de manière claire.

Les mêmes extraits ont été utilisés aussi pour interroger ChatGPT sur les variantes synonymiques du terme *fibre artificielle* dans les deux périodes envisagées :

Question 6 : Quelles sont les variantes synonymiques du terme *fibre artificielle* de l'extrait 1 ? Quelles sont les variantes synonymiques du terme *fibre artificielle* de l'extrait 2 ?

Réponse ChatGPT :

Extrait 1 : Dans cet extrait de 1929, les fibres artificielles sont évoquées dans un contexte historique de développement des fibres manufacturées pour imiter les fibres naturelles comme le coton ou la laine. Les variantes synonymiques de "fibre artificielle" à cette époque incluraient : fibre synthétique, soie artificielle, rayonne, fibre manufacturée.

Extrait 2 : Dans l'extrait de 1999, les fibres artificielles incluent des matériaux plus modernes et techniquement avancés, reflétant des progrès significatifs dans la chimie des polymères et des processus de production. Les variantes synonymiques incluraient : fibre synthétique, fibre manufacturée, fibre régénérée, fibre cellulosique.

À la différence de la question précédente, dans ce cas, la réponse de ChatGPT semble prendre en charge le facteur temporel. Cependant, les seules variantes synonymiques qui sont admissibles sont *soie artificielle* et *fibre manufacturée* pour *fibre artificielle* en 1929. Les explications données concernant la proposition des autres variantes ne sont pas appropriées²⁵. Par exemple, relativement au terme *fibre synthétique* indiqué comme variante synonymique de *fibre artificielle* en 1999, ChatGPT précise qu'il s'agit d'« un synonyme courant pour toutes les fibres produites chimiquement » : nous en pouvons déduire que, dans cette réponse, ChatGPT fait une confusion entre langue de spécialité et langue générale.

Notre expérimentation concernant l'emploi des agents conversationnels dans le processus de post-édition a montré que ces outils d'IA peuvent être utilisés surtout lors de deux tâches : la recherche de textes pour la construction d'un corpus – pour la collecte des termes ou, éventuellement, pour se documenter sur un domaine de spécialité – et pour l'extraction terminologique. Relativement à cette dernière, l'avantage principal de l'agent conversationnel tel que ChatGPT est celui de fournir une liste de candidats-termes en fonction d'une demande spécifique. Néanmoins, un extracteur terminologique tel que TermoStat Web 3.0 est en mesure de fournir des données plus complètes et précises : son emploi s'avère être donc plus efficace.

²⁵ Pour simplifier la présentation des données, les justifications des variantes synonymiques proposées ne figurent pas dans la réponse de ChatGPT citée ici.

En ce qui concerne l'analyse de la dimension conceptuelle de la terminologie, les essais effectués avec ChatGPT n'ont pas porté de résultats satisfaisants. Même si, dans certains cas, l'outil fournit des informations correctes, celles-ci semblent être données de manière fortuite : considérons, par exemple, la description correcte de *fibre artificielle* en réponse à la question 2 et des informations contradictoires présentées à ce sujet dans les réponses aux questions 5 et 6.

4. Conclusion

Notre étude vise à cerner et décrire les étapes lors desquelles l'emploi des outils de TA et d'IA s'avère adéquat en termes d'accélération et d'efficacité du processus de traduction spécialisée. Conscientes des enjeux relatifs à la performance de ces outils, influencée par une disponibilité des données très différente – en fonction, principalement, de la combinaison linguistique, du domaine de spécialité, de la typologie textuelle et de son degré de spécialisation –, nous avons cherché à mettre en évidence d'un point de vue méthodologique les apports et les limites possibles de leur emploi. Les agents conversationnels peuvent être utilisés – en plus de la génération d'une prétraduction – aussi pour la construction des corpus spécialisés ou pour l'extraction terminologique ; cependant, les données fournies doivent être toujours analysées avec attention pour n'en retenir que celles qui sont pertinentes au projet de traduction concerné.

L'évaluation de la qualité des traductions automatiques a relevé des difficultés liées notamment au traitement de la terminologie, dues à la complexité du terme (type de contexte, degré de spécialisation, évolution diachronique). Tandis que la forme linguistique du terme peut être traitée par l'IA, son contenu conceptuel semble ne pas être accessible : les réponses fournies par l'agent conversationnel ChatGPT par rapport à la nature des concepts désignés par certains termes clés du domaine textile – des termes pris isolément ou insérés dans de courts extraits – ont montré que les données générées ne peuvent pas être considérées comme fiables. Les réponses analysées présentent souvent des lacunes²⁶ et, dans certains cas, la différence entre le sens dans la langue générale et celui dans la langue de spécialité n'est pas prise en compte. C'est pour cette raison

²⁶ En raison de l'amélioration constante des outils d'IA, nous avons testé la performance du même agent conversationnel en novembre 2025, c'est-à-dire 17 mois après le premier test, en lui fournissant les mêmes questions. Nous pouvons confirmer que les réponses données ne sont pas précises et ne peuvent pas être employées pour l'étude de l'organisation conceptuelle du domaine considéré :

- Question 1 et 2 : les réponses présentent des imprécisions (Réponse ChatGPT : *Fibres synthétiques - Matières premières : artificielles/pétrochimiques* ; Commentaire : les fibres synthétiques ne sont pas fabriquées à partir des matières artificielles).
- Question 3 : l'organisation conceptuelle reste erronée (Réponse ChatGPT : les concepts du domaine textiles sont aussi *Coloris/largeur/prix, Usage mobilier* ; Commentaire : les couleurs, la largeur, les prix et l'usage des textiles ne sont pas considérés des concepts du domaine).
- Question 4 : l'identification des concepts désignés par la même forme linguistique – *fibre* – n'est pas correcte (Réponse ChatGPT : *Définition commune : Une fibre est un élément filiforme très fin, naturel ou fabriqué, destiné à être transformé en fil. Elle constitue la base des matériaux textiles et leur confère des*

que, pour la réalisation d'une post-édition de qualité, les compétences en terminologie jouent un rôle de premier plan : ce n'est qu'en se fondant sur ses connaissances terminologiques que le post-éditeur ou la post-éditrice peut fournir un texte traduit qui soit précis, cohérent d'un point de vue terminologique et approprié à la culture de la langue d'arrivée et aux besoins des clients. Nos observations ont porté à la proposition d'une modélisation du processus de post-édition des textes spécialisés, ayant au cœur le traitement de la terminologie. Les différentes étapes sont résumées dans une liste de contrôle présentée dans la Fig. 2.

Fig. 2 - Liste de contrôle pour la post-édition des textes de spécialité

Post-édition des textes de spécialité		
Liste de contrôle		
1. Macrostructure		
1.1. l'organisation générale du texte	☐	
1.2. les conventions typographiques	☐	
1.3. la présence des figures	☐	
2. Microstructure		
2.1. l'orthographe lexicale	☐	
2.2. les aspects grammaticaux	☐	
2.3. les choix lexicaux	☐	
2.4. les choix terminologiques :		
2.4.1. les besoins des destinataires	☐	
2.4.2. la typologie textuelle et son degré de spécialité	☐	
2.4.3. la pertinence des choix terminologiques considérant la variation	☐	
2.4.4. la présentation des concepts propres d'une culture	☐	
2.4.5. la cohérence terminologique	☐	

Références bibliographiques

- Alì M. et al., *Qualité de la traduction automatique dans le domaine de la mode durable, du changement climatique et de l'environnement*, in « De Europa », Special Issue, 2023, pp. 235-256.
- Anquetil J., *Le tissage. Encyclopédie contemporaine des métiers d'art*, Paris, Dessain et Tolra, 1977.

propriétés mécaniques, esthétiques ou fonctionnelles ; Commentaire : en terminologie, il n'est pas acceptable de proposer une même définition pour des concepts différents).

- Question 5 : les définitions de *fibres artificielles* dans les deux extraits (1929, 1999) sont incomplètes (Réponse ChatGPT : *Définition adaptée au texte de 1999 - Une fibre artificielle est une fibre textile fabriquée à partir de cellulose naturelle, transformée chimiquement en une solution puis régénérée sous forme de filament continu ou de fibres discontinues* ; Commentaire : les fibres artificielles peuvent être fabriquées aussi à partir d'autres matières telles que les polysaccharides).
- Question 6 : les variantes synonymiques de *fibres artificielles* dans les deux extraits (1929, 1999) ne sont pas correctes (Réponse ChatGPT : [extrait 1929] *soie artificielle, rayonnel/rayon, soie chimique, fibre de cellulose régénérée, fibre manufacturée (au sens ancien)*, [extrait 1999] *fibre cellulosique régénérée, fibre cellulosique artificielle, fibre de cellulose régénérée, MMCF / man-made cellulosic fibre, viscose/modal/lyocell (noms de fibres appartenant à cette famille)* ; Commentaire : les seules variantes synonymiques correctes sont *soie artificielle* et *fibre manufacturée* pour *fibres artificielles* dans l'extrait de 1929).

- Baum M., Boyeldieu C., *Dictionnaire encyclopédique des textiles*, Paris, Eyrolles, 2018.
- Calvi S., Dankova K., *Industrie de la langue et formation des traducteurs spécialisés*, in « Revue Traduction et Langues », 20 (1), 2022, pp. 8-21.
- Dankova K., *Les fibres textiles entre synchronie et diachronie : études terminologiques*, Berne, Peter Lang, 2023.
- Diderot D., Le Rond D'Alembert J., *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers par une société de gens de lettres*, 35 vol., Paris, Panckoucke, 1751-1772. URL : <<http://enccre.academie-sciences.fr/encyclopedie/>>, consulté le 24-11-2025.
- Drouin P., *Term extraction using non-technical corpora as a point of leverage*, in « Terminology », 9/1, 2003, pp. 99-117.
- Fauque C., Bramel S., *Une seconde peau : fibres et textiles d'aujourd'hui*, Paris, Éditions Alternatives, 1999.
- Flöter-Durr M., *Les limites épistémologiques des techniques numériques actuelles de l'intelligence artificielle en traduction/Epistemological limits of current digital techniques of artificial intelligence in translation*, in « Lebende Sprachen », 67(1), 2022, pp. 4-44.
- Garden M., *Ouvriers et artisans au XVIII^e siècle*, in R. Favier, L. Fontaine (éds.), *Un historien dans la ville*, Paris, Éditions de la Maison des sciences de l'homme, 2008, pp. 87-112. URL : <<https://books.openedition.org/editionsmsh/9958>>, consulté le 24-11-2025.
- Luc P., *Le tissage de la soie artificielle*, Paris, L'édition textile, 1929.
- Organisation internationale de la Francophonie [OIF], *La langue française dans le monde*, Paris, Gallimard, 2022.
- Popel M. et al., *Transforming machine translation: a deep learning system reaches new translation quality comparable to human professionals*, in « Nature Communications », 11, 2020. URL : <<https://doi.org/10.1038/s41467-020-18073-9>>, consulté le 24-11-2025.
- Robert A.-M., *La post-édition : l'avenir incontournable du traducteur ?*, in « Traduire », 222, 2010, pp. 137-144.
- Schumacher P., *Avantages et limites de la post-édition*, in « Traduire », 241, 2019, pp. 108-123.
- Villa M.L. et al., *Langages et savoirs : intelligence artificielle et traduction automatique dans la communication scientifique*, in « De Europa », Special Issue, 2022, pp. 107-127.
- Weidmann D., *Aide-mémoire Textiles techniques*, Paris, Dunod, 2010.
- Zanola M.T., *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*, Paris, L'Harmattan, 2014.
- Zanola M.T., *Che cos'è la terminologia?*, Roma, Carocci, 2018.

Un mare di termini: storia di un progetto terminologico come catalizzatore di collaborazioni interdisciplinari

ANA LOURDES DE HÉRIZ, JOACHIM GERDES, VERONICA MONTICELLI,
MARTINA ODELLO, VELIA PASCALE, MICAELA ROSSI, LAURA SANTINI

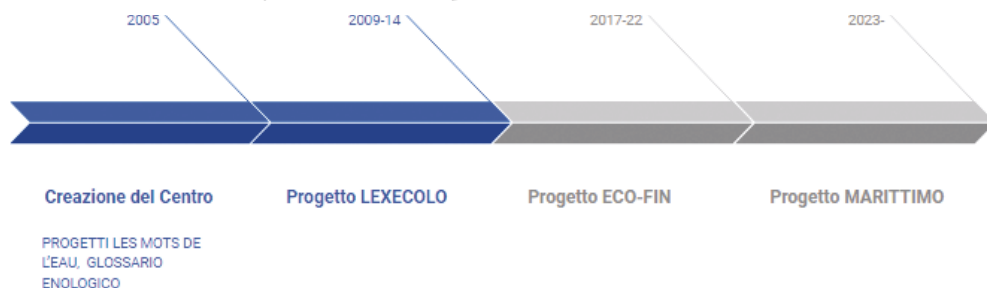
1. Presentazione del contesto: le attività del Ce.R.Te.M. dell'Università di Genova

Fondato nel 2005 presso il Dipartimento di Lingue e Culture moderne dell'Università di Genova, il Centro di Ricerca in Terminologia Multilingue Ce.R.Te.M. è un centro di ricerca interdipartimentale (vi partecipano, infatti, anche i Dipartimenti di Scienze Politiche e di Italianistica, Romanistica, Antichistica, Arti e Spettacolo), che ha come principale missione lo studio delle terminologie specialistiche in prospettiva comparatistica e interlinguistica. Il Ce.R.Te.M. coordina ricerche in ambito terminologico, redazione di glossari (Piccardo 2009), effettua attività di formazione alla terminologia nelle lingue attivate presso il Dipartimento, ovvero italiano, inglese, francese, spagnolo, tedesco, russo, svedese, portoghese, arabo, polacco, cinese.

Dalla sua fondazione, il Ce.R.Te.M. ha stretto accordi di collaborazione scientifica con gruppi e reti di ricerca nell'ambito terminologico, a livello nazionale (Ass.I.Term, Rete REI) e internazionale (REALITER, TermCoord, WIPO), realizzando inoltre progetti *ad hoc* con enti ed organismi del territorio ligure (Accademia della Marina Mercantile). Fanno parte del Centro ricercatrici e ricercatori, dottorande e dottorandi, e studentesse e studenti tirocinanti.

Tra i progetti condotti dal Ce.R.Te.M., l'elaborazione di glossari multilingui è l'attività principale; il sito del Centro (URL: <<https://certem.unige.it>>) contiene i principali repertori terminologici elaborati a partire dalla fine degli anni Novanta – precedenti in alcuni casi anche alla formalizzazione del Centro stesso.

Fig. 1 - Linea del tempo delle attività del Ce.R.Te.M.



Tra tutti i repertori, in ordine cronologico, si possono evidenziare:

- il *glossario enologico* (1998-2018), disponibile in italiano, francese, inglese, russo, spagnolo, tedesco, che riunisce tutta la terminologia legata alla vinificazione e alla degustazione del vino;
- il glossario *Les mots de l'eau* (2002-2003), nato da un progetto INTERREG e disponibile in italiano, spagnolo, francese, arabo (Bianchini *et al.*, 2008). Contiene schede terminologiche relative alla terminologia dell'acqua vista attraverso la prospettiva di diverse discipline (ingegneria, etnologia, biologia...), in ottica multilingue e multiculturale (Rossi, 2007; Grimaldi *et al.*, 2022; Zanola, 2017);
- il glossario sulle *energie rinnovabili* (2009-2014), anche questo disponibile in italiano, inglese, francese, tedesco, russo;
- il glossario sulla *banca e finanza* (2017-2022), disponibile in italiano, francese, inglese, spagnolo e tedesco.

2. Il glossario marittimo: un progetto strategico

Nel 2023, i membri del Ce.R.Te.M. decidono di intraprendere un nuovo progetto pluriennale, anche in ragione della collaborazione in atto con l'Accademia della Marina Mercantile, e selezionano come dominio di indagine il linguaggio marittimo nelle sue diverse declinazioni. Tale selezione è ulteriormente giustificata da una ragione istituzionale: l'esistenza all'interno dell'Università di Genova di un centro strategico, il Centro del Mare (URL: <<https://mare.unige.it/>>), che riunisce attività di formazione, ricerca e Terza Missione sulle tematiche marittime:

Fig. 2 - Il sito del Centro del Mare UniGE



Il Centro del Mare riunisce le discipline che nel mare trovano il proprio campo di indagine e che formano **competenze fortemente specialistiche e multidisciplinari**.

Qui più di **400 docenti e ricercatori UniGe** svolgono didattica e ricerca sul mare.

L'obiettivo di costituire un glossario multilingue sulla base degli assi di ricerca e dei sottodomini interessati dalle attività del Centro del Mare dell'Università di Genova appare

quindi strategico per poter raccogliere competenze ed esperienze dalle diverse strutture dell'Ateneo. Le figure attivamente coinvolte in questo progetto sono numerose, e hanno competenze diversificate: se partecipano linguiste, linguisti, traduttrici e traduttori afferenti al Ce.R.Te.M., collaborano fin dall'inizio anche ricercatrici e ricercatori con esperienze in ambiti molto vari (diritto, scienze della vita, ingegneria navale e nautica, storia e culture del mare...). Il Centro del Mare di UniGE riunisce, infatti, più di 400 unità tra docenti, ricercatrici e ricercatori che svolgono attività didattica, di ricerca o di Terza Missione intorno ai temi del mare. In questo contesto, il progetto assume una dimensione del tutto istituzionale, e le diverse professionalità come vedremo partecipano alle varie fasi del lavoro terminologico.

2.1 La metodologia di lavoro

La partecipazione ad un gruppo di lavoro ampio ed eterogeneo come quello che gravita intorno al Centro del Mare ha condotto fin dall'inizio del progetto all'elaborazione di una metodologia molto interdisciplinare (Bachimont, 2020; Cabré, 1991):

1. in una prima fase, sono stati identificati diversi sottodomini con l'aiuto di figure esperte del Comitato Tecnico Scientifico (CTS) del Centro del Mare. Per ogni sottodominio, coloro che hanno precise competenze settoriali sono stati invitati a formulare i concetti chiave che potessero fornire una base di lavoro per stabilire la struttura concettuale del glossario. I sottodomini, come si evince dall'immagine che segue, sono numerosi e diversificati, sia nell'ambito delle scienze fisiche e naturali che nell'ambito delle tecnologie, che nelle discipline delle scienze umane e sociali:

Fig. 3 - *Sottodomini identificati*

MEZZI MARINI, ROBOTICA E SUBACQUEA
 Mezzi di superficie e sottomarini
 Cantieristica
 Robotica, sistemi e tecnologie subacquee
SISTEMI DI TRASPORTO MARITTIMO, LOGISTICA ED ECONOMIA PORTUALE
INFRASTRUTTURE COSTIERE E OFFSHORE
AMBIENTE MARINO E COSTE
BIOLOGIA, ECOLOGIA E BIOTECNOLOGIE MARINE
RISORSE ENERGETICHE DAL MARE
TURISMO E CROCIERE
SPORT DEL MARE E ATTIVITÀ MOTORIA LUDICO-RICREATIVA IN AMBIENTE MARINO
DIRITTO DEL MARE E DELLA NAVIGAZIONE
STORIA E CULTURE DEL MARE

2. per ogni dominio, il lavoro di spoglio dei concetti chiave è stato frutto di un'interlocuzione tra i due macrogruppi (ovvero le figure esperte in linguistica e quelle con competenze in settori disciplinari specifici), e ha condotto all'elaborazione di tabelle riassuntive per ogni sottodominio (si veda l'immagine seguente per il sottodominio "risorse energetiche dal mare"):

Fig. 4 - *Tabella riassuntiva per sottodominio*

RISORSE ENERGETICHE DAL MARE	
– Energie rinnovabili dal mare (energia dal moto ondoso, energia dalle maree, eolico off-shore, energia talassotermica, energia da gradiente salino, geotermia marina, Sea Water Air Conditioning, impatto ambientale energia dal mare). – Marine Spatial Planning. – Tecnologie di conversione dell’energia marina (wave energy converter, tidal energy converters, turbine eoliche offshore, microbial fuel cells). – Accumulo energetico marino (produzione di idrogeno off-shore, CAES sottomarino, autoproduzione per acquacoltura autonoma, ricarica mezzi autonomi marini, rigassificazione di GNL). – Trasporto e distribuzione di energia in mare; Integrazione architetture in corrente alternata e continua off-shore.	– Energie rinnovabili dal mare – Marine Spatial Planning – Tecnologie di conversione dell’energia marina – Accumulo energetico marino – Distribuzione di energia in mare

3. le persone esperte nei diversi domini sono state di supporto anche nella selezione delle fonti di consultazione, così come in quella dei testi per la prima costituzione del corpus di analisi. Sulla base delle loro indicazioni, è stata effettuata una prima estrazione terminologica dal testo della norma UNCLOS (Convenzione delle Nazioni Unite sul Diritto del Mare) del 1982, disponibile nelle sei lingue ufficiali dell’ONU (inglese, francese, arabo, cinese, russo e spagnolo), ma anche in tutte le lingue dell’Unione europea in seguito alla traduzione ufficiale necessaria alla ratifica (per l’Italia, la Convenzione è stata ratificata con la Legge 689/1994);
4. per l’estrazione terminologica, sono state utilizzate in una fase preliminare le piattaforme Sketch Engine¹ (funzione Keywords) e Termostat². Le liste dei candidati termini estratti sono state sottoposte a verifica manuale, comparate nelle diverse lingue oggetto del glossario e suddivise nei sottodomini identificati a valle del confronto con le persone esperte di riferimento. Tra i molti sottodomini, il gruppo di lavoro del Ce.R.Te.M. si è focalizzato su due in particolare, ovvero il “diritto del mare e della navigazione” e la “biologia, ecologia e biotecnologie marine”. Su quest’ultimo si concentrerà la nostra analisi nelle pagine seguenti;
5. la possibilità di disporre di una versione multilingue ufficiale della Convenzione ha permesso di effettuare fin dall’inizio una comparazione interlinguistica dei candidati termini, al fine di verificare eventuali variazioni denominative o concettuali. Si veda ad esempio la tabella seguente, in cui vengono comparate le diverse specie ittiche oggetto di indicazioni particolari all’interno della normativa:

¹ URL: <<https://auth.sketchengine.eu/>> (consultato il 19-11-2025).

² URL: <TermoStat Web> (consultato il 19-11-2025).

Fig. 5 - Tabella di comparazione interlinguistica sulla base della norma UNCLOS

BIOLOGIA, ECOLOGIA E BIOTECNOLOGIE MARINE					
ITALIANO	INGLESE	FRANCESE	SPAGNOLO	TEDESCO	RUSSO
adocetus	adocetus	adocetus	paparda	adocetus	adocetus
alalunga	alalunga	alalunga	alalunga	alalunga	alalunga
albacares	tuna	albacares	albacares	Gelbflossen- thunfisch	albacares
albidus	albidus	albidus	albidus	albidus	albidus
alletteratus	alletteratus	alletteratus	alletteratus	alletteratus	alletteratus
alopiidae/ alopidi	alopiidae	alopiidae	alopiidae	Alopiidae	alopiidae
anadromi/ banco anadromi	anadromous stock	anadromes	anádromos	Anadrome/ anadromer Bestand	анадромные
angustirostris	angustirostris	angustirostris	angustirostris	angustirostris	angustirostris

6. una volta verificati i candidati termini, è stato definito un modello di scheda sulla base delle norme ISO (1087/2020), in seguito lievemente adattato agli obiettivi del progetto (si veda il paragrafo successivo) ed il lavoro di schedatura ha avuto inizio. Ad oggi, sono stati realizzati due glossari di circa 100 termini ognuno nell'ambito delle biodiversità marina (IT/ES/DE) e un glossario di circa 100 termini nell'ambito del diritto del mare (IT/EN/RU)³.

3. Sfide e difficoltà

3.1 I termini tra lingue e saperi

Il progetto ha da subito evidenziato la necessità di mediare tra le rappresentazioni della conoscenza nei discorsi esperti e le prassi dell'analisi linguistica e terminologica. La definizione stessa di "concetto" ha generato nelle fasi iniziali una difficoltà di comunicazione tra le varie componenti del gruppo di lavoro: spesso, infatti, ciò che per la figura specialista era un concetto-chiave in realtà rappresentava per la componente terminologica del gruppo un intero microdominio (è il caso, ad esempio, di "energie rinnovabili dal mare" nello schema di cui al paragrafo 2.1.).

I due domini selezionati per la prima fase della schedatura hanno inoltre posto problemi diversi. Nel caso del "diritto del mare", i concetti oggetto di schedatura vengono definiti in maniera univoca all'interno della norma UNCLOS e delle norme correlate. Il testo giuridico ha in questo senso un valore normativo, ed è possibile estrarre definizioni coerenti dal testo giuridico. La definizione dei termini sarà in questo caso coerente con

³ Le autrici di questi glossari sono Veronica Monticelli e Martina Odello (biodiversità) e Velia Pascale (diritto del mare).

il dominio analizzato, e la circolarità tra contesto, testo e concetti è garantita dalla natura stessa del corpus selezionato.

La terminologia della “biodiversità” si collega invece strettamente a categorizzazioni che appartengono ad altre discipline, già da lungo tempo normate e stabilizzate. È il caso della biologia marina, per la quale le tassonomie delle specie e dei generi sono ormai attestate e convalidate (seppure in costante evoluzione). In questo caso, chi si occupa della terminologia dovrà rispettare la categorizzazione dei saperi messa in atto nelle comunità specialistiche, anche se tale tassonomia varia nel tempo e può in alcuni casi essere particolarmente complessa e stratificata. Non sarà quindi possibile definire ad esempio un *Thunnus alalunga* come un semplice “pesce”, ma si dovrà permettere a chi legge di ricostruire la classificazione del *Thunnus alalunga* nella sua completezza, come si evince dall’immagine seguente, in cui il modello di schedatura è stato adattato a questa necessità con l’aggiunta del campo “relazioni concettuali”, che esplicita la tassonomia scientifica⁴:

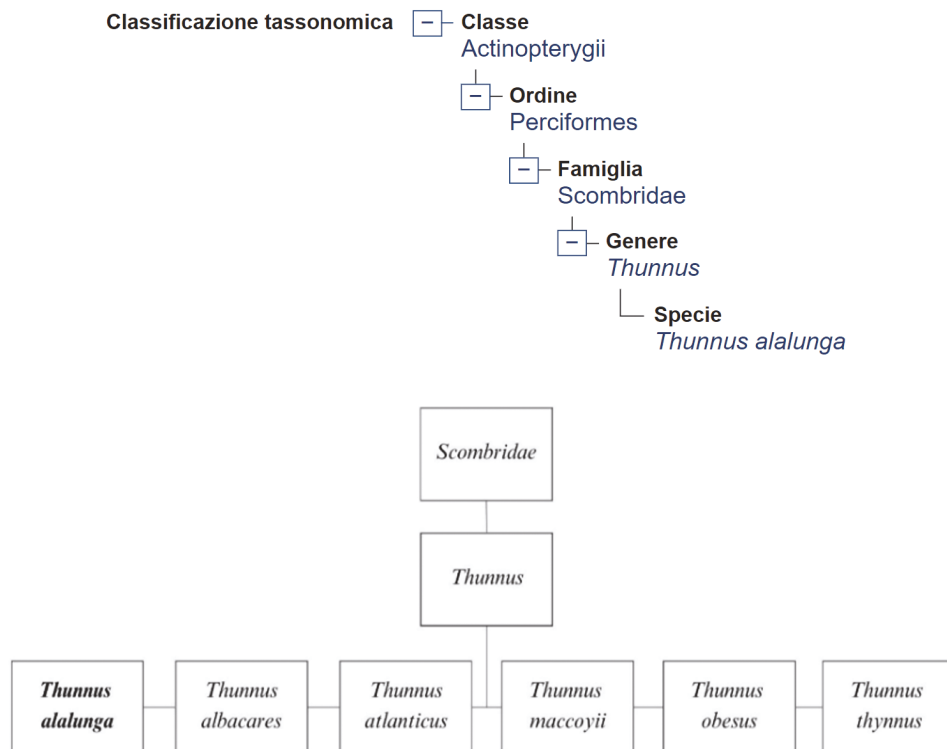
Fig. 6 - Esempio di modello per la schedatura

Lemma	<i>Thunnus alalunga</i>																
Sinonimi	Aa-lunga, Alalunga, Alalengu, Alalunga, Alilonga, Lalonga, Licia, Tonno, Tonno bianco (1)(2) Nomi dialettali <table><tr><td>GENOVA</td><td>Aa-lunga</td></tr><tr><td>ROMA</td><td>Alalunga</td></tr><tr><td>NAPOLI</td><td>Alalunga</td></tr><tr><td>TARANTO</td><td>Alalunga</td></tr><tr><td>GALLIPOLI</td><td>Alilonga</td></tr><tr><td>VENEZIA</td><td>Alalunga</td></tr><tr><td>CAGLIARI</td><td>Alalengu, Licia</td></tr><tr><td>SICILIA</td><td>Alalunga</td></tr></table> (3) Abbreviazioni: ALB (4)	GENOVA	Aa-lunga	ROMA	Alalunga	NAPOLI	Alalunga	TARANTO	Alalunga	GALLIPOLI	Alilonga	VENEZIA	Alalunga	CAGLIARI	Alalengu, Licia	SICILIA	Alalunga
GENOVA	Aa-lunga																
ROMA	Alalunga																
NAPOLI	Alalunga																
TARANTO	Alalunga																
GALLIPOLI	Alilonga																
VENEZIA	Alalunga																
CAGLIARI	Alalengu, Licia																
SICILIA	Alalunga																
Relazioni concettuali	- Classe: <i>Actinopterygii</i> - Ordine: <i>Perciformes</i> - Famiglia: <i>Scombridae</i> - Genere: <i>Thunnus</i> - Specie: <i>Thunnus alalunga</i> (5)																

Ulteriore elemento di complessità, le modalità di classificazione tassonomica non sempre corrispondono all’organizzazione dei concetti che può interessare chi si occupa di terminologia (Di Nunzio *et al.*, 2014, per un esempio interessante), come si evince dall’esempio che segue, nel quale la classificazione tassonomica non permette di differenziare le specie di *Thunnus*, valorizzate invece nella rappresentazione ad albero alla base della schedatura terminologica:

⁴ La necessità di adattare la redazione terminografica alle tassonomie scientifiche esistenti ha condotto alla scelta di adottare come lemma *pivot* il termine della classificazione linneana, per poi inserire nel campo sinonimi le diverse denominazioni vernacolari delle specie analizzate.

Fig. 7 - Esempio di classificazione



Questa fase del progetto ha permesso di condurre una riflessione sul legame tra ontologie e terminologie, che ancora oggi si rivela uno dei punti fondamentali della riflessione teorica in terminologia⁵, e di cercare una conciliazione tra l'approccio di organizzazione delle conoscenze propria delle discipline del *knowledge management* e l'approccio terminologico, che non può prescindere dall'espressione linguistica dei concetti.

Le difficoltà di classificazione emerse in precedenza sono state confermate dalla fase di redazione delle definizioni terminografiche, per le quali è stato necessario consultare il personale del Dipartimento DISTAV (Dipartimento di Scienze della Terra, dell'Ambiente e della Vita) dell'Università di Genova e la SIBM (Società Italiana di Biologia Marina).

Le definizioni sono state calibrate sulla base del modello aristotelico (*per genus et differentia*), compito facilitato dalla struttura tassonomica del microdominio specifico. All'interno dello stesso genere, le definizioni delle diverse specie sono state così comparsate ai fini dell'estrazione dei tratti differenziali pertinenti (forma del corpo, posizione delle pinne, colori, ecc.). Si noti come, nella tabella che segue, i tratti comuni sono stati evidenziati in giallo e hanno fornito materiale per la definizione del genere, mentre i tratti in verde sono stati identificati come i tratti pertinenti di ciascuna specie:

⁵ Si veda a questo proposito la prima sezione del volume collettaneo diretto da Faber e L'Homme (2022).

Fig. 8 - Tabella preparatoria per la definizione dei concetti

Albidus	Audax	Angustirostris	Belone	Georgei	Pfluegeri	Tetrapturus
Diagnostic Features : Body elongate and fairly compressed. Bill stout and long, round in cross section; nape fairly elevated, right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth. Two dorsal fins, the first with 38 to 46 rays, usually with a rounded anterior lobe, higher than body depth anteriorly, then	Diagnostic Features : Body elongate and fairly compressed. Bill stout and long, round in cross section; nape fairly elevated, right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth. Two dorsal fins, the first with 37 to 42 rays, usually with a pointed anterior lobe, higher than (or occasionally equal to) body	Diagnostic Features : Body elongate and fairly compressed. Bill short and slender, round in cross section; lower jaw shorter than upper jaw, but still projecting; nape nearly straight; right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file- shaped teeth. Two dorsal fins, the first with 45 to 50 rays and with a pointed	Diagnostic Features: Body elongate and fairly compressed. Bill rather short and slender, round in cross section; nape almost straight; right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth. Two dorsal fins, the first with 39 to 46 rays and a rounded anterior lobe higher than body depth anteriorly, the fin	Diagnostic Features : Body fairly robust and compressed. Bill long and slender, round in cross section; nape moderately humped; right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth. Two dorsal fins, the first with 43 to 48 rays, higher than the maximum body depth anteriorly, and lower	Diagnostic Features: Body elongate and remarkably compressed, its depth very low. Bill slender and rather long, round in cross section; nape nearly straight, right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth. Two dorsal fins, the first with 44 to 50 rays, and a rounded anterior	A fish genus characterised by an elongate and compressed body; usually with a slender bill, round in cross section; right and left branchiostegal membranes completely united to each other, but free from isthmus; no gillrakers; both jaws and palatines with small, file-like teeth; two anal fins, the second very similar in size to the second dorsal; slender pelvic fins; caudal peduncle with a

3.2 La gestione della variazione

La presenza di una classificazione tassonomica condivisa dal punto di vista scientifico non esclude, tuttavia, la realtà della variazione terminologica, che nel caso del microdominio oggetto del nostro studio si declina su diversi assi.

Il più evidente ad una prima analisi è senza dubbio il livello della variazione diatopica. Le nominazioni delle specie analizzate sono, infatti, estremamente variabili su base geografica, e si possono reperire variabili nazionali, ma anche regionali e addirittura locali, come nei due esempi che seguono:

Fig. 9 - Esempio di gestione della variazione diatopica

Término	<i>Tetrapturus albidus</i>
Sinónimos	- Kajikia albida (denominación correcta reclasificada post UNCLOS) - Aguja blanca (Cuba, Nicaragua, Puerto Rico, España, Uruguay, Venezuela) - Aguja blanca del Alántico (España) - Aguja de costa (España) - Aguja de paladar (Cuba) - Alfiler (España) - Alton (España) - Altón (España) - Blanca (Cuba) - Cabezona (Cuba) - Cometa (España) - Marlin (UNCLOS) - Marlin blanca (España) - Marlin blanco (México) - Pez aguja (España) [adaptado de (1)] Abreviaturas: WHM (FAO) (2)

Lemma	<i>Euthynnus alletteratus</i>																																		
Sinonimi	Alacurza, Aleterato, Alletterato, Allittirato, Allittiratu, Carcana, Cuvarito, Cuvaritu, Leterato, Letterato, Lettirado, Litterato, Littiratu, Nzirru, Palametidd, Pizziteddu, Sanguinaccio, Scampírru, Tonnella, Tonnetto, Tonnina, Tunna, Tunnina (1) Nomi dialettali <table><tr><td>GENOVA</td><td>Tunna, Tonnella</td></tr><tr><td>VIAREGGIO</td><td>Sanguinaccio</td></tr><tr><td>TERRACINA</td><td>Lettirado</td></tr><tr><td>NAPOLI</td><td>Alletterato</td></tr><tr><td>Reggio Calabria</td><td>Allitteratu, Zirru, tunnina</td></tr><tr><td>BRINDISI</td><td>Nzirru</td></tr><tr><td>BARI</td><td>Palametidd</td></tr><tr><td>MANFREDONIA</td><td>Letterato</td></tr><tr><td>VENEZIA</td><td>Carcana, Aletterato</td></tr><tr><td>RIMINI</td><td>Letterato Tonnina</td></tr><tr><td>TRIESTE</td><td>Letterato</td></tr><tr><td>CAGLIARI</td><td>Alacurza Scampírru</td></tr><tr><td>MESSINA</td><td>Littirati, Allitteratu Tunnina</td></tr><tr><td>CATANIA</td><td>Allitratu, Cuvarito</td></tr><tr><td>SIRACUSA</td><td>Pizziteddu Cavarito</td></tr><tr><td>Porto Empedocle</td><td>Alliterato, Cavarito</td></tr><tr><td>PALERMO</td><td>Allitratu</td></tr></table> (2) Abbreviazioni: LTA (3)	GENOVA	Tunna, Tonnella	VIAREGGIO	Sanguinaccio	TERRACINA	Lettirado	NAPOLI	Alletterato	Reggio Calabria	Allitteratu, Zirru, tunnina	BRINDISI	Nzirru	BARI	Palametidd	MANFREDONIA	Letterato	VENEZIA	Carcana, Aletterato	RIMINI	Letterato Tonnina	TRIESTE	Letterato	CAGLIARI	Alacurza Scampírru	MESSINA	Littirati, Allitteratu Tunnina	CATANIA	Allitratu, Cuvarito	SIRACUSA	Pizziteddu Cavarito	Porto Empedocle	Alliterato, Cavarito	PALERMO	Allitratu
GENOVA	Tunna, Tonnella																																		
VIAREGGIO	Sanguinaccio																																		
TERRACINA	Lettirado																																		
NAPOLI	Alletterato																																		
Reggio Calabria	Allitteratu, Zirru, tunnina																																		
BRINDISI	Nzirru																																		
BARI	Palametidd																																		
MANFREDONIA	Letterato																																		
VENEZIA	Carcana, Aletterato																																		
RIMINI	Letterato Tonnina																																		
TRIESTE	Letterato																																		
CAGLIARI	Alacurza Scampírru																																		
MESSINA	Littirati, Allitteratu Tunnina																																		
CATANIA	Allitratu, Cuvarito																																		
SIRACUSA	Pizziteddu Cavarito																																		
Porto Empedocle	Alliterato, Cavarito																																		
PALERMO	Allitratu																																		

La variazione denominativa può, inoltre, essere espressa sull’asse diafasico/diastratico: le denominazioni dialettali non avranno le stesse condizioni d’uso del termine scientifico, o della denominazione commerciale; così, il *Thunnus alalunga* avrà diverse denominazioni regionali e solo tre denominazioni commerciali autorizzate⁶:

Fig. 10 - Esempio di variazione diafasica nella denominazione



La variazione denominativa può, infine, manifestarsi anche sull’asse diacronico: nel tempo, alcuni nomi di specie vengono abbandonati, e sostituiti, come nell’esempio seguente, in cui *Tetrapturus albidus* è ormai sconsigliato e sostituito da *Kajikia albida*⁷:

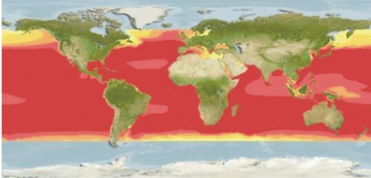
⁶ *Thunnus alalunga* (europa.eu), URL: <https://fish-commercial-names.ec.europa.eu/fish-names/species/thunnus-alalunga_it#commdes> (consultato il 19-11-2025).
⁷ Le tassonomie scientifiche vengono aggiornate; per quanto riguarda le specie marine le variazioni sono consultabili in repertori come il World Register of Marine Species (WoRMS), URL: <<https://www.marine-species.org/>> (consultato il 19-11-2025).

Fig. 11 - Esempio di variazione diacronica nella denominazione

Lemma	<i>Tetrapturus albidus</i>
Sinonimi	<i>Kajikia albida</i> (denominazione latina riclassificata post UNCLOS) (1) Auggia peddirina (Sicilia), Auggia pilligrina, Marlin bianco (2), Marlin (3) Abbreviazioni: WHM (1)

Come si evince dai pochi esempi riportati sopra, la sinonimia è in questo microdominio particolarmente sviluppata, e le strategie di nominazione si affidano spesso a tratti percettivi, quali la dimensione di pinne, occhi o altre parti del corpo (come in *bigeye tuna*), il colore (*tonno a pinne gialle*), la zona di diffusione (*tonnetto indopacifico*). Per questo motivo, nella schedatura sono state sempre inserite due immagini, una per la specie oggetto di schedatura e l'altra per la diffusione della specie a livello globale (si veda sotto, per il *Thunnus alalunga*):

Fig. 12 - Esempio di immagine per documentare la diffusione della specie

Lemma	<i>Thunnus alalunga</i>																
Sinonimi	Aa-lunga, Alalunga, Alalangu, Alalunga, Alilonga, Lalonga, Licia, Tonno, Tonno bianco (1)(2) Nomi dialettali <table><tr><td>GENOVA</td><td>Aa-lunga</td></tr><tr><td>ROMA</td><td>Alalunga</td></tr><tr><td>NAPOLI</td><td>Alalunga</td></tr><tr><td>TARANTO</td><td>Alalunga</td></tr><tr><td>GALLIPOLI</td><td>Alilonga</td></tr><tr><td>VENEZIA</td><td>Alalunga</td></tr><tr><td>CAGLIARI</td><td>Alalangu, Licia</td></tr><tr><td>SICILIA</td><td>Alalunga</td></tr></table> (3) Abbreviazioni: ALB (4)	GENOVA	Aa-lunga	ROMA	Alalunga	NAPOLI	Alalunga	TARANTO	Alalunga	GALLIPOLI	Alilonga	VENEZIA	Alalunga	CAGLIARI	Alalangu, Licia	SICILIA	Alalunga
GENOVA	Aa-lunga																
ROMA	Alalunga																
NAPOLI	Alalunga																
TARANTO	Alalunga																
GALLIPOLI	Alilonga																
VENEZIA	Alalunga																
CAGLIARI	Alalangu, Licia																
SICILIA	Alalunga																
Relazioni concettuali	Scombridae Thunnus Thunnus alalunga Thunnus albacares Thunnus atlanticus Thunnus maccoyii Thunnus obesus Thunnus thynnus (5)																
 (5)  (6)																	

La circolazione delle denominazioni delle specie più diffuse può in alcuni casi portare addirittura a denominazioni scientificamente errate: è il caso della specie *Katsuwonus*

pelamis che, pur non appartenendo al genere *Thunnus*, viene correntemente denominato *skipjack tuna*. Anche in questi casi particolarmente spinosi, l'apporto delle comunità esperte può rivelarsi fondamentale.

4. Per concludere

Il caso di studio brevemente commentato nelle pagine precedenti rappresenta un buon esempio di progetto terminografico che richiede un gruppo di lavoro fortemente interdisciplinare. Le competenze di analisi linguistica, infatti, rischiano di non essere sufficienti per analizzare nel dettaglio la classificazione scientifica e per orientarsi all'interno delle molte variazioni denominative esistenti (alcune delle quali, quali *tuna* per citare un esempio emblematico, spesso utilizzate in modo improprio e fuorviante).

Nel caso del progetto descritto in questo contributo, i membri e tirocinanti del Ce.R.Te.M. hanno potuto avvalersi delle competenze presenti nei diversi settori dell'Università di Genova, valorizzandole allo stesso tempo con la disseminazione dei risultati della propria ricerca. La valenza istituzionale del glossario marittimo costituisce un valore aggiunto importante per questo strumento di analisi e di comunicazione presso il grande pubblico, e la metodologia adottata può essere considerata come un primo esperimento per innovare le procedure attualmente esistenti: se, infatti, l'apporto della linguistica viene spesso considerato come accessorio rispetto a grandi progetti di ricerca disciplinari, il piccolo tentativo compiuto dal Ce.R.Te.M. dimostra che si può mettere l'aspetto linguistico al centro di un progetto interdisciplinare ampio e ambizioso.

Nel 2023, un ulteriore tassello della collaborazione è stato aggiunto grazie alla collaborazione tra il Ce.R.Te.M., il Centro del Mare dell'Università di Genova e il giro del mondo della Nave Vespucci: è, infatti, attualmente in corso la schedatura di un microdominio specifico alla navigazione a vela, con il supporto dei ricercatori imbarcati e del personale della Marina Militare – anche in questo caso, la capacità di lavorare in prospettiva interdisciplinare e interistituzionale rappresenta uno dei punti di forza del progetto, che prevede la schedatura di 100 termini in arabo, cinese, francese, inglese, italiano, portoghese, spagnolo, tedesco entro la fine del viaggio della Nave Vespucci (febbraio 2025)⁸.

Riferimenti bibliografici

Bachimont B., *L'interdisciplinarité comme pratique disciplinaire*, in « Revue intelligibilité du numérique », 1, 2020. URL : <<https://intelligibilite-numerique.numerev.com/index.php/numeros/n-1-2020/4-l-interdisciplinarite-comme-pratique-disciplinaire>>, consultato il 19-11-2025.

⁸ URL: <<https://life.unige.it/progetto-vespucci-sett2023>> (consultato il 19-11-2025). Al momento della pubblicazione di questo testo, il glossario del progetto è consultabile all'indirizzo <<https://certem.it/GlossarioVespucci/index.php?op=home>> (consultato il 19-11-2025).

- Bianchini L., Mabroul A., Rossi M., *Les mots de l'eau : entre terminologie spécialisée et analyse interculturelle*, in « Synergies Italie », 4, 2008, pp. 123-132.
- Cabré M.T., *Terminologie ou terminologies ? Spécialité linguistique ou domaine interdisciplinaire ?*, in « Meta », 36/1, 1991, pp. 55-63.
- Di Nunzio D., Grbac D., Corveddu M., Rolla C., *Terminologia e tassonomia: un'esperienza in Biblioteca*, in « Informatica e Diritto », XL annata, vol. XXIII, n. 1, 2014, pp. 193-223.
- Faber P., L'Homme M.C. (eds.), *Theoretical Perspectives on Terminology: Explaining terms, concepts and specialized knowledge*, Amsterdam/Philadelphia, John Benjamins, 2022.
- Grimaldi C. et al. (a cura di), *Terminologia e interculturalità. Problematiche e prospettive*, Città di Castello, I Libri di Emil, 2022.
- Piccardo G., *CeRTeM – Centro di Ricerca in Terminologia Multilingue dell'Università di Genova*, in « Publif@rum », 12, 2009. URL: <<https://riviste.unige.it/index.php/publiforum/article/view/1530/1658>>, consultato il 19-11-2025.
- Rossi M., *Langue et culture dans un verre. Pour une étude multilingue du langage du vin*, in E. Lavric (ed.), *Food and Language. Sprache und Essen*, Frankfurt, Peter Lang, 2007, pp. 161-170.
- Zanola M.T., *Terminologia, traduzione e comunicazione specialistica, diffusione delle conoscenze: le attività dell'Associazione Italiana per la Terminologia*, in « Publif@rum », 27, 2017. URL: <http://www.farum.it/publiforum/ezine_pdf.php?id=384>, consultato il 19-11-2025.

Imprécisions terminologiques et marketing : le cas de la bière *doppio malto*

LORENZO DEVILLA, NICLA MERCURIO

1. Introduction¹

La terminologie, en tant que discipline, traite de la normalisation des langues de spécialité afin d'assurer une communication efficace et sans ambiguïté, facilitant le dialogue entre les expert-es d'un domaine, ainsi que la transmission de connaissances à un public non professionnel (Cabré, 1998 [1993] ; Zanola, 2011, 2018 : 18, 97, 98). Toutefois, cette définition se heurte à la complexité des sociétés actuelles, consuméristes et numérisées, qui se caractérisent entre autres par une démocratisation de l'information – au sens de la facilité d'accès mais aussi de la possibilité de créer des contenus (Beauvisage *et al.*, 2014 : 163, 164) – et par l'horizontalisation participative, à savoir le processus par lequel les médias numériques redistribuent la responsabilité énonciative étant auparavant l'apanage de sources légitimes et institutionnelles, vers un public désormais actif et producteur de contenus (Bertin, Granier, 2015 : 121). S'opposant au précédent modèle de la communication fondée sur une « idéologie “transmissive” et souveraine » (Bertin, Granier, 2015 : 121), les nouveaux canaux de communication permettent à tout le monde d'accéder à certaines terminologies et aux connaissances qu'elles véhiculent, et de contribuer à leur diffusion, même si l'on n'est pas spécialiste d'un domaine donné.

Ainsi, encore plus qu'auparavant, les termes voyagent d'un secteur à l'autre, des professionnel·les aux non professionnel·les, qui se les approprient et les réutilisent. Devenant parfois des mots du lexique courant (déterminologisation ou banalisation, Meyer, Mackintosh, 2000a, 2000b) ou passant dans une autre langue de spécialité (nomadisation), ces termes peuvent acquérir des significations plus larges ou s'éloigner de leur sens originel (Bertaccini *et al.*, 2008 : 51). Dans certains cas, il s'agit de mécanismes typiques des procédés de néologie, mais on court également le risque de tomber dans la polysémie et l'imprécision ou de faire des usages impropres.

Il n'est donc pas étonnant que la déterminologisation et d'autres phénomènes similaires suscitent l'intérêt des chercheurs et chercheuses, qui soulignent, d'un côté, un besoin accru de connaissances spécialisées de la part des « gens ordinaires » (Scarpa, 2002 : 35 ; Bertaccini *et al.*, 2008 : 50) et, de l'autre, la présence de ces phénomènes dans le

¹ Lorenzo Devilla a rédigé les chapitres 1, 4.2 et 5, alors que Nicla Mercurio a rédigé les chapitres 2, 3, 4. et 4.1. Cette contribution s'inscrit dans le cadre du projet de recherche *Plurilinguismo, patrimonio culturale e sviluppo sostenibile* (« Plurilinguisme, patrimoine culturel et développement durable »), dirigé par Lorenzo Devilla (Université de Sassari) et financé par la *Fondazione di Sardegna-Progetti di ricerca di base dipartimentali 2022 e 2023*.

contexte technologique et numérique (Ungureanu, 2006) ainsi que dans le domaine médiatique et publicitaire (Corbolante, 2010, 2013). En effet, comme exemples de dé-terminologisation en italien, Corbolante cite entre autres *Internet*, *algoritmo*, *hashtag*, *meme* et *drone*.

Le terme italien que nous étudions – [*birra*] *doppio malto*, littéralement « [bière] double malt » –, bien que ne relevant pas du domaine informatique comme ceux que nous venons de mentionner, relève pourtant de ce type de phénomène et peut aussi représenter un cas de « fausse terminologie » (*false terminology*)², à l’instar de ce que l’on observe de plus en plus dans le marketing des aliments et des boissons. En Italie, *birra doppio malto* est une dénomination de vente (loi n° 1354/1962, art. 2) (Normattiva, 2016) qui a une incidence sur les impôts à payer par la brasserie productrice (décr.-L. n° 504/1995, art. 34) (Normattiva, 2023), mais dans le langage courant ce terme tend à désigner un style de bière possédant certaines caractéristiques. Les idées reçues sur l’univers brassicole, les publicités et la présence du terme sur les étiquettes de différentes bières italiennes ou étrangères importées en Italie (PE, 2011 ; Di Siero, 2023 : 3), qu’elles soient industrielles ou artisanales, contribuent à la diffusion de cette acception.

Dans cet article, nous allons examiner un corpus hétérogène composé de sites Internet de grandes surfaces et de sites de vente spécialisés, ainsi que d’étiquettes de bières en langue italienne, afin de comprendre dans quel sens le terme est le plus fréquemment employé et s’il s’accompagne de stratégies promotionnelles. Pour ce faire, en nous appuyant sur l’analyse du discours publicitaire (Adam, Bonhomme, 2012 ; Berthelot-Guiet, 2015) et sur l’onomastique commerciale (Altmanova, 2013 ; Temmerman, 2018), nous mènerons une analyse qualitative selon une perspective terminologique et discursive. Nous essayerons de répondre aux questions suivantes : compte tenu de l’écart existant entre le marché industriel et la production artisanale, quelles sont les brasseries qui utilisent le plus le terme *doppio malto* dans son acception étendue ? Pourquoi ? Nous tenterons également de réfléchir aux conséquences que cela peut avoir ainsi qu’au rôle joué par la terminologie, qui, dans ce contexte, dialogue avec le discours publicitaire (numérique) et les secteurs spécialisés évoqués plus haut, de la filière agroalimentaire à la restauration.

2. Le domaine brassicole

Le domaine brassicole est en plein essor à l’échelle globale (Aquilani *et al.*, 2015 : 214), surtout en ce qui concerne une catégorie spécifique : la bière artisanale, à savoir une bière non filtrée, non pasteurisée et produite en quantité limitée, même s’il faut préciser que les terminologies et les dispositions législatives en la matière varient d’un pays à l’autre (Mercurio, 2023 : 40).

² Alors que le contenu sémantique d’un mot peut être adapté en fonction du contexte, celui d’un terme est lié à un domaine spécialisé spécifique. Par conséquent, les termes qui s’écartent du domaine originel peuvent provoquer des malentendus et diffuser des connaissances erronées, tout en restant des termes à part entière : *ecopelle* (Zanola, 2017) est un exemple célèbre à cet égard.

En effet, ces dernières années, on assiste à un développement significatif de la filière tant au niveau de la consommation que de la production. Cette situation est d'autant plus intéressante qu'elle concerne des territoires plutôt associés à la viticulture dans l'imaginaire collectif, mais qui sont en revanche caractérisés par une croissance considérable du secteur de la bière. En effet, d'après Brewers of Europe, en Italie la production brassicole a augmenté de 22 % dans la période 2015-2023³, alors qu'elle a diminué en Allemagne (-11,2 %) et dans le Royaume-Uni (-17,1 %), qui ont été les plus grands producteurs en 2020 et 2021 (BoE, 2024 : 8, 9 ; BoE, 2022 : 6). D'ailleurs, l'Italie est l'un des pays européens qui compte le plus grand nombre de microbrasseries (1.010), après les géants du secteur (BoE, 2024 : 22). En 2023, elle est classée parmi les principaux pays consommateurs (BoE, 2024 : 10).

Une production scientifique importante témoigne du regain d'attention à l'égard de la bière. Outre les études en sciences économiques (Garavaglia, Swinnen, 2018 ; Miller *et al.*, 2019) et en nutrition (Broz *et al.*, 2022 ; Zugravu *et al.*, 2023), de nombreux travaux en sciences du langage traitent du discours de dégustation de bière, un genre discursif que Konnelly (2020) dénomme *brutoglossia* (*craft beer talk*) sur le modèle de *oinoglossia* (Silverstein, 2003), terme désignant le discours sur le vin. Sans prétendre à l'exhaustivité, nous citons aussi les travaux en terminologie – principalement en langue anglaise (Konnelly, 2020 ; Malin, 2019), mais aussi en langue française (Mercurio, 2023 ; Parizot *et al.*, 2023) –, le marketing, le discours publicitaire (Parizot, 2022 ; Devilla, Mercurio, 2025), l'onomastique commerciale (Temmerman, 2018 ; Mercurio, 2024), l'ethnographie du goût (Fletcher, 2016 ; Hamer, 2017) et la perception des consommateurs et consommatrices (Lelièvre *et al.*, 2008 ; Aquilani *et al.*, 2015).

2.1 La bière artisanale en Italie

Comme nous l'avons souligné lors d'une étude précédente (Mercurio, 2023 : 146), s'il est vrai que la France et l'Italie possèdent toutes deux une tradition viticole très importante et réputée dans le monde entier, la France est davantage orientée vers les grands producteurs de bière comme l'Allemagne, la Belgique, le Royaume-Uni et l'Irlande. En revanche, en Italie, à cause de préjugés hérités de l'époque romaine (UB, 2020 : 9), on a encore tendance à reléguer la bière au statut de boisson simple, voire grossière, à boire froide avec une pizza, un plat considéré autrefois comme pauvre. De plus, la diffusion du phénomène *crafty*, qui désigne l'imitation des bières artisanales par l'industrie à travers l'emploi d'un certain emballage⁴ ou de mots-clés, critiqué au départ aux États-Unis (BA, 2012 ; Rice, 2016 : 245, 246), a entraîné la nécessité de mettre en place des initiatives visant à protéger le produit et la filière artisanale.

En effet, à la suite de nos recherches et des informations fournies par l'association italienne de catégorie Unionbirrai, nous pouvons affirmer que l'Italie est le seul pays

³ Un scénario similaire peut être constaté en France, où la production a progressé de 8,4 %.

⁴ Altmanova (2013 : 53) souligne l'importance du *packaging* en tant qu'outil participant à la création de l'identité d'une marque.

qui a élaboré dans sa législation une définition de *birra artigianale*. Aux États-Unis, en Australie, au Royaume-Uni ou en Allemagne, des définitions de *birra artigianale* ont été formulées par des associations (Baiano, 2020 : 1830-1831). La dénomination *birra artigianale* figure plus précisément dans l'article 2 comma 4-bis de la *Disciplina igienica della produzione e del commercio della birra*, la loi n°1354 du 16 août 1962, mis à jour en 2016, en tant que produit de petites brasseries indépendantes, dont on donne aussi une définition (Normattiva, 2016) :

Si definisce birra artigianale la birra prodotta da piccoli birrifici indipendenti e non sottoposta, durante la fase di produzione, a processi di pastorizzazione e di microfiltrazione. Ai fini del presente comma si intende per piccolo birrificio indipendente un birrificio che sia legalmente ed economicamente indipendente da qualsiasi altro birrificio, che utilizzi impianti fisicamente distinti da quelli di qualsiasi altro birrificio, che non operi sotto licenza di utilizzo dei diritti di proprietà immateriale altrui e la cui produzione annua non superi 200.000 ettolitri, includendo in questo quantitativo le quantità di birra prodotte per conto di terzi.

En France, en revanche, on se limite à mentionner la dénomination *bière*, qui figure dans le décret n° 92-307 du 31 mars 1992 (Légifrance, 2020), dans la version en vigueur au 2 février 2020. Pour définir avec plus de précision ce que l'on entend par *bière artisanale*, il faut aussi avoir recours à la définition d'*artisanat* donnée par l'Institut national de la statistique et des études économiques (INSEE, 2019).

2.1.1 La birra doppio malto en Italie : une étude de cas

Si l'on prend en considération le terme [*birra*] *doppio malto*, on se rend compte que des idées fausses concernant cette dénomination circulent chez les consommateurs et consommatrices italien-nes. Au-delà du monde du whisky, que cette expression peut facilement évoquer, l'adjectif *double* évoque quant à lui les célèbres bières belges *Dubbel* ou *Double*⁵. Nous avons ainsi constaté que certaines cartes de restaurants proposent des bières belges *doppio malto*, comme s'il s'agissait d'un style brassicole précis⁶. De plus, une marque très connue dont les brasseries sont diffusées partout en Italie s'appelle *Doppio Malto*. Il s'agit d'un nom de marque qui est bien capable de « séduire le consommateur » et de s'inscrire dans l'imaginaire collectif (Altmanova, 2013 : 25, 28). Au contraire, une brasserie italienne précise sur un panneau que la *birra doppio malto* n'existe pas (sous-entendu en tant que style de bière) et que le terme est issu de la sphère juridique (Image 1)⁷.

⁵ Bières caractérisées par une couleur ambrée profonde et cuivrée, un arôme de malt et un corps moyen. La teneur en alcool est comprise entre 6 et 7,6 % (BJCP, 2021 : 53-54).

⁶ Les styles universellement reconnus ont été codifiés par le *Beer Judge Certification Program* (BJCP, 2021), une organisation américaine qui s'occupe de systématiser les compétences en matière de dégustation et d'évaluation de la bière.

⁷ « Miti da sfatare sulla birra : la birra doppio malto non esiste ! È un'invenzione del legislatore italiano ».

Un message semblable s'est également répandu sur la toile à travers des mêmes partages par les passionné-es de bière artisanale.

Image 1 - Panneau affiché dans une brasserie de Naples (Italie) (photo prise par nous, 2019)



Les dictionnaires n'aident pas non plus à éclaircir la question, laissant supposer que la *birra doppio malto* est brassée avec deux types de malts (orge et blé), d'où l'adjectif *doppio*. Dans le dictionnaire en ligne *Treccani*, le terme apparaît sous l'entrée *malto* : « birra doppio m., preparata con un miscuglio di malto d'orzo e di frumento ». Le dictionnaire du *Corriere della sera* donne un exemple similaire : « birra doppio m., ottenuta da malto d'orzo e frumento ». Dans le *Nuovo De Mauro* sur *Internazionale*, figure l'entrée *doppio malto* : « di bevanda alcolica, spec. di birra o whisky, preparata con una mescolanza di malto d'orzo e di frumento ».

En revanche, l'encyclopédie *Treccani* inclut le terme dans le paragraphe *Disposizioni legislative*. De fait, l'article 2 de la loi de la *Disciplina igienica della produzione e del commercio della birra* (Normattiva, 2016) déjà citée mentionne la classification « tutta italiana » (UB, 2020 : 132), qui distingue entre *birra analcolica*, *birra leggera* ou *light*, *birra speciale* et *birra doppio malto* sur la base du degré Plato ou degré saccharimétrique, à savoir l'unité de mesure qui exprime la quantité de sucres présents dans le moût avant la fermentation (Normattiva, 2016). La dénomination *birra doppio malto* est réservée au produit dont le degré Plato est supérieur ou égal à 14,5.

Il s'agit d'une dénomination de vente qui ne révèle aucune caractéristique de la recette ou du style de la bière, mais qui détermine le droit d'accise à payer par la brasserie, selon l'article 34 du décret-loi n° 504/1995 (Normattiva, 2023). Le terme peut suggérer quelque chose à propos de la teneur en alcool, puisqu'une *birra doppio malto* présente un titre alcoométrique volumique supérieur à 3,5 %, à l'instar d'ailleurs des boissons qui peuvent être classées dans les catégories *birra* et *birra speciale* (UB, 2020 : 132). Ce ne sont donc pas nécessairement des bières très fortes en alcool comme on pourrait le penser, ni des bières brassées avec deux types de malts.

3. Collecte et description du corpus

Le corpus d'étude, collecté entre mai et août 2023, est constitué de 6 sites de la toile de grandes surfaces⁸ et 9 sites de vente spécialisés – dont la boutique en ligne d'une entreprise agricole et brassicole –, considérés comme des références par les principaux réseaux du secteur en Italie (Fermento Birra)⁹. Les premiers sont destinés à un public large et varié, alors que les autres s'adressent à une clientèle plus experte ou à des personnes intéressées par des produits brassicoles spécifiques. Le corpus a été sélectionné aussi par une recherche avancée sur Google, en saisissant comme mot-clé *doppio malto*. À partir de ces sites, nous avons récupéré les catégorisations dans les interfaces, les descriptions, les présentations des bières et leurs étiquettes.

Nous avons considéré les bières italiennes et étrangères, commercialisées sur le marché italien. Au total, nous avons retenu 104 bières et 70 marques (nous ne comptabilisons pas les mêmes bières si elles sont présentes dans plus d'un site ou avec des formats différents : bouteilles, canettes, fûts, packs), dont 26 marques italiennes et 44 étrangères. Les autres pays où ces bières sont brassées sont la Belgique, l'Allemagne, le Danemark, la France, l'Irlande, les Pays-Bas, le Royaume-Uni, l'Écosse, les États-Unis, l'Autriche, la Suisse et l'Espagne. Il y a 20 marques industrielles et 50 marques artisanales, mais il faut préciser que ce comptage ne peut pas être précis car à l'étranger la notion de bière artisanale est moins stricte qu'en Italie.

De plus, même s'ils ne font pas partie de notre corpus, les sites officiels et les pages sur les réseaux sociaux (Facebook et X) de certaines brasseries ont été consultés, y compris en d'autres langues, pour effectuer des comparaisons. D'ailleurs, ces sites sont parfois la source directe des informations que les supermarchés et les distributeurs insèrent dans les fiches des produits.

4. Analyse du corpus

Nous nous concentrerons d'abord sur les différents points dans lesquels le terme [*birra*] *doppio malto* figure. Nous proposerons ensuite des exemples significatifs, des bières industrielles aux bières artisanales, étrangères et italiennes.

Comme nous l'avons anticipé, le terme apparaît d'abord dans la description du produit (71 fois) : par exemple, « una birra doppio malto rossa a bassa fermentazione » ou « è una speciale bionda doppio malto ». Les citations sur les étiquettes sont très nombreuses (41 – dans la moitié des cas, le terme est répété ou accentué par des astuces graphiques) et dans le titre du produit qui accompagne le nom de la bière (18) : « LEFFE BLONDE Birra bionda belga d'abbazia doppio malto », « Tuborg doppio malto Strong », *Menabrea Rossa Doppio Malto*, etc.

⁸ Carrefour, Coop, Emi Supermercati, Lidl, Pam, Sole 365.

⁹ Beer Paradise Shop, Beer Wulf, Bernabei, Biraghi a casa, Birraland, Birre da manicomio, Mastri Birrai Umbri, Rivoldrink, Trova birre. Dans 11 autres sites spécialisés consultés, le terme [*birra*] *doppio malto* ne figure pas.

[Birra] *doppio malto* figure également dans la fiche technique du produit, notamment en tant que *denominazione [legale] di vendita* (13) ou dans les caractéristiques de la bière concernée (2). Moins fréquentes mais assez significatives sont les citations dans certains textes qui introduisent les produits (3), où nous lisons, par exemple, que « la birra *doppio malto* è una delle tipologie più amate dagli appassionati di birra artigianale », et qui associent des « birre importanti » à des « *doppio malto* ». En tant que style ou catégorie, le terme figure aussi dans des listes déroulantes permettant de rechercher le produit (2).

4.1 *Doppio malto* entre industrie et artisanat

Certaines bières étrangères, distribuées et promues par leur division italienne, ont des étiquettes traduites où le terme [birra] *doppio malto* est en évidence, sur la contre-étiquette de la canette ou de la bouteille. Comme il s'agit d'un concept éminemment italien, cette expression ne figure pas sur le même produit en vente à l'étranger (comme dans le cas de la bière *Tennent's*) et il n'y a pas non plus d'équivalents dans les autres langues proposées. Nous constatons une différence similaire, voire davantage marquée, pour la plupart des bières industrielles étrangères du corpus : [birra] *doppio malto* apparaît sur l'étiquette italienne de la bière *Du Demon*, mais pas dans la fiche technique du site officiel en français.

La situation change lorsque le site en ligne d'une brasserie est disponible en italien : par exemple, [birra] *doppio malto* figure sur la contre-étiquette de la bière hollandaise 8.6, dans la fiche technique en italien et sur le profil X. À l'inverse, le terme est absent dans les versions anglaise et française.

Sur la bouteille de la *Ceres*, le terme est renforcé par un jeu de mots – « ci vedi doppio (malto) » – suggérant le fort degré d'alcool qui devrait relever d'une (prétendue) double quantité de malt. En revanche, *doppio malto* figure de manière plus discrète sur les étiquettes qui ne sont pas entièrement traduites, comme celle des bières *Leffe* et *La Trappe*, où il représente une simple indication de la dénomination de vente.

Quant aux bières industrielles italiennes, nous constatons que la promotion repose beaucoup sur le recours au terme *doppio malto*, bien présent sur les étiquettes et parfois répété : dans le cas de *Peroni*, il est surligné en caractères dorés et l'information est encore plus visible que l'indication du style.

Un autre exemple significatif est donné par la bière *Perlenbacher*, brassée en Italie pour certains supermarchés et vendue dans des pays différents. Sur le site italien, nous avons le produit « Birra Strong *doppio malto* », alors que, dans la version en langue italienne du site suisse, cette même bière est tout simplement indiquée comme « Strong 7,9 % Vol. ».

Dans le cas de nombreuses bières artisanales étrangères, on constate que le terme [birra] *doppio malto*, proposé comme catégorisation sur les sites de vente et dans les descriptions, n'apparaît pas sur les étiquettes ni sur les sites officiels, les brasseries *Riegele* et *Side Effect* étant de bons exemples à cet égard.

En ce qui concerne les bières artisanales italiennes, le cadre est plus varié. Certaines sont indiquées comme des *doppio malto* sur le site de vente ainsi que sur le site officiel

et sur l'étiquette (par exemple, *Farchioni, Gilac*), alors que d'autres sont décrites comme des *doppio malto* par le vendeur (*birra bionda doppio malto, Doge*), mais pas par le producteur sur le site officiel (« Strong lager », *Doge*). Certaines microbrasseries, telles que *Mastio* et *Tre Fontane* l'indiquent de manière très simple sur l'étiquette.

4.2 *Doppio malto* en discours

Lorsque le terme n'est pas seulement employé en tant que dénomination de vente, il peut constituer une véritable stratégie publicitaire visant à convaincre les acheteurs et acheteuses potentiel·les et à alimenter un certain imaginaire associé aux bières *doppio malto*.

En effet, on a remarqué que dans les descriptions des produits, élaborées par les vendeurs ou par les producteurs, le terme est souvent accompagné d'autres mots-clés renvoyant à des caractéristiques spécifiques, tels que *corposo-a, intenso-a* (intensité de goût, d'arôme, de teneur en alcool ou de couleur), *ricchezza, ricco-a di...* (richesse de goût, d'arôme ou de corps), *pieno-a*, (s'inscrivant dans les catégories déjà mentionnées), *forte, alto-a, elevato-a* faisant référence au degré d'alcool, bien que, comme nous l'avons dit au chapitre 2.1.1., ce ne soit pas toujours le cas pour une bière *doppio malto*.

Parfois *doppio malto* est précédé d'un article défini ou indéfini et suivi d'une désignation territoriale : *la/una doppio malto + danese, italiana, belga, tedesca*. Ou encore, dans le sillage du marché *crafty*, certaines brasseries industrielles exploitent la terminologie associée aux caractéristiques des bières artisanales, qui ne subissent pas de filtration lors de la postproduction et, surtout, ne sont pas pasteurisées. Nous observons ainsi une récurrence de termes comme *non filtrata, non pastorizzata* et *cruda*.

Qu'il s'agisse de produits artisanaux ou industriels, on remarque la tentative de mettre en avant la qualité de la bière également à travers la référence à la tradition, à l'utilisation d'ingrédients locaux et au lien avec le territoire : *migliori malti selezionati, di puro malto, malto da orzo 100 % italiano, fascino tutto italiano, malto d'orzo selezionato, ingredienti naturali, metodo tradizionale [belga]¹⁰, in Italia con il metodo artigianale, in perfetta tradizione tedesca, fedele alla tradizione e al territorio*.

Dans certains cas, le terme *malto* vise à indiquer l'emploi d'une quantité supérieure de cet ingrédient : *ben due diverse tipologie di malti, 6 varietà di malti, prodotta con doppio malto d'orzo/di frumento¹¹*.

Nous avons aussi repéré des définitions qui, contrairement aux autres exemples cités, sont totalement imprécises. *Doppio malto* y figure comme synonyme de plusieurs styles brassicoles :

- *doppio malto, cioè che sviluppa una gradazione importante* ;
- *Birra speciale Belga, in stile Strong Ale, o comunemente denominata come birra bionda doppio malto* (*birra speciale* et *birra doppio malto* sont deux dénominations de vente différentes mentionnées dans la loi) ;
- *doppio malto, denominata anche Doppelbock* ;

¹⁰ On observe le terme *metodo* qui revoie à l'univers du champagne.

¹¹ D'ailleurs, dans la fiche originale d'une bière industrielle espagnole, on lit que le produit se caractérise par « el doble de malta [...] la doble cantidad de malta en la receta ».

- *è una Strong Dark Ale, cioè una birra scura doppio malto* ;
- *Bock italiana, ovvero una “Doppio Malto”*.

Il ne faut donc pas s'étonner de lire des commentaires insistant sur l'idée que le terme *doppio malto* désigne un style spécifique ou une bière forte en alcool :

È una vita che il sabato sera mi divido equamente tra le bionde e le rosse, non faccio distinzioni, l'unica prerogativa è che devono essere doppio malto. #Birra #DoppioMalto (X)

5. Remarques conclusives

Cet article visait à mettre en évidence l'influence que la publicité, notamment à l'ère du numérique, peut avoir sur les terminologies et leur précision. L'analyse d'un corpus hétérogène a montré que le terme italien examiné, [*birra*] *doppio malto*, une dénomination de vente mentionnée dans la loi italienne, est assez ambigu, l'adjectif *doppio* suggérant en effet qu'il y aurait quelque chose de plus (plus de malt, plus d'alcool, etc.) et évoquant aussi les bières belges *Dubbel*. Comme nous avons essayé de le mettre en évidence, l'imaginaire collectif associé à la bière et le marketing, qui mise largement sur la mention du terme sur l'étiquette, alimentent cette fausse terminologie.

Comme il ressort de notre analyse, [*birra*] *doppio malto* est utilisé dans plusieurs sections du corpus et sous différentes formes, du titre du produit à la description, de la fiche technique à l'étiquette. Dans la plupart des cas, il est utilisé de manière correcte, même si on précise rarement qu'il s'agit d'une dénomination de vente. Au sens que lui attribue la loi italienne se superposent des sens étendus, diffusés par la publicité, les « expert-es » du secteur et la toile. Les explications ou précisions accompagnant ce terme sont le plus souvent imprécises ou fausses.

Dans son acception étendue, *doppio malto* est utilisé la plupart du temps dans un contexte discursif plus large et accompagné d'astuces graphiques ou de mots-clés renvoyant à un ensemble de valeurs partagées associées à la bière. Même si on ne peut pas déterminer une constante, on observe néanmoins que ce sont surtout les brasseries industrielles (italiennes) et les grandes surfaces qui mobilisent cette expression. Toutefois, certaines grandes brasseries artisanales ne sont pas en reste, ainsi que quelques sites spécialisés, qui diffusent eux aussi des informations imprécises.

Compte tenu de ces résultats, nous estimons qu'une interaction entre les professionnels des secteurs et des disciplines concernés, y compris les terminologues, serait souhaitable afin non seulement de véhiculer des connaissances et des terminologies précises, mais aussi d'assurer la transparence pour les consommateurs et consommatrices, connaisseur-euse-s ou non. Dans tous les cas, à consommer avec modération !

Références bibliographiques

Adam J.-M., Bonhomme M., *L'argumentation publicitaire*, Paris, Armand Colin, 2012.

- Altmanova J., *Du nom déposé au nom commun. Néologie et lexicologie en discours*, Milano, EDU-Catt, 2013.
- Aquilani B., Laureti T., Poponi S., Secondi L., *Beer choice and consumption determinants when craft beers are tasted: An exploratory study of consumer preferences*, in « Food Quality and Preference », 41, 2015, pp. 214-224.
- Baiano A., *Craft Beer: An Overview*, in « Comprehensive Reviews in Food Science and Food Safety », 20, 2021, pp. 1829-1856.
- Beauvisage T., Beuscart J., Mellet K., Trespeuch M., *Une démocratisation du marché : notes et avis de consommateurs sur le Web dans le secteur de la restauration*, in « Réseaux », 183(1), 2014, pp. 163-204.
- Beer Judge Certification Program (BJCP), *2021 Beer Style Guidelines*, 2021. URL : <https://www.bjcp.org/download/2021_Guidelines_Beer.pdf>, consulté le 12-11-2025.
- Bertaccini F., Lecci C., Bono V., *Processi di terminologizzazione e determinologizzazione nel dominio della diffusione e distribuzione del libro*, in « AIDAinformazioni », 1-2, 2008, pp. 47-61.
- Berthelot-Guiet K., *Analysier les discours publicitaires*, Paris, Armand Colin, 2015.
- Bertin E., Granier J.-M., *La société de l'évaluation : nouveaux enjeux de l'âge numérique*, in « Communication & langages », 184(2), 2015, pp. 121-146.
- Brewers Association (BA), *Craft vs. Crafty: A Statement from the Brewers Association*, 2012. URL : <<https://www.brewersassociation.org/press-releases/craft-vs-crafty-a-statement-from-the-brewers-association/>>, consulté le 12-11-2025.
- Brewers of Europe (BoE), *European Beer Trends. Statistics Report – 2022 Edition*, 2022. URL : <<https://brewersofeurope.eu/uploads/mycms-files/documents/publications/2022/european-beer-trends-2022.pdf>>, consulté le 12-11-2025.
- Brewers of Europe (BoE), *European Beer Trends. Statistics Report – 2024 Edition*, 2024. URL : <<https://brewersofeurope.eu/wp-content/uploads/2024/12/eu-beer-trends-2024-web.pdf>>, consulté le 12-11-2025.
- Broz P., Rajdl D., Racek J., Trefil L., Stehlík P., *Effect of Beer Consumption on Methylation and Redox Metabolism*, in « Physiol Res », 271(4), 2022, pp. 573-582.
- Cabré M.T., *Terminologie : théorie, méthode et applications*, Paris/Ottawa, Armand Colin/Presses de l'Université d'Ottawa, 1998 [1993].
- Corbolante L., *Internet e determinologizzazione*, in « Terminologia etc. », 2010. URL : <<https://www.terminologiaetc.it/2010/05/14/internet-e-determinologizzazione/>>, consulté le 12-11-2025.
- Corbolante L., *Risemantizzazione*, in « Terminologia etc. », 2013. URL : <<https://www.terminologiaetc.it/2013/09/03/meccanismi-formazione-neologismi-semantic/>>, consulté le 12-11-2025.
- Corriere della sera, *Malto*. URL : <https://dizionari.corriere.it/dizionario_italiano/M/malto.shtm>, consulté le 12-11-2025.
- Devilla L., Mercurio N., *Valoriser le territoire à travers la bière artisanale : une analyse exploratoire de brasseries italiennes (Campanie, Sardaigne), françaises (Corse) et suisses (Jura)*, in « Repères-DoRiF », 31, 2025. URL : <<https://www.dorif.it/reperes/lorenzo-devilla-nicla-mercurio-valoriser-le-territoire-a-travers-la-biere-artisanale-une-analyse-discursive-exploratoire-des-sites-web-de-brasseries-italiennes-campanie-sardaigne-francaises-co/>>, consulté le 12-11-2025.
- Di Siero V., *Le informazioni in etichetta. Lo stato dell'arte con uno sguardo al futuro*, in « UB news », 2, 2023, pp. 4-5. URL : <https://www.unionbirrai.it/admin/public/area_download/7d912be2ba077248f3b91f9f1bf05112/UB_NEWS_GIUGNO_2023.pdf>, consulté le 12-11-2025.

- Fermento Birra, *Comprare birra artigianale online: i migliori beer-shop italiani e stranieri*. URL : <<https://www.fermentobirra.com/birra-sconti-e-offerte/>>, consulté le 12-11-2025.
- Fletchall A.M., *Place-making through Beer-Drinking: A Case Studies of Montana's Craft Breweries*, in « *Geographical Review* », 4(106), 2016, pp. 539-566.
- Garavaglia C., Swinnen J. (eds.), *Economic Perspectives on Craft Beer. A Revolution in the Global Beer Industry*, Cham, Palgrave Macmillan, 2018.
- Hamer A., *Speaking of Qualia: Examining a Craft Beer Microcommunity's Membership Identity through Speech*, MA Thesis, University of South Carolina, 2017.
- Il Nuovo De Mauro, *Doppio malto*. URL : <<https://dizionario.internazionale.it/parola/doppio-malto>>, consulté le 12-11-2025.
- Institut national de la statistique et des études économiques (INSEE), *Artisanat*, 2019. URL : <<https://www.insee.fr/fr/metadonnees/definition/c1137>>, consulté le 12-11-2025.
- Konnolly L., *Brutoglossia: Democracy, Authenticity, and the Enregisterment of Connoisseurship in « Craft Beer Talk »*, in « *Language & Communication* », 75, 2020, pp. 69-82.
- Légifrance, *Décret n° 92-307 du 31 mars 1992 portant application de l'article L. 412-1 du code de la consommation en ce qui concerne les bières*, 2020. URL : <<https://www.legifrance.gouv.fr/loda/id/LEGITEXT000033406888/2020-02-02>>, consulté le 12-11-2025.
- Lelièvre M., Choller S., Abdi H., Valentin D., *What is the Validity of the Sorting Task for Describing Beers? A Study Using Trained and Untrained Assessors*, in « *Food Quality and Preference* », 19, 2008, pp. 697-703.
- Malin N., *Terminology of Beer Reviews*, BA Thesis, University of Gävle, 2019.
- Mercurio N., *Terminologie(s) et territorialité : la description des bières artisanales en français et en italien*, in Ch. Roche (éd.), *TOTh 2022. Terminologie & Ontologie : Théories et Applications*, Chambéry, Presses universitaires Savoie Mont Blanc, 2023, pp. 139-159.
- Mercurio N., *Promouvoir le territoire à travers les appellations commerciales : une analyse comparative des microbrasseries de Provence-Alpes-Côte d'Azur et de Campanie*, in D. Fadda, C. Saggiomo (éds.), *Un coup de dés 12*, Paris, La Renaissance française, 2024, pp. 133-146.
- Meyer I., Mackintosh K., « *L'étiement* » du sens terminologique : aperçu du phénomène de la détermination, in H. Béjoint, P. Thoiron (éds.), *Le sens en terminologie*, Lyon, Presses Universitaires de Lyon, 2000a, pp. 198-217.
- Meyer I., Mackintosh K., *When Terms Move into Our Everyday Lives: An Overview of De-terminologization*, in « *Terminology* », 6(1), 2000b, pp. 111-138.
- Miller S.R., Sirrine J.R., McFarland A., Howard P.H., Malone T., *Craft Beer as a Means of Economic Development: An Economic Impact Analysis of the Michigan Value Chain*, in « *Beverages* », 5(2), 35, 2019. URL : <<https://www.mdpi.com/2306-5710/5/2/35>>, consulté le 12-11-2025.
- Normattiva, *Legge 16 agosto 1962, n° 1354. Disciplina igienica della produzione e del commercio della birra*, 2016. URL : <<https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1962-08-16;1354-art2>>, consulté le 12-11-2025.
- Normattiva, *Decreto legislativo 26 ottobre 1995, n° 504. Testo unico delle disposizioni legislative concernenti le imposte sulla produzione e sui consumi e relative sanzioni penali e amministrative*, 2023. URL : <<https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:1995-10-26;504!vig>>, consulté le 12-11-2025.
- Parizot A., *Stratégies communicationnelles et marketing. Mise en scène discursive et storytelling des produits monastiques. Étude de cas*, in « *Revue de communication sociale et publique* », 34, 2022, pp. 53-69.

- Parizot A., Verdier B., Leroyer P., *À travers le(s) verre(s) à bière : Ceci n'est pas (qu')un verre !*, in A. Moutat, C. Duteil (éds.), *Les terminologies professionnelles de la gastronomie et de l'œnologie : arts de la table, dimensions créatives et culturelles*, Limoges, Presses Universitaires de Limoges, 2023, pp. 73-83.
- Parlamento europeo (PE), *Regolamento (UE) n. 1169/2011 del Parlamento europeo e del Consiglio del 25 ottobre 2011 relativo alla fornitura di informazioni sugli alimenti ai consumatori*, 2011. URL : <<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0018:0063:it:PDF>>, consulté le 12-11-2025.
- Rice J., *Professional Purity: Revolutionary Writing in the Craft Beer Industry*, in « Journal of Business and Technical Communication », 30(2), 2016, pp. 236-261.
- Scarpa F., *Terminologie e lingue speciali*, in M. Magris et al. (a cura di), *Manuale di terminologia. Aspetti teorici, metodologici e applicativi*, Milano, Hoepli, 2022, pp. 27-47.
- Silverstein M., *Indexical Order and the Dialectics of Sociolinguistic Life*, in « Language & Communication », 23(3-4), 2003, pp. 193-229.
- Temmerman R., *Co-création et web 2.0. Noms de marques pour de nouvelles bières artisanales dans la ville multilingue de Bruxelles*, in « Éla. Études de linguistique appliquée », 57(192), 2018, pp. 417-434.
- Treccani. Dizionario online, *Malto*. URL : <<https://www.treccani.it/vocabolario/malto1/>>, consulté le 12-11-2025.
- Treccani. Enciclopedia online, *Birra*. URL : <<https://www.treccani.it/enciclopedia/birra>>, consulté le 12-11-2025.
- Ungureanu L., *L'interpénétration langue générale-langue spécialisée dans le discours d'Internet*, Paris, Connaissances et Savoirs, 2006.
- Unionbirrai (UB), *Corso di degustazione birra – Primo livello. Conoscere e degustare le birre*, Milano, Associazione Unionbirrai, 2020.
- Zanola M.T., *La terminologia, un ponte fra i saperi*, in *Treccani*, 2011. URL : <https://www.treccani.it/magazine/lingua_italiana/speciali/termini/Zanola.html>, consulté le 12-11-2025.
- Zanola M.T., *Terminologia, traduzione e comunicazione specialistica, diffusione delle conoscenze: le attività dell'Associazione Italiana per la Terminologia*, in « Publif@rum », 27, 2017, URL : <<https://riviste.unige.it/index.php/publiforum/article/view/1728/2058>>, consulté le 12-11-2025.
- Zanola M.T., *Che cos'è la terminologia*, Roma, Carocci, 2018.
- Zugravu C.A., Medar C., Manolescu L.S.C., Constantin C., *Beer and Microbiota: Pathways for a Positive and Healthy Interaction Nutrients*, in « Nutrients », 15(4), 844, 2023, doi : 10.3390/nu15040844.

Terminologie et traduction au prisme de l'interdisciplinarité : des débuts de la Terminologie textuelle à la Traduction automatique neuronale

CÉCILE FRÉROT*

1. Introduction

Les besoins de normalisation qui ont présidé au développement de la terminologie se sont accompagnés, dans les années 1980, d'un rapprochement étroit de la terminologie avec la traduction. Celle-ci constituait alors un terrain d'application privilégié, la traduction spécialisée et la terminologie partageant l'objectif commun de répondre aux besoins de la communication spécialisée (Cabré, 1998), en réponse aux besoins des traducteurs (Gouadec, 2005). Le début des années 1990 marque un tournant dans l'histoire de la terminologie en lien avec de nouvelles orientations (Savatovsky, Candel, 2007 ; Faber, L'Homme, 2022). Les voies empruntées rendent alors compte de la diversité des théories, objectifs et situations analysées (Delavigne, De Vecchi, 2021). Outre-Atlantique, ce tournant se manifeste par les débuts de la terminotique (Auger *et al.*, 1991) et, en France, ce virage donne naissance à la terminologie textuelle (Bourigault, Slodzian, 1999).

Ainsi, si la terminologie a longtemps été pleinement envisagée en rapport avec la traduction (Humbley, 2011), l'influence de nouvelles approches en terminologie développées dans un cadre interdisciplinaire comme c'est le cas de la terminologie textuelle (Bourigault, Slodzian, 1999) a modifié l'existant. Dans ce contexte, nous envisageons l'interdisciplinarité comme la rencontre de la terminologie avec d'autres disciplines scientifiques, écartant en cela l'interdisciplinarité appliquée à un domaine scientifique, technique ou juridique, ce que la terminologie nous semble porter de manière intrinsèque. À une période de renouvellement et de questionnements épistémologiques de la terminologie – en témoigne, par exemple, le Colloque LTT 2024¹ – nous explorons l'interdisciplinarité par le prisme des outils issus de disciplines connexes ; nous menons ces explorations dans une perspective appliquée en terminologie et en traduction spécia-

* LIDILEM, Univ. Grenoble Alpes.

¹ Colloque LTT, Lexicologie, Traduction, Terminologie organisé à Paris les 30 et 31 octobre, intitulé « Approches épistémologiques de la Terminologie. Enjeux actuels », URL : <<https://litt-usn2024.sciencesconf.org/>> (consulté le 19-11-2025).

lisée en adoptant un point de vue didactique, et au regard de l'évolution automatisée des environnements de travail, en l'occurrence la traduction automatique.

C'est ainsi que nous commençons par caractériser brièvement la nature interdisciplinaire d'un courant terminologique qui a marqué un tournant dans l'histoire de la terminologie, la terminologie textuelle (partie 2). Nous explorons ensuite l'évolution interdisciplinaire par le prisme des outils en portant notre attention sur le traitement automatique des langues et sur les outils d'analyse de corpus (partie 3). Le changement de paradigme en linguistique de corpus induit par l'avènement de la traduction automatique nous amène à questionner le traitement de la terminologie dans le domaine de la traduction automatique neuronale (partie 4). Nous concluons en dessinant des perspectives de recherche qui renforcent l'interdisciplinarité de la terminologie textuelle et ouvrons l'interdisciplinarité au-delà du périmètre exploré en faisant appel aux sciences de l'information et de la communication.

2. La Terminologie textuelle à ses débuts : caractérisation interdisciplinaire

Le début des années 1990 marque un tournant dans l'histoire de la terminologie (Faber, L'Homme, 2022) en lien avec de nouvelles orientations. Ce tournant, qualifié de « tournant textuel » (Roche, 2018), se manifeste dans l'Hexagone par la naissance de la terminologie textuelle (Bourigault, Slodzian, 1999). La terminologie textuelle émerge comme un courant à même de renouveler la discipline dans ses fondements théoriques et méthodologiques, alors fortement ancrée dans une approche conceptuelle héritée des travaux d'Eugen Wüster. Intervenant dans un contexte de révolution textuelle marquée par les débuts d'internet, la terminologie textuelle se construit en réponse à des besoins terminologiques liés à des applications textuelles. Dans les années 2000, les besoins en terminologie dans les entreprises et les institutions se multiplient au-delà des produits que sont les bases de données terminologiques pour la traduction. Ces produits concernent les nouvelles applications de la terminologie en entreprise comme, par exemple, les thésaurus pour les systèmes d'indexation automatique ou les réseaux lexicaux pour les moteurs de recherche thématique sur la toile. À cette période, les outils qui accompagnent la terminologie textuelle² se développent au contact d'un environnement terminologique bilingue ; ils sont utilisés pour alimenter la base de données terminologiques des traducteurs d'un organisme scientifique ou pour constituer un lexique bilingue des droits humains.

2.1 Extraction automatique de terminologie et linguistique de corpus

La terminologie textuelle repose sur l'analyse textuelle comme point de départ de l'analyse conceptuelle. Cette analyse s'établit dans une perspective syntagmatique permettant d'étudier le fonctionnement des termes dans les textes. L'extraction des connaissances, et donc des termes, à partir d'un ensemble de textes constitués en corpus s'opère dans le cadre de la linguistique de corpus dans une dimension outillée. Mise en œuvre au

² Nous faisons référence ici au logiciel d'extraction de terminologie LEXTER (Bourigault, 1993).

Canada à travers la terminotique (Auger *et al.*, 1991), cette dimension se développe en France avec des travaux fondateurs pour l'extraction automatique de terminologie (Bourigault, 1994 ; Daille, 1994) dans le cadre de la terminologie computationnelle. La publication du premier ouvrage dédié, *Recent Advances in Computational Terminology* (Bourigault *et al.*, 2001), témoigne du rôle joué par la terminologie textuelle dans le développement des outils. Le logiciel LEXTER, développé par Didier Bourigault (1993), a été très largement utilisé pour la construction de ressources terminologiques³. Son extension vers un analyseur syntaxique automatique, SYNTAX (Bourigault *et al.*, 2005), a montré la nécessité de déployer l'extraction au-delà du terme dans diverses applications, notamment pour observer des phénomènes phraséologiques dans un contexte appliqué de traduction spécialisée.

2.2 L'Ingénierie des connaissances

La dimension interdisciplinaire de la terminologie textuelle s'exerce également à travers sa rencontre avec l'Ingénierie des connaissances. L'objectif commun que poursuivent ces deux disciplines vise à représenter les connaissances sous la forme de réseaux terminologiques et conceptuels, guidés par les corpus pour identifier termes et relations. En France, la naissance du groupe TIA (Terminologie et Intelligence Artificielle) en 1993, créé par Didier Bourigault et Anne Condamines, marque les débuts d'une période interdisciplinaire très vive, réunissant chercheurs en linguistique, terminologie, intelligence artificielle et traitement automatique des langues. Les recherches dans le domaine de la construction de connaissances à partir de textes ont tout particulièrement interrogé la notion de contextes riches en connaissances avec les travaux fondateurs d'Ingrid Meyer (*knowledge-rich contexts*, 2001). Mis au jour à l'aide de marqueurs de relations ou de patrons de connaissances, ces contextes ont fait l'objet de travaux montrant le rôle du domaine et du genre textuel dans leur identification (Condamines, 2002 ; Marshman, L'Homme, 2008). D'autres études, appliquées à la traduction spécialisée, ont mis en évidence le rôle joué par ces contextes : c'est le cas du projet CRISTAL qui a exploité des corpus comparables pour alimenter une base de données destinée à des traducteurs. Les résultats ont montré la préférence des utilisateurs pour des contextes riches en connaissances par rapport à d'autres contextes (Josselin-Leray *et al.*, 2014 ; Planas *et al.*, 2014). Plus récemment, des travaux ont porté sur le rôle d'un corpus parallèle aligné (Condamines *et al.*, 2021) dans l'identification de marqueurs de relations conceptuelles.

3. Évolution interdisciplinaire et terminologie outillée

Ainsi, l'assise interdisciplinaire marque la naissance et le développement de la terminologie textuelle. Son évolution se caractérise par des éléments de stabilité et de changement, mais toujours dans le cadre d'une approche outillée fondée sur l'extraction automatique de ter-

³ Le numéro 19 de la revue « Terminologies Nouvelles » 1999, issu du colloque TIA, concentre nombre d'études utilisant LEXTER.

minologie (Frérot, 2023). Au regard de cette évolution, un domaine nous semble mériter une attention particulière : il s'agit du Traitement Automatique des Langues (TAL).

3.1 Le Traitement Automatique des Langues

Les dialogues féconds des années 2000 entre spécialistes de TAL et terminologues-linguistes se concentrent en particulier sur l'analyse distributionnelle qui est au cœur du projet de la terminologie textuelle. Mise en œuvre à ses débuts dans un outil dédié (UPERY, Bourigault, 2002), cette collaboration, voire cette synergie, entre TAL et terminologie s'est émue au fil des années (Condamines, 2022).

Dans le cadre de l'extraction automatique de terminologie, l'intelligence artificielle (IA) et le déploiement des techniques d'apprentissage machine/apprentissage profond dans les méthodes d'extraction terminologique (Foo, Merkel, 2010 ; Drouin *et al.*, 2018 ; Yadav, Bethard, 2018) ont modifié le rapport aux outils par des non-spécialistes de TAL. Dans ce contexte, l'IA a pour effet de limiter l'utilisation par des terminologues-linguistes de logiciels dédiés aux seuls outils rendus publics et dotés d'une certaine utilisabilité⁴, apte à les rendre exploitables par des non-spécialistes. Parmi les logiciels les plus établis⁵ figure en particulier TERMOSTAT (Drouin, 2003). Plus globalement, des compétences informatiques s'avèrent indispensables pour s'engager dans la voie des langages de programmation (qu'il s'agisse de Perl, Python ou R⁶). La maîtrise de la programmation par des non-spécialistes constitue un enjeu de taille qui interroge l'interdisciplinarité et ce, d'autant plus si l'on adopte un point de vue didactique (dans les formations en terminologie ou en traduction, faut-il par exemple intégrer des enseignements visant l'acquisition de compétences en programmation ?).

Cette voie de recherche est toutefois directement liée au traitement de corpus de très grande taille (plusieurs dizaines de millions de mots) comprenant des données hétérogènes (*Humanitarian Encyclopedia*⁷), marquant un changement de paradigme dans le traitement de corpus et s'éloignant de la notion de corpus spécialisé opératoire en terminologie et en langues de spécialité (Bowker, Pearson, 2002). Or, le traitement de corpus spécialisés constitués selon des critères bien définis, que nous qualifierons ici de corpus « maîtrisés » (contrôle des paramètres de constitution), constitue un pan de recherche en linguistique de corpus appliquée à la traduction spécialisée et à la terminologie dans une perspective didactique, mais aussi sur le terrain, dans les pratiques professionnelles.

3.2 Outils d'analyse de corpus

C'est par la voie des corpus que se déploie l'interdisciplinarité à travers les outils et méthodes issus de la linguistique de corpus. Les outils qui permettent d'interroger un

⁴ Le site de la langue française définit l'utilisabilité comme la « capacité d'un produit à être employé efficacement, et avec satisfaction dans un contexte donné » (consulté le 19-11-2025).

⁵ Nous nous concentrons ici sur les logiciels en accès libre, écartant en cela des logiciels tels que SKETCH ENGINE.

⁶ Perl et Python étant davantage connus des linguistes (González Granado *et al.*, 2022).

⁷ *Humanitarian Encyclopedia*, URL : <<https://humanitarianencyclopedia.org>> (consulté le 19-11-2025).

corpus sont représentés par les concordanciers (Pincemin *et al.*, 2006). Leur intégration dans l'enseignement a fait l'objet de très nombreux travaux en linguistique de corpus appliquée à la traduction et à la terminologie. Toute une école s'est développée dès les années 1990, notamment avec les conférences CULT (*Corpus Use and Learning to Translate*) et TaLC (*Teaching and Language Corpora*). Les travaux dans ce domaine ont massivement porté sur l'exploitation des corpus et d'outils d'interrogation de corpus comme des *documentation tools* (Marco, van Lawick, 2009). De façon très synthétique, les principales études montrent l'apport des corpus monolingues et comparables comme aide à la résolution de problèmes terminologiques et phraséologiques (Aston, 1999 ; Bowker, Pearson, 2002 ; Maia, 2003 ; Kübler *et al.*, 2018) et comme outil d'aide à la prise de décision, notamment à travers l'utilisation de corpus ad hoc ou *DIY corpora* (Varantola, 2003). Les corpus parallèles viennent combler les lacunes des dictionnaires et des bases de données terminologiques (Olohan, 2004 ; Frankenberg-Garcia, 2005) et complètent l'utilisation de corpus comparables (Pearson, 2003).

La littérature est dense jusqu'à la période qui voit la démocratisation des outils de Traduction Automatique (TA). L'avènement de la TA Neuronale (TAN) en 2016, avec la performance qu'on lui reconnaît aujourd'hui et le développement de la post-édition, appelle inévitablement à un repositionnement de la linguistique de corpus outillée sur le plan didactique⁸. Les travaux récents qui s'appuient sur la post-édition (Kübler *et al.*, 2018) montrent l'apport que représente les corpus comparables pour post-éditer des sorties de TAN, en l'occurrence pour la traduction de syntagmes nominaux spécialisés à laquelle se heurte les systèmes neuronaux.

Une autre voie de recherche directement en prise avec les réalités professionnelles des traducteurs et des terminologues repose sur l'utilisation d'outils d'analyse de corpus « prêts à l'emploi » (González Granado *et al.*, 2022), dont la prise en main par des non-spécialistes est opératoire. Nous citerons ici ANTCONC et SKETCH ENGINE. Dans ce dernier outil, le *Corpus Query Language* (CQL) permet de formuler des requêtes plus complexes et les dernières recherches s'orientent vers l'utilisation de *Word sketches* produits dans SKETCH ENGINE avec des requêtes formulées avec le langage de requêtes CQL. Dans ce contexte, les recherches visent à réexploiter les requêtes dans différents corpus via des adaptations pour prendre en compte les variations terminologiques entre corpus (San Martín *et al.*, 2020).

Sur le plan didactique, ces adaptations sont peu « coûteuses » en termes de compétences informatiques (programmation), ce qui constitue une piste prometteuse pour les formations en traduction et en terminologie. La capacité de combiner des analyses statistiques avec des requêtes faciles à programmer ou à adapter présente un réel intérêt notamment en terminologie.

⁸ En d'autres termes, la question se pose aujourd'hui de savoir comment, sur le plan didactique, articuler l'intégration des outils d'analyse de corpus aux nouvelles pratiques des étudiants et des professionnels. Lors d'un Colloque, les interrogations de Marianne Starlander (2022) *L'enseignement des outils d'aide à la traduction est-il passé de mode en cette ère neuronale ?* illustrent ce changement de paradigme. Sur le plan des outils, au-delà des concordanciers, les questionnements sont liés au contenu des cours de TAO traditionnels et à la place des outils comme les mémoires de traduction ou les SGBD terminologiques aux côtés de la TA.

graphie orientée vers la traduction. En effet, dans la démarche de travail du terminologue-traducteur, l'extraction de candidats-termes s'accompagne d'une étape d'extraction de données terminologiques (notamment les CRC évoqués plus tôt), données qui sont utilisées pour alimenter des ressources terminologiques. Enfin, l'utilisation de patrons de connaissances pour la construction de corpus ouvre des perspectives de recherche (Marshman, 2022⁹) et ce, au vu de l'utilisation reconnue des corpus tant au niveau linguistique que conceptuel mais également étant donné la nature chronophage liée à la constitution de corpus.

4. Terminologie et Traduction Automatique

4.1 De quelques considérations générales

Comme nous l'avons vu précédemment, l'empreinte de l'IA en terminologie n'est pas nouvelle (voir 2.2. avec la naissance du groupe TIA). Dans un cadre interdisciplinaire, l'IA appliquée au TAL s'est traduite par des travaux en Traduction Automatique et aujourd'hui, le constat est sans appel : la TAN et l'IA font désormais largement partie du paysage traductionnel, dans un contexte où l'IA appliquée au traitement des langues produit des outils de plus en plus sophistiqués.

L'avènement de la TAN en 2016 a constitué un tournant majeur dans le monde la traduction. La TAN est venue bouleverser les processus de traduction, précipitant le milieu de la traduction dans la « révolution textuelle » (Grass, 2022 : 12) que propose l'IA, créant de nouveaux marchés et soulevant de nombreux débats éthiques liés à en particulier à la protection des données. Le monde de la recherche en traductologie s'est emparé de cet objet comme en témoignent les très nombreuses publications des dernières années, que ce soit dans une perspective linguistique, didactique, cognitive ou outillée (par exemple, Forcada, 2010 ; Cadwell *et al.*, 2016 ; Bowker, Buitrago-Ciro, 2019).

Sur le terrain, les traducteurs apportent leurs expériences, craintes et questionnements ; en témoignent par exemple la revue « Traduire » (2022) consacrée à la traduction automatique, dont les premiers mots de l'édito sont formulés en ces termes « Le biotraducteur et la biotraductrice sont-ils des espèces menacées d'extinction ? », ou bien encore les rencontres de la Société française des traducteurs¹⁰ (2023) dont le programme a laissé une large place à la question de la TA et de l'IA à travers des conférences sur « L'IA en question : amie ou ennemie ? » ou bien « Post-éditer ou humaniser du contenu robotisé ».

Face à ces mutations, des réflexions s'engagent afin d'appréhender les progrès des systèmes de TAN au regard de l'activité de post-édition, d'une part, et au regard des domaines dans lesquels la TAN échoue à produire une traduction de qualité, d'autre part. C'est sur ce dernier aspect que nous portons notre attention. Dans les domaines où le style et le sens sont prépondérants, la TAN ne peut produire une traduction à l'égale d'une traduction humaine dans la-

⁹ « The inclusion of KRCs containing KPs in specialized language resources for translators, writers and other language professionals – as well as trainees in these professions – is an area ripe for expansion » (Marshman, 2022 : 309).

¹⁰ Société française des traducteurs, URL : <<https://www.sft.fr/>> (consulté le 19-11-2025).

quelle la créativité du traducteur a toute sa place. Le domaine de la transcréation, par exemple, illustre parfaitement cette situation. Par ailleurs, devant « l'incapacité des systèmes [neuronaux] à détecter des erreurs de sens relevant de la connaissance du monde et de rapports de causalité » (Grass, 2022), la valeur ajoutée de l'humain ne fait aucun doute. Et la fluidité sur le plan linguistique à laquelle parvient la TAN aujourd'hui, fluidité « trompeuse », exige un travail humain de post-édition afin d'identifier les erreurs de sens. Il en va de même pour les discours spécialisés où la TAN se heurte à des difficultés de terminologie, comme dans le cas de la traduction de termes complexes anglais-français ou celui de la variation (Calvi, Dankova, 2022).

L'adaptation des sorties de TA à la terminologie d'un domaine spécialisé fait l'objet de nombreuses recherches en TA/TAL mais pas en terminologie (Hasler *et al.*, 2018). Il nous semble que c'est précisément sur cette question que la terminologie a un rôle à jouer et peut intervenir dans ce champ de recherches.

4.2 Les systèmes de TA et la terminologie

Le traitement de la terminologie d'un domaine spécialisé a fait l'objet de travaux en TA/TAL tout d'abord dans les systèmes basés sur les règles puis dans les systèmes statistiques. Les principes de certaines études menées dans le contexte de l'intégration de la terminologie dans les systèmes statistiques ont ensuite été repris pour la TAN (Chen *et al.*, 2017 ; Chu, Wang, 2018)¹¹.

Dans ces derniers systèmes, la qualité de la terminologie est réduite en raison de la difficulté des systèmes à repérer des corpus parallèles dans le domaine cible. « Les corpus varient considérablement d'un corpus à l'autre, et les traductions faites à partir de corpus portant sur différents domaines sont peu fiables » (Chen *et al.*, 2017 : 40), ce qui a pour effet de dégrader la précision au bénéfice de la fluidité (Koehn, Knowles, 2017¹²). En France, de récents travaux placent au cœur des développements en traduction automatique des questions de nature terminologique (Bénard *et al.*, 2023¹³ ; Bénard, Bordet, Kübler, 2023).

Plus globalement, et dans un contexte directement lié aux utilisateurs, les approches proposées pour résoudre le problème de la terminologie du domaine différente de celle du corpus d'entraînement consistent notamment à utiliser un système neuronal déjà entraîné en s'appuyant, par exemple, sur des mémoires de traduction ou sur l'utilisation de moteurs de TA en ligne : c'est notamment le cas de DeepL Pro qui dès 2020 a permis aux utilisateurs d'intégrer leur propre terminologie sous la forme de bases de données terminologiques ou glossaires, utilisés ensuite dans la traduction. Cette approche, qui permet d'évaluer des sorties de TA avec et sans glossaire, est à notre connaissance une voie très peu explorée en terminologie (Guerrerro, 2020 ; Baldassarre, 2021), voie qui peut précisément contribuer à renforcer l'interdisciplinarité de la terminologie.

¹¹ Pour une synthèse des travaux, cf. Baldassarre, 2021 : 28-32.

¹² « NMT systems have lower quality out of domain, to the point that they completely sacrifice adequacy for the sake of fluency » (Baldassarre, 2021 : 28).

¹³ Le projet MaTOS (Machine Translation for Open Science) vise à développer de nouvelles méthodes pour la traduction automatique (TA) intégrale de documents scientifiques entre le français et l'anglais, ainsi que des métriques automatiques pour évaluer la qualité des traductions produites (Bénard *et al.*, 2023).

Dans le prolongement de ces derniers travaux, des pistes de recherches se dessinent dans le cadre de la terminologie textuelle avec une approche qui place au cœur des travaux la terminologie outillée (extraction automatique de terminologie, concordancier) et l'intervention du terminologue-linguiste pour sélectionner les candidats-termes et construire une ressource bilingue intégrée au moteur de TA. Sur un plan linguistique, cette intégration pose en particulier la question de la réalisation ou manifestation des termes en discours avec le phénomène de la variation ainsi que la question de la synonymie. Les résultats de Baldassarre (2021) montrent en effet la défaillance du système dans le traitement de la synonymie et de la variation terminologique, avec un effet de répétition¹⁴.

5. Conclusions et perspectives

Au regard des travaux largement éprouvés dans l'approche basée sur corpus dont relève la terminologie textuelle, la terminologie pourrait bien trouver dans le contexte de la TA une place qui renforce son interdisciplinarité. La constitution de ressources terminologiques pour la TAN offre un espace de recherche interdisciplinaire de nature à renouveler les dialogues entre TAL et terminologie pour la TA. Un carrefour interdisciplinaire qui s'annonce fructueux, étant donné le contexte de renouvellement et de questionnements épistémologiques que la terminologie traverse aujourd'hui.

Par-delà le périmètre traductionnel, la terminologie semble également pouvoir dessiner l'avenir dans une dynamique interdisciplinaire avec les Sciences de l'Information et de la Communication (SIC). En France, cette dynamique s'est déjà engagée à travers des travaux sur la médiatisation de la science et la diffusion des connaissances (Ledouble, 2023). En effet, les questions de vulgarisation et de médiatisation des savoirs spécialisés entrent de plein champ dans des problématiques terminologiques et des questions de recherche liées à la circulation des savoirs. Dans une perspective outillée, l'ouverture souhaitée de la terminologie à d'autres disciplines peut également s'exercer à travers les méthodes et outils dévolus à chaque discipline¹⁵.

Ainsi, que la terminologie revisite et renforce ses liens avec des disciplines au contact desquelles elle s'est solidement constituée (la traduction), où qu'elle développe l'interdisciplinarité dans un mouvement plus vaste (les SIC), nul doute que les questions soulevées de part et d'autre constituent des axes de recherche porteurs pour l'avenir.

Références bibliographiques

Aston G., *Corpus use and learning to translate*, in « Textus », 12, 1999, pp. 289-314.

¹⁴ Le moteur DeepL prend en charge des ressources bilingues constituées d'une équivalence parfaite terme source-terme cible.

¹⁵ IRaMuTeQ, logiciel libre d'analyse statistique de corpus utilisés, est majoritairement utilisé en SIC ; en terminologie, il s'agit principalement de TERMOSTAT, d'ANTCONC et de SKETCH ENGINE.

- Auger P. et al., *Un projet d'automatisation des procédures en terminographie*, in « Meta », 3(1), 1991, pp. 121-127.
- Baldassarre V.V., *Intégration de la terminologie en traduction automatique : le cas de la Confédération suisse avec DeepL Pro*, Mémoire de master, Université de Genève, 2021.
- Bénard M. et al., *MaTOS : Traduction automatique pour la science ouverte*, in *Actes de CO-RIA-TALN 2023. Actes de l'atelier « Analyse et Recherche de Textes Scientifiques » (ARTS)@ TALN 2023*, Paris, ATALA, 2023, pp. 8-15.
- Bénard M., Bordet G., Kübler N., *Réflexions sur la traduction automatique et l'apprentissage en langues de spécialité*, in « Asp », 82, 2023, pp. 81-98.
- Bourigault D., *Analyse syntaxique locale pour le repérage de termes complexes dans un texte*, in « Traitement automatique des langues », 34(2), 1993, pp. 105-117.
- Bourigault D., *LEXTER, un logiciel d'Extraction de Terminologie. Application à l'acquisition des connaissances à partir de textes*, Thèse de doctorat, École des hautes études en sciences sociales, Paris, 1994.
- Bourigault D., *Upéry : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus*, in *Actes de la 9^e conférence annuelle sur le Traitement Automatique des Langues (TALN 2002)*, Nancy, 2002, pp. 75-84. URL : <<https://aclanthology.org/2002.jeptalnrecital-long.5/>>, consulté le 19-11-2025.
- Bourigault D. et al. (eds.), *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia, John Benjamins, 2001.
- Bourigault D. et al., *Syntex, analyseur syntaxique de corpus*, in *Actes des 12^e journées de la Conférence TALN*, Dourdan, 2005. URL : <<https://hal.science/hal-00005571v1/document>>, consulté le 19-11-2025.
- Bourigault D., Slodzian M., *Pour une terminologie textuelle*, in « Terminologies nouvelles », 19, 1999, pp. 29-32.
- Bowker L., Buitrago Ciro J., *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, Bingley, Emerald Publishing, 2019.
- Bowker L., Pearson J., *Working with Specialized Language: A Practical Guide to Using Corpora*, London/New York, Routledge, 2002.
- Cabré M.T., *La terminologie : théorie, méthode et applications*, Ottawa, Presses de l'Université d'Ottawa, 1998.
- Cadwell P. et al., *Human Factors in Machine Translation Post-Editing among Institutional Translators*, in « Translation Spaces », 5(2), 2016, pp. 222-243.
- Calvi S., Dankova K., *Industrie de la langue et formation des traducteurs spécialisés*, in « Traduction et Langues », 21, 2022, pp. 190-204.
- Chen B. et al., *Cost Weighting for Neural Machine Translation Domain Adaptation*, in *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver, Association for Computational Linguistics, 2017, pp. 40-46.
- Chu C., Wang R., *A Survey of Domain Adaptation for Neural Machine Translation*, in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 1304-1319. URL : <<https://aclanthology.org/C18-1111/>>, consulté le 19-11-2025.
- Condamines A., *Corpus Analysis and Conceptual Relation Patterns*, in « Terminology », 8(1), 2002, pp. 141-162.
- Condamines A., *Terminologie, Intelligence Artificielle et Psychologie cognitive : réflexions sur les interactions possibles dans l'étude de la variation en langues spécialisées*, in « De Europa », 15, 2022, pp. 131-148.

- Condamines A. et al., *Apport de la traduction dans l'analyse des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien*, in C. Frérot, M. Pecman (éds.), *Des corpus numériques à l'analyse linguistique en langues de spécialité*, Grenoble, UGA Editions, 2021, pp. 313-335.
- Daille B., *Approche mixte pour l'extraction de terminologie : statistiques lexicales et filtres linguistiques*, Thèse de doctorat, Université Paris VII, 1994.
- Delavigne V., De Vecchi D., *Termes en discours, entreprises et institutions*, Paris, Presses Sorbonne Nouvelle, 2021.
- Drouin P., *Term extraction using non-technical corpora as a point of leverage*, in « Terminology », 9(1), 2003, pp. 99-115.
- Drouin P. et al., *Lexical profiling of environmental corpora*, in *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, Paris, 2018, pp. 3419-3425. URL : <<https://aclanthology.org/L18-1539/>>, consulté le 19-11-2025.
- Faber P., L'Homme M.-C. (eds.), *Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized knowledge*, Amsterdam/Philadelphia, John Benjamins, 2022.
- Foo J., Merkel M., *Using machine learning to perform automatic term recognition*, in *LREC 2010 Workshop on Methods for automatic acquisition of Language Resources and their evaluation methods*, Valletta, Malta, 2010, pp. 49-54. URL : <https://www.ida.liu.se/~jodfo01/publications/foo-merkel-using_machine_learning_to_perform_automatic_term_recognition.pdf>, consulté le 19-11-2025.
- Forcada M., *Machine translation today*, in Y. Gambier, L. Van Doorslaer (eds.), *Handbook of Translation Studies*, Amsterdam/Philadelphia, John Benjamins, 2010, pp. 215-223.
- Frankenberg-Garcia A., *Pedagogical Uses of Monolingual and Parallel Concordances*, in « ELT Journal », 59(3), 2005, pp. 189-198.
- Frérot C., *Des origines de la terminologie textuelle. Fondements, ancrage théorique et méthodologique, extraction automatique de terminologie*, TOTh, Université Savoie Mont Blanc, 2023.
- González Granado N. et al., *De l'analyse statistique à l'apprentissage automatique : le langage R au service de la terminologie*, in « ELA. Études de linguistique appliquée », 208(4), 2022, pp. 447-467.
- Gouadec D., *Terminologie, traduction et rédaction spécialisées*, in « Langues », 157, 2005, pp. 14-24.
- Grass T., *L'erreur n'est pas humaine*, in « Traduire », 246, 2022, pp. 10-22.
- Guerrero L., *NMT plus a Bilingual Glossary: Does this Really Improve Terminology Accuracy and Consistency?*, in « Translating and the Computer Conference », 2020. URL : <<https://asling.org/tc42/wp-content/uploads/TC42-20Nov-NMT-plus-a-bilingual-glossary.pdf>>, consulté le 19-11-2025.
- Hasler E. et al., *Neural Machine Translation Decoding with Terminology Constraints*, in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, 2018, pp. 506-512. URL : <<https://aclanthology.org/N18-2081/>>, consulté le 19-11-2025.
- Humbley J., *Terminologie et traduction : une complémentarité oubliée ?*, in « Colloque Tralogy, Session 1 – Terminologie et Traduction », Paris, 2011. URL : <https://hal.science/hal-02495532/file/Tralogy_S1_A1_Humbley.pdf>, consulté le 19-11-2025.
- Josselin-Leray A. et al., *Good Contexts for Translators – A First Account of the Cristal Project*, in *XVI EURALEX International Congress: The User in Focus*, Bolzano, Italie, 2014. URL : <<https://hal.science/hal-00986391v1/document>>, consulté le 19-11-2025.

- Koehn P., Knowles R., *Six Challenges for Neural Machine Translation*, in *Proceedings of the First Workshop on Neural Machine Translation*, 2017, pp. 28-39. URL : <<https://aclanthology.org/W17-3204/>>, consulté le 19-11-2025.
- Kübler N. et al., *Teaching Specialised Translation Through Corpus Linguistics: Translation Quality Assessment and Methodology Evaluation and Enhancement by Experimental Approach*, in « Meta », 63(3), 2018, pp. 807-825.
- Ledouble H., *Médiatisation de la science et diffusion des connaissances : approche terminologique et cognitive en agroécologie*, London, ISTE Editions, 2023.
- Maia B., *Training translators in terminology and information retrieval using comparable and parallel corpora*, in F. Zanettin, S. Bernardini, D. Stewart (eds.), *Corpora in Translator Education*, Manchester, St. Jerome Publishing, 2003, pp. 47-58.
- Marco I., van Lawick H., *Using corpora and retrieval software*, in A. Beeby, P. Sánchez-Gijón, P. Rodríguez Ines (eds.), *Corpus use and translation: Corpus use for learning to translate and learning corpus use to translate*, Amsterdam/Philadelphia, John Benjamins, 2009, pp. 9-28.
- Marshman E., *Knowledge patterns in corpora*, in P. Faber, M.-C. L'Homme (eds.), *Theoretical Perspectives in Terminology. Explaining terms, concepts and specialized knowledge*, Amsterdam/Philadelphia, John Benjamins, 2022, pp. 291-310.
- Marshman E., L'Homme M.-C., *Portabilité des marqueurs de la relation causale : étude sur deux corpus spécialisés*, in F. Maniez, P. Dury, N. Arlin, C. Rougemont (éds.), *Corpus et dictionnaires de langues de spécialité*, Grenoble, Presses Universitaires de Grenoble, 2008, pp. 87-110.
- Meyer I., *Extracting knowledge-rich contexts for terminography*, in D. Bourigault, C. Jacquemin, M.-C. L'Homme (eds.), *Recent advances in computational terminology*, Amsterdam/Philadelphia, John Benjamins, 2001, pp. 279-302.
- Olohan M., *Introducing Corpora in Translation Studies*, London/New York, Routledge, 2004.
- Pearson J., *Using parallel texts in the translator training environment*, in F. Zanettin, S. Bernardini, D. Stewart (eds.), *Corpora in Translator Education*, Manchester, St. Jerome Publishing, 2003, pp. 15-24.
- Pincemin B. et al., *Concordanciers : thème et variations*, in J.-M. Viprey et al. (éds.), *Actes des JADT 2006*, Besançon, Presses Universitaires de Franche-Comté, 2006, pp. 773-784.
- Planas E. et al., *Exploring the Use and Usefulness of KRCs in Translation: Towards a Protocol*, in *11th International Conference on Terminology and Knowledge Engineering: Ontology, Terminology and Text Mining*, Berlin, 2014. URL : <<https://hal.science/hal-01005839v1/document>>, consulté le 19-11-2025.
- Roche J., *Le tournant ontologique de la Terminologie*, Thèse de doctorat, Université Grenoble Alpes, 2018.
- San Martín A. et al., *Extraction of Hyponymic Relations in French with Knowledge-Pattern-Based Word Sketches*, in *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC)*, Marseille, 2020, pp. 5953-5961. URL : <<https://aclanthology.org/2020.lrec-1.729/>>, consulté le 19-11-2025.
- Savatovsky D., Candel D., *Présentation*, in « Langages », 168/4, 2007, pp. 3-10.
- Starlander M., *L'enseignement des outils d'aide à la traduction est-il passé de mode en cette ère neuronale ?*, intervention lors du Colloque international « Enseigner la traduction à l'heure neuronale », Bruxelles, 2022.
- Varantola K., *Translators and Disposable Corpora*, in F. Zanettin, S. Bernardini, D. Stewart, (eds.), *Corpora in Translator Education*, Manchester, St. Jerome Publishing, 2003, pp. 55-70.

Yadav V., Bethard S., *A survey on recent advances in named entity recognition from deep learning models*, in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, 2018, pp. 2145-2158. URL : <<https://arxiv.org/abs/1910.11470>>, consulté le 19-11-2025.

Terminologie de l'endométriose et interdisciplinarité : propositions méthodologiques et étude de cas

JULIE HUMBERT-DROZ

1. Introduction et objectifs

Dans cet article, nous nous intéressons à l'interdisciplinarité telle qu'elle intervient dans le domaine médical, notamment dans le contexte de maladies chroniques, et aux difficultés spécifiques qu'elle peut poser d'un point de vue terminologique. Il est en effet reconnu que l'interdisciplinarité peut être source de variation, au sein de la communauté des spécialistes et en dehors (p. ex. Condamines, Dehaut, 2011 ; Picton, Dury, 2017 ; Delavigne, 2022). Nous nous focalisons en particulier sur la terminologie de l'endométriose, qui est une maladie gynécologique chronique et complexe. La recherche sur cette maladie, son diagnostic et la prise en charge des patientes sont aujourd'hui interdisciplinaires (HAS, CNGOF, 2017 ; EndoFrance, 2025), ce qui peut se traduire par des variations dans l'usage des termes.

Dans ce contexte, nous cherchons à explorer les rapports entre interdisciplinarité et variation et à observer dans quelle mesure ces deux aspects peuvent être corrélés. L'objectif de cette étude est en réalité double : dans un premier temps, nous visons à discuter les difficultés méthodologiques liées à l'interdisciplinarité pour une analyse terminologique, en particulier ses implications pour la constitution d'un corpus adéquat, qui doit refléter les usages existants, dans toute leur diversité, et qui doit dans le même temps rester opérationnel pour une analyse outillée, et nous effectuons une proposition en ce sens. Dans un second temps, nous cherchons à observer les conséquences de cette interdisciplinarité sur la variation définitionnelle du terme *endométriose*, en identifiant d'abord les variations dans le corpus puis en les mettant en relation avec les disciplines dont les discours sont représentés dans le corpus.

Cette étude s'inscrit par ailleurs dans un projet plus vaste qui s'intéresse à la circulation des termes de l'endométriose entre spécialistes, patientes et non-spécialistes, et à leur appropriation par les patientes et par le grand public en France¹. La constitution d'un corpus spécialisé et le repérage de la variation autour de la définition du terme *endométriose* en constituent les premières étapes.

¹ Ce projet, intitulé « La circulation des termes de l'endométriose : de la compréhension à l'appropriation des termes », est financé par le Fonds national suisse (FNS).

Cet article est structuré de la manière suivante : dans la section 2, nous définissons l'endométriose en mettant l'accent sur les défis liés à l'interdisciplinarité dans ce contexte et nous faisons un court état des lieux des problématiques terminologiques qui peuvent se poser. La section 3 pose les bases méthodologiques sous-tendant notre proposition, en détaillant d'abord les critères qui ont guidé la constitution du corpus, puis en précisant les outils exploités et la méthode adoptée pour le repérage et l'analyse des contextes délimitaires du terme *endométriose*. Dans la section 4, nous décrivons les résultats obtenus, d'abord quantitativement puis qualitativement, en insistant sur les points qui varient le plus dans le corpus. Enfin, nous discutons ces résultats en relation avec notre objectif et proposons quelques remarques conclusives et pistes pour poursuivre nos recherches.

2. Contexte théorique

2.1 L'endométriose

L'endométriose est une maladie gynécologique complexe. Elle se définit par « la présence en dehors de la cavité utérine de tissu semblable à la muqueuse utérine » qui subit « l'influence des modifications hormonales » au cours de chaque cycle menstruel (EndoFrance, 2025). Il s'agit d'une maladie inflammatoire chronique et incurable qui touche une à deux femmes sur dix, selon les estimations actuelles, et qui provoque de nombreux symptômes douloureux (Young *et al.*, 2015 ; Ilschner *et al.*, 2022). Elle est généralement qualifiée de complexe en raison des différentes théories élaborées pour expliquer ses causes, de la combinaison de facteurs génétiques, environnementaux et hormonaux qui la font évoluer et de la multiplicité des formes qu'elle peut prendre (Galès, 2020 ; EndoFrance, 2025). Parallèlement, la méconnaissance de cette maladie entraîne un retard diagnostique important, entre 7 et 10 ans (Young *et al.* 2015), ce qui retarde d'autant la mise en place de soins adaptés (EndoFrance 2025).

La nature multifactorielle de l'endométriose et l'étendue possible des lésions à plusieurs organes exigent la collaboration de spécialistes de plusieurs disciplines médicales pour la diagnostiquer et prendre en charge les femmes atteintes. De même, la recherche sur l'endométriose s'effectue au sein de plusieurs disciplines, et non uniquement en gynécologie. Cette forte interdisciplinarité, dans la pratique et dans la recherche, entraîne une multiplication des points de vue sur la maladie, qui peuvent par ailleurs se traduire par certaines variations (sémantiques, lexicales) dans les textes. C'est précisément ce qui rend une analyse terminologique pertinente, selon nous, qui permettra de mieux saisir ces variations, à l'intérieur de la communauté des spécialistes et en dehors. Bien que cet article se focalise sur la variation autour de la définition du terme *endométriose* chez les spécialistes, d'autres travaux en cours interrogent les phénomènes qui se produisent lorsque les termes circulent en dehors des discours spécialisés (Humbert-Droz, 2024).

2.2 Interdisciplinarité et problématiques terminologiques

La question de l'interdisciplinarité est bien connue des terminologues. Il s'agit en premier lieu de l'une des difficultés caractéristiques de la phase de constitution d'un corpus servant à étudier ou décrire la terminologie d'un domaine, particulièrement lorsque celui-ci doit être circonscrit, étant donné que la plupart des domaines sont justement interdisciplinaires (Meyer, Mackintosh, 1996 : 268 ; Bowker, Pearson, 2002 : 50 ; Péraldi, 2011 : 72). On retrouve également cette question dans le cas de domaines transversaux, c'est-à-dire des domaines qui ont des implications pour d'autres domaines, comme l'environnement (p. ex. Delavigne, 2022), ou des domaines dont les innovations deviennent aujourd'hui essentielles pour d'autres domaines, comme l'informatique (Bordet, 2013 ; Grimaldi, 2024). L'interdisciplinarité intervient aussi « en situation d'innovation disciplinaire » (Condamines, Dehaut, 2011 : 266), c'est-à-dire lorsque des chercheurs de disciplines distinctes collaborent dans le contexte de l'émergence d'un nouveau domaine.

Dans toutes ces situations, un domaine partage plusieurs termes avec des domaines proches (Péraldi, 2011 : 153), ce qui cause nécessairement certaines variations dans l'usage des termes. Par exemple, Condamines et Dehaut mettent en lumière des fonctionnements synonymiques et polysémiques, qui sont potentiellement conflictuels et qui peuvent avoir un impact sur la communication entre les spécialistes des différents domaines (Condamines, Dehaut, 2011 : 278*sqq.*). D'autres cas de variation sont observés lorsque différents groupes de spécialistes travaillent dans un même domaine. Par exemple, Picton et Dury observent la variation dénomminative en médecine nucléaire et en pédagogie universitaire, en fonction des différents groupes de spécialistes impliqués (respectivement techniciens et médecins, et chercheurs et conseillers pédagogiques) (Picton, Dury, 2017).

Dans le cas de l'endométriose, l'interdisciplinarité intervient au moins à deux niveaux, en raison des différentes disciplines médicales concernées (de la gynécologie à la radiologie, de la pédiatrie à l'algologie) et des différents groupes de spécialistes impliqués (médecins, sages-femmes, infirmiers, manipulateurs, etc.). Ces deux niveaux peuvent par ailleurs « se combiner », car chaque discipline regroupe des spécialistes de statuts différents (par exemple, des infirmiers et des médecins en gynécologie, des médecins et des manipulateurs en radiologie). Enfin, l'une des particularités de l'endométriose, notamment dans le contexte français, s'illustre à travers une volonté explicite d'interdisciplinarité, promue principalement par les recommandations de la Haute Autorité de santé (HAS) et du Collège national des gynécologues et obstétriciens français (CNGOF) en matière de diagnostic et de traitement de cette maladie (voir *infra*).

3. Aspects méthodologiques

3.1 Construire un corpus spécialisé autour de l'endométriose

De manière générale, notre approche s'inscrit dans le champ de la terminologie outillée (Condamines, Picton, 2022) et l'analyse s'articule autour d'un corpus spécialisé consti-

tué spécifiquement pour notre projet. Bien que celui-ci se fonde sur les discours de quatre communautés-clés dans la circulation des termes de l'endométriose, nous nous focalisons ici uniquement sur les discours des spécialistes. Ceux-ci sont rassemblés dans un corpus nommé *EndoExp* qui regroupe 166 textes et compte 1 004 108 occurrences. Sa constitution a représenté un défi majeur dans la réalisation de ce projet et a fait l'objet d'une réflexion approfondie, principalement sur les moyens de refléter, dans le corpus, l'interdisciplinarité caractéristique du contexte de l'endométriose.

Plus précisément, vu « l'éclatement » des discours sur l'endométriose qui proviennent de disciplines médicales variées (imagerie médicale, urologie, gynécologie, pharmacie, etc.), considérer un seul domaine, ou sous-domaine médical, n'est pas satisfaisant pour refléter cette diversité. Inversement, considérer « le » domaine médical dans son ensemble comporte le risque de récupérer des textes dont le propos n'est pas suffisamment ciblé sur l'endométriose ou dont les auteurs ne sont pas suffisamment spécialistes de cette maladie. Pour cette raison, nous avons compilé un corpus non pas à partir de textes d'un domaine bien délimité, comme c'est souvent le cas en terminologie (p. ex. L'Homme, 2020 : 133), mais de « discours sur » l'endométriose, c'est-à-dire « un ensemble de pratiques qui reflètent la réalité des échanges et des genres textuels en présence » (Delavigne, 2022 : 28), reprenant ainsi une proposition de la socioterminologie.

Suivant ces principes, ce corpus regroupe des textes de différent genres (articles de recherche, thèses, mémoires de fin d'études, etc.), produits par des locuteurs de différents statuts (médecins, sages-femmes, infirmiers, etc.) dans 16 disciplines différentes. La taille des échantillons varie selon les disciplines, parfois considérablement. Notons par exemple qu'une grande majorité des textes s'inscrit en gynécologie, alors qu'une minorité provient de disciplines comme les maladies respiratoires ou la pédiatrie (Tableau n. 1). Cependant, comme l'endométriose est une maladie avant tout gynécologique, qui peut avoir des conséquences sur d'autres organes que les organes reproducteurs, ces déséquilibres ne sont pas surprenants et ils reflètent sans doute les disparités qui existent dans les discours des spécialistes. Il sera tenu compte de ces déséquilibres dans l'analyse.

Tableau n. 1 - *Composition du corpus EndoExp*

<i>Discipline</i>	<i>Nombre de textes</i>	<i>Nombre d'occurrences</i>
Gynécologie	103	532 419
Médecine générale	14	187 711
Radiologie	17	109 002
Obstétrique	9	58 303
Pharmacie	3	33 135
Algologie	3	21 731
Chirurgie viscérale	5	19 398
Urologie	2	11 620
Pathologie	1	8 559
Sciences infirmières	1	5 681

<i>Discipline</i>	<i>Nombre de textes</i>	<i>Nombre d'occurrences</i>
Pédiatrie	2	5 214
Autres (sexologie, médecine d'urgence, médecine interne, maladies respiratoires, médecine nucléaire)	6	11 335
<i>Total</i>	166	1 004 108

Le corpus couvre une période courte et récente, de 2018 à 2022. Ce choix se fonde sur la publication fin 2017 des recommandations de bonne pratique pour la prise en charge de l'endométriose par la HAS et le CNGOF, largement diffusées dès 2018. Ce texte remplace les recommandations du CNGOF de 2006, tient compte des dernières avancées de la recherche et promeut un accompagnement interdisciplinaire des patientes (HAS, CNGOF, 2017). Il a vocation à être un texte de référence pour le diagnostic et le traitement de l'endométriose par les gynécologues, les chirurgiens, les médecins généralistes, les sages-femmes et les autres professionnels de santé impliqués. La diffusion de ces recommandations se veut large et leur impact ainsi que leur retentissement dans les médias importants. Enfin, en tant que premier texte officiel recommandant une approche de l'endométriose qui doit faire collaborer plusieurs disciplines médicales, l'année de sa diffusion, 2018, est une année charnière ; c'est pourquoi elle a été choisie comme point de départ.

3.2 Outils et méthodes pour l'exploration des données

Dans cet article, nous nous intéressons principalement à la définition du terme *endométriose* telle qu'elle est formulée dans les discours des spécialistes. Pour repérer les contextes qui explicitent une définition, entière ou partielle, dans le corpus, nous nous basons sur la notion de contextes riches en connaissances, c'est-à-dire des contextes qui indiquent « at least one item of domain knowledge that could be useful for conceptual analysis » (Meyer, 2001 : 281).

Nous identifions ces contextes à l'aide de Sketch Engine (Kilgariff *et al.*, 2004), dont la fonctionnalité de recherche en CQL (*Corpus Query Language*) permet d'effectuer des requêtes très précises et donc de cibler la recherche uniquement sur les patrons à l'étude. Dans notre cas, il s'agit des définitions introduites par le verbe *être*, qu'il s'agisse de définitions hyperonymiques (p. ex. le syntagme « l'endométriose est une maladie » suivi des caractéristiques spécifiques) ou de définitions introduites par un verbe à la forme passive (p. ex. « l'endométriose est définie comme » ou « l'endométriose est caractérisée par »). Bien que ces patrons soient particulièrement productifs en discours (p. ex. Rebeyrolle, 2000 ; Meyer, 2001 ; Lefeuve, 2017 ; Sierra *et al.*, 2022), tous les résultats obtenus avec Sketch Engine ne sont pas pertinents. Une évaluation manuelle de chacun des contextes retournés permet de ne conserver que ceux qui attestent une définition.

Les hyponymes d'endométriose (par exemple *endométriose profonde*, *endométriose superficielle*) sont exclus de l'analyse, de même que les reprises anaphoriques. Seuls les contextes définitoires apparaissant dans la même phrase que le terme *endométriose* sont retenus.

4. Quand les spécialistes définissent l'endométriose

4.1 Aperçu quantitatif et observations globales

L'analyse du corpus a retourné 1275 contextes correspondant au patron recherché, dont 153 (12 %) seulement sont pertinents, c'est-à-dire qu'ils explicitent une définition, entière ou partielle. Le détail du nombre de contextes repérés et les proportions de contextes pertinents pour chaque discipline représentée dans le corpus sont illustrés dans le Tableau n. 2.

On remarque ainsi d'emblée que les disciplines pour lesquelles un plus grand nombre de contextes pertinents a été extrait sont la gynécologie, la médecine générale et l'obstétrique. Ce constat n'est pas surprenant dans la mesure où ce sont les disciplines les plus directement concernées par l'endométriose et celles dont les textes sont les plus nombreux dans le corpus (cf. Tableau n. 1). Notons également la pharmacie, avec 12 contextes pertinents, et l'imagerie, avec 11 contextes pertinents. Là encore, ces disciplines sont parmi les mieux représentées dans le corpus.

Nous remarquons également qu'aucun contexte définitoire correspondant aux patrons recherchés n'a été repéré pour 4 des 16 disciplines prises en compte (médecine interne, pédiatrie, sciences infirmières, urologie).

Tableau n. 2 - *Nombre et proportion de contextes pertinents par rapport au nombre de contextes repérés pour chaque discipline*

<i>Discipline</i>	<i>Nombre de contextes repérés</i>	<i>Nombre de contextes définitoires</i>	<i>Proportion (%)</i>
Algologie	22	2	9,1
Chirurgie viscérale	22	2	9,1
Gynécologie	654	60	9,2
Imagerie	84	11	13,1
Maladies respiratoires	4	1	25
Médecine d'urgence	9	1	11,1
Médecine générale	232	34	14,7
Médecine nucléaire	1	1	100
Obstétrique	132	26	19,7
Pathologie	16	1	6,3
Pharmacie	85	12	14,1
Sexologie	14	2	14,3
<i>Total</i>	1275	153	12

Les contextes définitoires ainsi relevés peuvent être classés en trois catégories principales, selon le type de définition exprimé. On observe deux types de définitions caractéristiques du domaine médical, à savoir des définitions *symptomatiques*, qui mettent l'accent sur les symptômes associés à la maladie, et des définitions *histologiques*, qui s'attachent à la description de la maladie sur les plans anatomique et biologique. Une troisième catégorie regroupe les contextes qui ne correspondent ni au type symptomatique ni au type histologique et que nous nommons « autre » dans la suite de cet article. Le nombre de contextes repérés pour chaque type est illustré dans le Tableau n. 3. À quatre reprises, les deux points de vue sont exprimés dans la même phrase ; c'est pourquoi le total de ce Tableau est supérieur au total du Tableau n. 2.

Tableau n. 3 - *Nombre de contextes définitoires repérés par type de définition*

<i>Type de définition</i>	<i>Nombre de contextes</i>
Histologique	37
Symptomatique	41
Autre	79
<i>Total</i>	157

Le Tableau n. 4 illustre la répartition des contextes définitoires par type et par discipline. On observe que :

- pour la médecine d'urgence et les maladies respiratoires, les deux contextes repérés correspondent au type de définition histologique ;
- pour la sexologie, les deux contextes correspondent au type symptomatique ;
- pour l'algologie, la chirurgie viscérale, la médecine nucléaire et la pathologie, les contextes repérés sont de type « autre » ;
- pour l'imagerie, les contextes sont répartis de manière équitable entre le type histologique et le type « autre » ;
- pour la gynécologie, la médecine générale, l'obstétrique et la pharmacie, les trois types de contextes définitoires ont été repérés.

À l'exception de l'imagerie, les disciplines qui ont retourné la majorité des contextes définitoires (gynécologie, médecine générale, obstétrique et pharmacie) comprennent des contextes des trois types. Par ailleurs, parmi ces quatre disciplines, le type « autre » n'est minoritaire qu'en médecine générale et ce sont toujours les contextes de type symptomatique qui sont majoritaires (sauf en médecine générale, où l'on observe le même nombre de contextes de type histologique que de type symptomatique). En revanche, même si les contextes de type symptomatique sont plus nombreux dans l'ensemble du corpus, les contextes de type histologique sont mieux répartis (dans 7 disciplines, contre 5 pour le type symptomatique).

Tableau n. 4 - Nombre de contextes par type de définition et par discipline

Discipline	Nombre de contextes correspondant au type de définition « histologique »	Nombre de contextes correspondant au type de définition « symptomatique »	Nombre de contextes correspondant au type de définition « autre »
Algologie	0	0	2
Chirurgie viscérale	0	0	2
Gynécologie	10	14	37
Imagerie	6	0	5
Maladies respiratoires	1	0	0
Médecine d'urgence	1	0	0
Médecine générale	13	13	11
Médecine nucléaire	0	0	1
Obstétrique	5	8	13
Pathologie	0	0	1
Pharmacie	1	4	7
Sexologie	0	2	0

Ces quelques remarques mettent ainsi en lumière une certaine variabilité dans la manière de définir l'endométriose dans les discours des spécialistes, chaque type de contexte étant susceptible de mettre en avant différents aspects de la maladie ou de refléter un point de vue particulier. Dans la section suivante, nous adoptons une perspective plus qualitative et présentons les différences les plus saillantes qui émergent de l'observation fine des contextes.

4.2 Une diversité de points de vue

Les définitions symptomatiques ne variant que très peu dans le corpus *EndoExp*, notre propos ne portera ici que sur les contextes correspondant aux définitions de type « autre » et aux définitions histologiques².

4.2.1 Contextes définitoires de type « autre »

Les contextes définitoires correspondant au type « autre » sont particulièrement hétérogènes et de nature plus descriptive que les deux autres types. Les aspects les plus fréquemment mentionnés concernent :

- le fonctionnement de la maladie, avec les adjectifs *hormono-dépendant* et *œstrogéno-dépendant* qui modifient l'hyperonyme d'*endométriose* (*maladie, pathologie, affection*) ;
- le caractère chronique de la maladie ;
- la complexité, l'hétérogénéité et la multifactorialité de l'endométriose.

² Voir Humbert-Droz (2024) pour une analyse des contextes définitoires de type symptomatique dans le corpus *EndoExp*, en comparaison avec un corpus de presse.

Ces aspects sont tout à fait consensuels et aucune différence particulière selon la discipline ne peut être observée dans le corpus. D'autres aspects varient cependant de manière plus significative. C'est le cas des contextes qui explicitent les conséquences possibles de l'endométriose. Ces contextes mentionnent :

- l'impact général de la maladie, avec les adjectifs *invalidant* et *handicapant* ;
- l'infertilité ;
- l'altération de la qualité de vie des femmes en général ou avec des précisions sur les aspects impactés (vie sociale, personnelle, professionnelle ou encore sexuelle) ;
- les conséquences sur la société, notamment le coût socio-économique de l'endométriose.

Ces conséquences s'observent dans des textes de gynécologie, d'obstétrique et de médecine générale, qui sont, comme nous le disons *supra*, les disciplines qui ont retourné le plus de contextes pertinents. Ce constat n'est donc pas surprenant. Par ailleurs, à ce stade, nous pensons que cette diversité permet surtout de saisir l'ampleur de l'impact de cette maladie, aussi bien au niveau individuel qu'au niveau collectif.

Un autre aspect qui varie de manière plus importante concerne la fréquence de la maladie dans la population, qui est mentionnée dans 31 contextes. Dans 20 d'entre eux, on observe simplement l'adjectif *fréquent* et, dans 17, un pourcentage approximatif est indiqué (les deux types d'informations apparaissent parfois simultanément, le second précisant le premier). On observe le plus de variation dans ces derniers, avec des intervalles allant de 2-6 % jusqu'à 10-15 % de femmes atteintes d'endométriose, en passant par 5-10 %, 5-15 % et 6-10 %, dans des contextes tout à fait similaires par ailleurs, comme le montrent les exemples suivants (la discipline associée à chaque exemple est indiquée entre parenthèses).

- a. L'endométriose est une maladie bénigne qui représente un problème de santé publique touchant **entre 10 et 15 %** des femmes en âge de procréer. (Gynécologie)
- b. L'endométriose est une maladie qui touche **entre 6 et 10 %** des femmes en âge de procréer. (Gynécologie)

À ces divergences s'ajoute un point de vue différent, dans un contexte qui explicite que la fréquence de l'endométriose n'est en réalité pas connue. Toutes les autres indications sont, de fait, des estimations.

- c. L'endométriose est une maladie féminine très fréquente [...] **même si la prévalence exacte n'est pas connue**. (Gynécologie)

Un dernier point de divergence concerne le caractère évolutif ou non de l'endométriose. Bien que seuls cinq contextes du type « autre » précisent cet aspect, leur observation révèle l'existence de deux points de vue distincts, voire contradictoires.

- d. L'endométriose étant une maladie **évolutive**. (Médecine générale)
- e. L'endométriose est une maladie **stable dans le temps et non progressive**. (Gynécologie)

Une légère tendance semble se dégager, même si les contextes sont peu nombreux. En effet, un second contexte associé à la gynécologie explicite le fait que l'endométriose n'est pas une maladie évolutive. Par ailleurs, le « caractère récidivant » de la maladie, qui sous-tend une certaine évolutivité, est exprimé dans un contexte associé à la discipline

« imagerie ». À ce stade, il semble que cet aspect puisse effectivement varier selon la discipline dans laquelle il est appréhendé.

Si nous nous plaçons dans une perspective de circulation des termes et des connaissances, ces deux derniers points de divergence sont susceptibles de se diffuser également vers le grand public, et plusieurs points de vue, caractéristiques d'une discipline en particulier ou non, peuvent coexister dans des textes non spécialisés. Cette coexistence est susceptible de créer certaines ambiguïtés pour les non-spécialistes et dans les représentations de la maladie qui sont véhiculées notamment par les médias.

4.2.2 Contextes définitoires de type histologique

Bien que les contextes correspondant au type de définition histologique soient les moins nombreux dans le corpus, ils présentent le plus de variabilité, notamment dans la manière de décrire ce qui compose l'endométriose (c'est-à-dire le « tissu semblable à la muqueuse utérine », repris de la définition de l'association EndoFrance, mentionnée en 2.1). On dénombre 15 descriptions différentes :

- *tissu endométrial, tissu endométrial composé de glandes et de stroma, tissu endométrial fonctionnel ;*
- *glandes endométriales et stroma, glandes et stroma endométrial, glandes et stroma endométriaux ;*
- *tissu de type endométrial, lésions ressemblant à du tissu endométrial ;*
- *tissu endométriosique ;*
- *cellules ectopiques, tissu endométrial ectopique ;*
- *cellules endométriales ;*
- *tissu de l'endomètre, cellules de l'endomètre, endomètre.*

Dans la plupart des cas, ces variations n'ont que peu d'impact au niveau conceptuel. Par exemple, *tissu endométrial ectopique* ajoute une précision quant à la localisation anatomique, en comparaison avec *tissu endométrial*, mais le concept reste fondamentalement le même. Par ailleurs, lorsque *tissu endométrial* est utilisé seul, la localisation anatomique est tout de même exprimée, par d'autres syntagmes, tels que *localisation ectopique* et *présence en dehors de l'utérus*.

En revanche, deux points de vue distincts semblent se dégager. D'abord, avec *tissu de type endométrial* et *lésions ressemblant à du tissu endométrial*, l'accent est placé plutôt sur une idée de ressemblance entre ce qui compose l'endométriose et le tissu endométrial « normal » (c'est-à-dire l'endomètre). Une nuance est donc ajoutée, en comparaison avec *tissu endométrial* notamment. Ensuite, trois descriptions associent clairement ce qui compose l'endométriose avec l'endomètre : *tissu de l'endomètre, cellules de l'endomètre, endomètre*. Étant donné qu'aucune notion de ressemblance n'est attestée, ce point de vue se distingue fortement du précédent, voire est carrément contradictoire. Si, dans le premier cas, les descriptions tendent à distinguer clairement ce qui compose l'endométriose de l'endomètre, dans le second cas, en revanche, ces deux éléments sont tout à fait associés, sans qu'aucune nuance ou modulation n'apparaisse dans le contexte plus large.

Il n'est pas sûr à ce stade que ces manières de décrire cet aspect de l'endométriose soient réellement contradictoires ou incompatibles, mais elles révèlent tout de même certaines nuances dans la manière de concevoir l'endométriose chez les spécialistes, voire reflètent

une certaine ambiguïté. Cette ambiguïté semble par ailleurs renforcée par le fait que les différents points de vue sont attestés dans les textes des mêmes disciplines, par exemple en gynécologie et en médecine générale. Autrement dit, aucune corrélation entre un point de vue particulier et une discipline ne peut être observée et les différents points de vue semblent coexister au sein des différentes disciplines, du moins d'après nos données.

Enfin, si, comme l'affirment Estopà et Montané, l'ambiguïté est l'un des facteurs pouvant affecter négativement la compréhension d'un contenu spécialisé par les patients (Estopà, Montané, 2020 : 216) et, par extension, par les non-spécialistes, alors les ambiguïtés repérées dans le corpus *EndoExp* (composition des lésions d'endométriose, évolutivité et fréquence de la maladie) sont susceptibles de se refléter dans d'autres textes, de genres et de degrés de spécialisation différents, qui participent à la diffusion des connaissances sur cette maladie.

5. Conclusion et travaux ultérieurs

Dans cet article, nous avons présenté plusieurs défis liés à l'interdisciplinarité pour l'étude de la terminologie de l'endométriose en corpus. Une proposition de corpus composé de textes provenant des multiples disciplines médicales qui s'intéressent à cette maladie a été faite en ce sens. La pertinence de ce corpus a été démontrée à travers notre étude des contextes définatoires du terme *endométriose* en discours. D'un point de vue quantitatif, nous avons montré que les aspects de la maladie sur lesquels les spécialistes se concentrent varient selon la discipline, ce qui a été observé à travers le type de définition privilégié (histologique, symptomatique, autre) par discipline (cf. Tableau n. 4). Notons tout de même que, les données étant peu nombreuses pour certaines disciplines (p. ex. médecine nucléaire, sexologie), ces résultats sont à nuancer et devront être confirmés par des études ultérieures. D'un point de vue qualitatif, notre analyse a mis en lumière certaines différences dans les contextes définatoires d'*endométriose*, en ce qui concerne la composition des lésions, la fréquence et l'évolutivité de la maladie. À l'exception de ce dernier aspect, cependant, les données ne permettent pas d'observer de corrélation entre un point de vue particulier et une discipline ; d'autres facteurs que l'interdisciplinarité sont probablement à la source de ces variations. D'autres pistes devront donc être explorées, par exemple en adoptant une perspective diachronique (Dury, 2013 ; Zanola, 2014) afin de retracer l'évolution du terme *endométriose* dans le temps et d'identifier les éléments permettant de contextualiser et d'expliquer les variations constatées ici en synchronie.

Dans le même temps, bien qu'elle ait permis un premier regard sur la variation définitionnelle du terme *endométriose* chez les spécialistes, l'approche proposée dans cet article doit être complétée. Une phase ultérieure de ce projet pourra ainsi inclure les contextes définatoires qui dépassent les limites de la phrase et qui comprennent des reprises anaphoriques (par exemple avec un hyperonyme, comme *maladie* ou *pathologie*) ou encore des contextes définatoires qui portent sur un méronyme d'*endométriose* (par exemple *lésion*, *foyer*). D'autres patrons pourront également être analysés pour compléter ces descriptions, notamment lorsque les définitions sont introduites par d'autres verbes.

Enfin, cette première étude constitue la base d'un projet plus vaste, dont la poursuite bénéficiera de ces premiers résultats. Une attention particulière sera notamment portée sur les ambiguïtés relevées dans les descriptions de ce qui compose l'endométriose, dans les indications de fréquence et dans le caractère évolutif de l'endométriose. Nous chercherons ainsi à observer dans quelle mesure ces ambiguïtés se diffusent auprès du grand public et selon quelles modalités, dans l'objectif global de mieux comprendre leur impact, entre autres, sur les représentations de l'endométriose véhiculées dans les discours non spécialisés.

Références bibliographiques

- Bordet G., *Brouillage des frontières, rencontre des domaines : quelles conséquences pour l'enseignement de la terminologie et de la traduction spécialisée*, in « ASp », 64, 2013, pp. 95-115.
- Bowker L., Pearson J., *Working with Specialized Language. A Practical Guide to Using Corpora*, Londres/New York, Routledge, 2002.
- Condamines A., Dehaut N., *Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie*, in « Meta : journal des traducteurs », 56(2), 2011, pp. 266-283.
- Condamines A., Picton A., *Textual Terminology: Origins, Principles and New Challenges*, in P. Faber, M.-C. L'Homme (eds.), *Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized knowledge*, Amsterdam/Philadelphia, John Benjamins, 2022, pp. 219-236.
- Delavigne V., *La notion de domaine en question. À propos de l'environnement*, in « Neologica », 16, 2022, pp. 27-59.
- Dury P., *Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine médical*, in « Debate terminológico », 9, 2013, pp. 2-10.
- EndoFrance, *Endométriose*, 2025. URL : <<https://www.endofrance.org/>>, consulté le 19-11-2025.
- Estopà R., Montané A., *Terminology in medical reports: textual parameters and their lexical indicators that hinder patient understanding*, in « Terminology », 26(2), 2020, pp. 213-236.
- Galès M.-R., *Endométriose : ce que les autres pays ont à nous apprendre*, Paris, Éditions Josette Lyon, 2020.
- Grimaldi C., *L'univers de la blockchain : étude d'une terminologie émergente aux contours nébuleux à partir d'un corpus secondaire*, in « Studia Universitatis Babeş-Bolyai Philologia », 69(1), 2024, pp. 127-141.
- HAS, CNGOF. *Prise en charge de l'endométriose. Recommandations de bonne pratique*, 2017.
- Humbert-Droz J., *Terminologie de l'endométriose et représentations de la maladie : regards croisés entre presse généraliste et discours spécialisés*, in 9^e Congrès Mondial de Linguistique Française, 2024. URL : <https://www.shs-conferences.org/articles/shsconf/abs/2024/11/shsconf_cmlf2024_05003/shsconf_cmlf2024_05003.html>, consulté le 19-11-2025.
- Ilchner S. et al., *Communicating Endometriosis Pain in France and Australia: An Interview Study*, in « Frontiers in Global Women's Health », 3, 2022. URL : <<https://pubmed.ncbi.nlm.nih.gov/35400132/>>, consulté le 19-11-2025.
- Kilgariff A. et al., *The Sketch Engine*, in *Proceedings of the 11th EURALEX International Congress*, 2004, pp. 105-116.
- L'Homme M.-C., *La terminologie : principes et techniques*, Montréal, Les Presses de l'Université de Montréal, 2020.

- Lefeuve L., *Analyse des marqueurs de relations conceptuelles en corpus spécialisé : Recensement, évaluation et caractérisation en fonction du domaine et du genre textuel*, Thèse soutenue à l'Université Toulouse – Jean Jaurès, 2017.
- Meyer I., *Extracting Knowledge-Rich Contexts for Terminography: A conceptual and methodological framework*, in D. Bourigault, C. Jacquemin, M.-C. L'Homme (eds.), *Recent Advances in Computational Terminology*, Amsterdam, John Benjamins, 2001, pp. 279-302.
- Meyer I., Mackintosh K., *The Corpus from a Terminographer's Viewpoint*, in « International Journal of Corpus Linguistics », 1(2), 1996, pp. 257-285.
- Péraldi S., *Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique. Analyse à partir d'un corpus spécialisé de langue anglaise*, Thèse soutenue à l'Université Paris Diderot, 2011.
- Picton A., Dury P., *Diastratic Variation in Language for Specific Purposes. Observations from the Analysis of two Corpora*, in P. Drouin, A. Francoeur, J. Humbley, A. Picton (eds.), *Multiple Perspectives in Terminological Variation*, Amsterdam/Philadelphia, John Benjamins, 2017, pp. 57-79.
- Rebeyrolle J., *Forme et fonction de la définition en discours*, Thèse soutenue à l'Université Toulouse 2, 2000.
- Sierra G. et al., *Exploration de contextes définitoires en français : Cas d'étude sur la COVID-19*, in M. Misuraca, G. Scepi, M. Spano (eds.), *Proceedings of the 16th International Conference on Statistical Analysis of Textual Data*, Cosenza, Edizioni Erranti, 2022, pp. 817-826.
- Young K. et al., *Women's experiences of endometriosis: A systematic review and synthesis of qualitative research*, in « Journal of Family Planning and Reproductive Health Care », 41(3), 2015, pp. 225-234.
- Zanola M.T., *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*, Paris, L'Harmattan, 2014.

La terminologie de la télémédecine : enjeux et perspectives d'un domaine interdisciplinaire

CAROLINA IAZZETTA

1. Introduction

L'essor des technologies de l'information et de la communication (TIC) à partir des années 1990 a engendré la prolifération d'une pléthore d'unités lexicales, parfois employées de manière interchangeable, relevant de domaines interdisciplinaires et pluridisciplinaires tels que l'informatique, la technologie et le marketing dans le secteur médical (*télémédecine*, *e-santé*, *santé numérique*, *santé connectée*, *santé digitale*, *cybersanté*, pour n'en citer que quelques-unes).

En raison de son interdisciplinarité, de ses nombreuses facettes et de son évolution rapide, la télémédecine¹ est un domaine difficile à saisir qui connaît un grand flou terminologique et conceptuel également signalé par le Conseil national de l'Ordre des médecins. À ce propos, l'Organisation Mondiale de la Santé (OMS) et d'autres organismes institutionnels ont du mal à donner des définitions univoques de ces unités lexicales, souvent utilisées comme des synonymes, même si elles ne le sont pas d'un point de vue sémantique. D'ailleurs, malgré l'essor des TIC dans tous les domaines de la vie humaine (entre autres, l'éducation, la culture, la santé et le travail), la notion même de TIC reste encore vague en raison de son application à des réalités différentes et à l'invention d'outils de plus en plus sophistiqués.

Dans cette contribution, après avoir présenté les procédés principaux de formation de néologismes appartenant au domaine de la télémédecine et avoir sélectionné un échantillon de termes tirés des guides de télémédecine disponibles sur le site du Ministère du travail, de la santé et des solidarités, nous reviendrons sur les définitions de ces termes dans les textes de l'OMS et quelques ressources lexicographiques et terminographiques en ligne (entre autres, le *Petit guide d'exploration de la santé numérique*, le *Glossaire de l'innovation en santé*, le *Grand dictionnaire terminologique-GDT*, *TERMIUM Plus*). Notre objectif est de mettre en évidence les frontières très floues et la confusion tant conceptuelle que terminologique caractérisant ce domaine de connaissances fortement influencé par l'anglais. De cette manière, nous envisageons de proposer une cartographie sémantique précise des termes les plus courants de la télémédecine, un secteur qui ne

¹ Dans cette contribution, nous avons fait le choix d'utiliser le mot *télémédecine* en raison de son plus grand nombre d'occurrences dans les discours de vulgarisation scientifique en ligne.

cesse d'évoluer et s'élargir et qui n'a pas encore fait l'objet d'études systématiques d'un point de vue linguistique.

2. Les préfixes à la base des créations néologiques en télémedecine

Les préfixes représentent une porte d'accès privilégiée à la connaissance d'un domaine spécialisé et, par conséquent, leur relation avec la terminologie est essentielle. En fait, la préfixation est l'une des typologies de « motivation » la plus employée à la base de la néologie terminologique (Kocourek, 1991 : 175 ; Humbley, 2018). Au fil des années, plusieurs études abordant les préfixes parmi les autres formes de création lexicale et terminologique ont été menées dans le domaine médical et dans d'autres disciplines scientifiques et techniques (entre autres, Raus, 2001 ; Bonadonna, 2012 ; Dury, 2021). Autrement dit, les préfixes contribuent de manière considérable et pertinente à la création de nouveaux termes pour communiquer la « multiplication des techniques, le rythme accéléré des innovations et des découvertes » (Dubuc, 2002 : 1) qui caractérisent les sciences.

À une époque de plus en plus connectée, la création de néologismes à partir de préfixes et de suffixes d'origine grecque ou latine² est encore fréquente dans le domaine médical, mais la tendance dominante est celle de créer des termes nouveaux à partir de préfixes dérivés des langues vivantes, et notamment de l'anglais. En particulier, parmi les principaux préfixes à l'origine de la formation de néologismes en télémedecine, *télé-*, *cyber-* et *e-* sont les plus productifs. Ils désignent des activités basées sur l'utilisation des réseaux informatiques et de télécommunications, ce qui représente une preuve évidente de l'interdisciplinarité de notre domaine d'étude.

Le formant *télé-*, provenant du grec *têle* signifiant « loin, au loin », a été juxtaposé au terme *medecine*, où sa position initiale indique le moyen par lequel est réalisée l'activité désignée, à savoir un service médical dispensé à distance. Son expansion va de pair avec la multiplication ou le perfectionnement des techniques et des outils de transmission à distance. Ce formant apparaît ainsi dans de nombreuses appellations du domaine de la télémedecine, soit pour indiquer des spécialités prises en charge par cette branche (entre autres, *téléradiologie*, *télécardiologie*, *télédermatologie*), soit pour renvoyer à une pathologie ou un traitement qui bénéficie de cette nouvelle prise en charge (par exemple, *télé dialyse*). Par ailleurs, *télé-* a montré toute sa vitalité pendant la crise sanitaire de COVID-19 dans la désignation de nouvelles notions liées aux activités professionnelles à distance (entre autres, *télétravail*, *téléécole*, *téléenseignement*, cf. Grimaldi, 2022).

Pour ce qui est du confixe *cyber-*, le développement d'internet et des TIC dans tous les domaines de la vie quotidienne a contribué à sa diffusion. Il dérive du grec

² Du côté des préfixes, citons, entre autres, *a-* signifiant l'absence ou la privation (*amnésie*, *anémie*, *ablation*), *poly-* et *oligo-* indiquant la quantité (*polydactylie*, *oligurie*), *tachy-* et *brady-* marquant la fréquence (*tachycardie*, *bradycardie*) et *iso-* renvoyant à l'égalité ou à la similitude (*isotherme*). Du côté des suffixes, citons, entre autres, *-algie* désignant la douleur (*névralgie*), *-ose* se référant à une dégénérescence (*arthrose*), *-rrhée* s'associant à un écoulement (*rhinorrhée*), *-tomie* concernant l'incision (*laparotomie*), *-pénie* marquant la diminution (*cytopénie*) et *-logie* se référant aux disciplines scientifiques (*histologie*, *cardiologie*).

κυβερνήτης (*kybernetes*), littéralement « timonier, pilote d'un navire », et par extension « celui qui guide et gouverne une ville ou un État » (cf. le vocabulaire de l'Accademia della Crusca). C'est le mathématicien et ingénieur britannique James Watt qui, à la fin du XVIII^e siècle, a employé pour la première fois le mot *cybernetic* dans un contexte purement technique pour désigner le régulateur centrifuge automatique des machines à vapeur. Le suffixe anglais *-ic* a été ensuite rendu en français par l'équivalent *-ique* (Raus, 2001 : 74), d'où le substantif *cybernétique*. Or, le sens actuel du mot tire son origine du terme anglais *cyberspace*, utilisé en 1984 par William Gibson, auteur américain de science-fiction, dans son livre *Neuromancer* pour désigner un « espace de communication créé par l'interconnexion mondiale des ordinateurs et par les données qui y sont traitées ; espace, milieu dans lequel naviguent les internautes » (dictionnaire *Le Robert*). Ce confixe s'avère être très productif dans la formation de mots nouveaux désignant aussi bien des lieux (*cybercafé*), que des objets (*cyberradio*), des personnes (*cyberartiste*), des activités (*cybertéléphoner*) ou l'espace virtuel des réseaux (*cyberie*). Par ailleurs, *cyber-* est de plus en plus utilisé aujourd'hui pour former des néonymes non seulement dans le secteur médical (par exemple, *cybermédecine*, *cyberbistouri*, *cybermalaise*, etc.), mais aussi dans les domaines de l'informatique (par exemple, *cyberassistance*, *cyberlogistique*, *cyberéthique*), de l'économie (par exemple, *cybermarketing*, *cyberbanque*, *cyberéconomie*), du droit (par exemple, *cybercriminalité*, *cyberdélinquance*), de la psychologie (par exemple, *cybermanie*, *cyberphobie*, *cyberpsychologie*) et de la sociologie (par exemple, *cyberpropagande*, *cyberinclusion* et *cybermilitantisme*). Il est ainsi évident que ce confixe a subi un glissement sémantique par rapport à sa valeur d'origine. D'après l'Office québécois de la langue française, sa productivité ainsi que la tendance aux abréviations a fait que le monème *cyber* est devenu libre et s'utilise soit comme un déterminant lexical de type épithète, par exemple dans des syntagmes tels que *culture cyber*, *ère cyber*, *génération cyber*, où il figure en apposition, soit tout seul pour désigner, selon les cas, le *cyberespace*, la *cyberculture* ou l'ensemble des nouvelles technologies (GDT). Par conséquent, des questions légitimes se posent : *cyber-* est-il vraiment un préfixe ? Est-il utilisé dans la formation de mots dérivés ou de mots composés ?

Finalement, bien que moins utilisé par rapport aux préfixes *télé-* et *cyber-*, le préfixe *e-* (par exemple, en *e-santé* et *e-médecine*) est également récurrent. Au-delà d'une simple mode qui affecte le monde médiatique, économique et administratif (*e-commerce*, *e-business*, *e-book* et *e-ministère*), ce préfixe tend à se répandre de plus en plus dans l'usage français et notamment en télémédecine. Il est emprunté à l'anglais en tant que forme raccourcie de l'adjectif *electronic*, au sens de « électronique » en langue française. Or, l'Office québécois de la langue française n'en recommande pas son usage et précise que le préfixe *e-* se révèle superflu et n'a pas de véritable raison d'être dans la mesure où plusieurs équivalents sont disponibles en français pour le remplacer. D'après la Commission générale de terminologie et de néologie (CGNT, 2005) :

E- est un néologisme hybride entre lettre, mot et concept, porteur de difficultés de tous ordres. S'il est aisément employé en anglais, notamment pour des raisons phonétiques (voyelle longue et accentuée), il est difficilement identifiable en fran-

çais. De plus la signification en est confuse et fluctuante, puisqu'il s'emploie pour désigner indifféremment tout ce qui est lié aux techniques de l'information et de la communication : technique, procédure, missions ou organismes.

La Commission générale de terminologie et de néologie déconseille donc l'emploi du préfixe *e-* sous toutes ses graphies (*e-*, *é-* ou *i-*) et suggère d'utiliser le préfixe *cyber-*, « dans les cas où la réalité à désigner a un caractère concret » (CGTN, 2005), *télé-*, « avec des notions relevant strictement du domaine de la télévision ou des activités à distance » (CGTN, 2005), et propose de privilégier, le cas échéant, les formules *en ligne* où à *distance*.

3. Les mots-clés de la télémédecine : entre flou terminologique et conceptuel

Dans un souci de mettre en évidence les frontières très floues et la confusion tant conceptuelle que terminologique caractérisant ce domaine de connaissances, nous avons sélectionné quelques termes clés du domaine de la télémédecine à partir d'un dépouillement manuel de guides de télémédecine disponibles sur le site du Ministère du travail, de la santé et des solidarités³. En particulier, nous nous sommes penchés sur les concepts de *télesanté*, *télémédecine* et *e-santé*, dont les dénominations sont souvent utilisées dans les mêmes contextes discursifs et textuels, ce qui pourrait se prêter à ambiguïté pour un public de non-experts du domaine médical, comme l'atteste cette capture d'écran du *Guide méthodologique pour l'élaboration du programme régional de télémédecine* (2011).

Fig. 1 - Capture d'écran du Guide méthodologique pour l'élaboration du programme régional de télémédecine (2011)

3.3.1.2 La nouvelle organisation et son impact sur les pratiques professionnelles

Gouvernance du projet de télémédecine

Dans le cadre du GCS Télésanté Midi-Pyrénées (dont les Hôpitaux de Lannemezan sont membres), le projet « Télémédecine en UCSA CH de Lannemezan » est piloté par une UF de télémédecine, avec un responsable médical et un responsable technique, dans le cadre d'un GCS « e-santé » régional dont les Hôpitaux de Lannemezan sont membres.

- Le GCS Télésanté Midi-Pyrénées met à disposition des acteurs du projet les ressources nécessaires :
 - le matériel informatique ;
 - une ligne de connexion entièrement sécurisée ;
 - une plateforme informatique régionale permettant d'accéder au dossier patient informatisé sur site ;
 - des adaptations de débit en fonction des besoins ;
 - une expertise concernant l'entretien et l'utilisation du matériel.
- L'UF de télémédecine gère l'organisation locale du projet.

Le GCS Télésanté Midi-Pyrénées et l'UF de télémédecine se réunissent officiellement au moins une fois par année. Les membres de l'UF de télémédecine se réunissent plus régulièrement de façon informelle.

Les autres acteurs concernés par le projet sont :

- les médecins requis
- les médecins demandeurs
- les infirmières.

³ URL : <<https://sante.gouv.fr/>> (consulté le 19-11-2025).

Il nous semble que l'ambiguïté potentielle existant entre les concepts de *télésanté* (ou *e-santé*) et de *télémédecine* remonte à la parution en 2008 et 2009 de deux rapports ministériels « La place de la télémédecine dans l'organisation des soins » (Simon, Acker, 2008) et « Un plan quinquennal éco-responsable pour le déploiement de la télésanté en France » (Lasbordes, 2009), le premier portant sur la *télémédecine* clinique, à distinguer de la *télésanté*, et le second traitant de la *télésanté* sous l'angle du plan organisationnel, et qui y intègre aussi la *télémédecine clinique*. Il n'est pas étonnant que ces nouveaux paradigmes de soins commençant par le préfixe *télé-* soient confondus, surtout lorsque le Ministère américain de la santé et des services sociaux (U.S. Department of Health and Human Services) associe la *télésanté* à la *télémédecine*. En outre, force est de constater que *télésanté* est un calque de l'anglais *telehealth*, utilisé dans les pays anglo-saxons pour décrire surtout les services de la télémédecine informative, alors qu'en France la *télésanté* intègre plutôt tous les domaines de la santé numérique, ce qui peut prêter à confusion et renforcer davantage l'ambiguïté terminologique et conceptuelle entre les notions de *télémédecine* et *télésanté*. D'après la Federal Communications Commission, la notion de *telehealth* est définie comme :

[...] Telehealth is similar to telemedicine but includes a wider variety of *remote healthcare services beyond the doctor-patient relationship*. It often involves services provided by nurses, pharmacists or social workers, for example, who help with patient health education, social support and medication adherence, and troubleshooting health issues for patients and their caregivers. (Federal Communications Commission)

Parmi les nombreuses définitions de *télémédecine* disponibles dans différentes ressources lexicographiques, nous avons choisi de prendre en considération celle proposée par le Glossaire de l'innovation en santé, qui nous semble la plus claire, exhaustive et pertinente :

[...] forme de pratique médicale à distance mettant en rapport, entre eux ou avec un patient, un ou plusieurs professionnels de santé, parmi lesquels figure nécessairement un professionnel médical et, le cas échéant, d'autres professionnels apportant leurs soins au patient. (*Glossaire de l'innovation en santé*)

Pour ce qui est de la notion de *télésanté*, ses définitions et ses frontières sont mal définies et parfois contradictoires aussi bien dans la communication spécialisée que dans la vulgarisation scientifique. Labordes, dans son rapport sur la télésanté de 2009, propose pourtant des définitions permettant de bien distinguer les concepts de *télésanté* et de *télémédecine*. D'après lui, la *télésanté* est une notion nouvelle conçue comme « l'utilisation des outils de production, de transmission, de gestion et de partage d'informations numérisées au bénéfice des pratiques tant médicales que médico-sociales » (Lasbordes, 2009 : 37), alors que pour le concept de *télémédecine*, il s'appuie sur la définition proposée par le Code de la santé publique (par la Loi n° 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires) qui la considère

comme « une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication ». À partir de ces définitions, il est ainsi évident que la télémédecine constitue la partie médicale de la télésanté.

Pour d'autres acteurs, notamment les praticiens appartenant à la Société française de télémédecine (SFT-Antel), le terme *télésanté* a une connotation plutôt commerciale, c'est-à-dire qu'elle « englobe toutes les applications que l'on trouve sur internet et qui sont, tout ou partie, liées à la santé » (Simon, Acker, 2008). Cette seconde acception est conforme à celle déjà suggérée en 1998 par l'OMS selon laquelle la télésanté correspond à « la compréhension d'un moyen d'intégration des systèmes de télécommunication pour protéger et faire avancer la santé ». Il est clair que le concept de *télésanté* est hybride, étant donné qu'il comporte deux dimensions complémentaires : d'une part, il s'agit d'un service, c'est-à-dire d'une façon de garantir des prestations de santé à distance, mais, d'autre part, il s'agit aussi d'un outil, à savoir d'une application technologique au domaine de la santé (Commission de l'éthique en science et technologie, 2014 : 17).

En novembre 1999, un informaticien australien a contribué à renforcer cette confusion terminologique et conceptuelle en introduisant la locution anglaise *e-Health* (traduit en français par *e-santé* ou *santé électronique*), par analogie avec *e-commerce*, recouvrant tous les types de pratiques à distance portant sur les soins et la santé, avec une démarche essentiellement technico-commerciale à destination des professionnels comme des patients. Autrement dit, la *e-santé* représente un domaine émergent à la croisée de l'informatique médicale, de la santé publique et du monde du commerce et de l'entreprise. La locution s'est ensuite banalisée pour qualifier tout ce qui contribue à la transformation numérique du système de santé voire, au-delà du seul secteur santé, le médico-social.

L'OMS définit la *e-santé* comme « l'utilisation des technologies de l'information et de la communication pour la santé ». Cependant, il est intéressant de noter que dans les documents institutionnels, la locution anglaise *e-health* est généralement traduite en français par *cybersanté* et, plus rarement, par *e-santé* :

Par *eHealth* (*cybersanté*) ou services de santé en ligne, on entend l'utilisation des technologies de l'information et de la communication (TIC) pour l'organisation, le soutien et la mise en réseau de tous les processus et personnes impliqués dans le système de santé.

(Confédération Suisse, Office Fédéral de la Santé Publique, 2007)

Partant de ce constat l'OMS et l'UIT se sont engagées à vulgariser et rendre accessibles les services de la cybersanté (eHealth) en prenant plusieurs résolutions et recommandations. (ITU/BDT Regional Economic and Financial Forum of Telecommunications/ICTs for Arab States, 2017)

Selon le *Petit guide d'exploration au pays de la santé numérique*, par *e-santé* on désigne « l'application des technologies de l'information et de la communication (TIC) à l'ensemble des activités en rapport avec la santé » (Béjean, Dumond, Habib, 2015 : 6), alors que d'après le *Glossaire de l'innovation en santé*, cette notion correspond plutôt à :

[l']utilisation des outils de production, de transmission, de gestion et de partage d'informations numérisées (services numériques) au bénéfice des pratiques tant médicales que médico-sociales. Soit tout ce qui contribue à la transformation numérique du système de santé et du secteur médico-social.

Les différentes tentatives de définition de *e-santé* montrent qu'il existe un décalage entre la *e-santé* en tant que produit, mis en avant par les acteurs économiques, et la *e-santé* en tant que service, promue par les autorités publiques. La *e-santé* est, en effet, à la fois un marché et une façon de concevoir les politiques publiques pour les autorités sanitaires. Cette locution est, en outre, utilisée de manière interchangeable avec celle de *santé numérique* par le Gouvernement français aussi bien que dans le portail <esante.gouv.fr>.

Face à ce flou terminologique et conceptuel, l'OMS précise dans son rapport de 2010 que la télémédecine et la *e-santé* sont des sous-ensembles de la télé-santé et que les dénominations de ces concepts ne sont pas des synonymes. La *e-santé* renvoie, en fait, à l'ensemble du contenu numérique en libre accès lié à la santé, facilitant sa diffusion que ce soit dans la sphère médicale ou domestique, alors que la télémédecine renvoie à la prise en charge clinique.

Pour ce qui est de la télé-santé, elle se définit plutôt par son rôle dans le domaine de la santé publique (éducation à la santé, développement de la santé communautaire et des systèmes de santé, épidémiologie, cf. Simon, 2015). Autrement dit, chaque terme revêt un caractère singulier. Par ailleurs, la distinction entre ces concepts est importante, étant donné qu'ils ne sont pas soumis aux mêmes réglementations juridiques. Sur la même ligne, le GDT précise :

La télémédecine fait partie du grand ensemble appelé *télé-santé*. En fait, la télémédecine a constitué les débuts de la télé-santé. Cependant, aujourd'hui, elle s'en distingue et est maintenant considérée comme une application de celle-ci.

Dans ce sens, le mot *télé-santé* peut être considéré comme hyperonyme de *télémédecine*. Or, le terme qui a le plus d'occurrences en français aussi bien dans les ressources lexicographiques que dans la vulgarisation scientifique en ligne est *télé-santé*, comme le prouve le GDT (Fig n. 2), alors que les attestations de *e-santé* sont généralement limitées aux domaines de l'ingénierie et de l'informatique. En fait, *e-santé* ne figure même pas dans la Banque de données terminologiques et linguistiques du gouvernement du Canada (*TERMIUM Plus*®). C'est le GDT qui propose plutôt trois fiches du terme *télé-santé*, selon les différents domaines d'application (hygiène et santé, internet et télématique, titre des programmes fédéraux).

lisés et confondus en langue italienne aussi bien dans les documents institutionnels que dans la vulgarisation en ligne sont *e-health*, *digital health*, *digital medicine*, *telemedicina*, *teleassistenza* et *mobile health* (Iazzetta, 2024). Il est ainsi évident que le lexique italien de la télémédecine, plus que le français, tire une bonne partie de son vocabulaire de l'anglais.

4. L'anglicisation de la télémédecine : la querelle *digital*/numérique

Avec la globalisation du système sanitaire, la création d'outils de diagnostic très sophistiqués et précis, une robotisation croissante et un accès à l'information médicale simple et rapide, la langue de la médecine est de plus en plus influencée par l'anglais. Cette anglicisation est liée aussi bien à l'hégémonie scientifique des pays de l'Amérique du Nord qu'aux propriétés de la langue anglaise possédant la flexibilité linguistique nécessaire (par exemple, troncation et concision) pour faciliter la création de nouvelles lexies (Faure, 2010).

Dans le cas spécifique de la télémédecine, les anglicismes circulent à la fois dans la communication spécialisée et dans la vulgarisation scientifique. Le mot *digital*, par exemple, qui est très utilisé, fait référence en français, à tout ce qui est lié aux doigts. D'après l'Académie française :

L'adjectif digital en français signifie « qui appartient aux doigts, se rapporte aux doigts ». Il vient du latin *digitalis*, « qui a l'épaisseur d'un doigt », lui-même dérivé de *digitus*, « doigt ». « C'est parce que l'on comptait sur ses doigts que de ce nom latin a aussi été tiré, en anglais, *digit*, « chiffre », et digital, « qui utilise des nombres ». On se gardera bien de confondre ces deux adjectifs qui appartiennent à des langues différentes et dont les sens ne se recouvrent pas : on se souviendra que le français a à sa disposition l'adjectif *numérique*.

Digital est désormais utilisé à tort pour désigner tout ce qui est numérique et plusieurs ressources lexicographiques françaises (entre autres, *Le Robert* et *Le Larousse*) considèrent les mots *digital* et *numérique* comme des synonymes, même s'ils ne le sont pas d'un point de vue sémantique. Par ailleurs, *numérique* est un adjectif polysémique, d'où certaines ambivalences qui expliquent cette tendance à utiliser *digital* sur le modèle anglo-saxon.

Les appellations différentes du domaine de la télémédecine suscitent l'intérêt et occupent la Commission d'enrichissement de la langue française qui, de son côté, préconise l'usage de mots français au lieu des nombreux anglicismes (par exemple, *numérique* pour l'anglais *digital*, *sciences des données* pour *data science*, etc.). En particulier, à propos du substantif *numérique*, la Commission le définit comme « [l'] ensemble des disciplines scientifiques et techniques, des activités économiques et des pratiques sociétales fondées sur le traitement de données numériques » (*Journal officiel*, 2021). Compte tenu de cette définition, les locutions *médecine numérique* ou *santé numérique* au lieu de *médecine digitale* ou *santé digitale* s'avèrent les choix les plus indiqués et pertinents.

5. Conclusion

Basée sur l'interdisciplinarité entre technologies, informatique, marketing et sciences humaines, la télémédecine a désormais le vent en poupe et s'impose comme un outil facilitant l'accès aux soins, notamment depuis la crise sanitaire de la COVID-19. Des études interdisciplinaires et pluridisciplinaires dans un domaine « jeune » et en évolution continue s'avèrent ainsi fondamentales.

D'un point de vue linguistique, notre travail a, en premier lieu, mis en exergue qu'à une époque de plus en plus influencée par les TIC, les préfixes les plus productifs pour la création de néologismes dans le domaine médical sont *cyber-*, *télé-* et *e-*, relevant des nouvelles techniques liant l'automatisme et la communication, et que *cyber-* et *télé-* sont ceux recommandés par les instances de politique linguistique francophones.

En deuxième lieu, cette étude nous a permis de démontrer que les termes *télémédecine*, *télésanté* et *e-santé* non seulement se chevauchent et sont parfois utilisés comme des synonymes, mais aussi qu'ils peuvent désigner dans certains cas une pratique médicale à distance et définissent donc l'acte de soins, alors que dans d'autres cas ils font référence aux outils et aux technologies utilisés pour la mise en place à distance de l'acte médical.

Bien qu'il existe des définitions officielles et malgré la présence d'un nombre important de ressources lexicographiques sur la télémédecine (entre autres, *Le Glossaire terminologique en télésanté du CHU Sainte-Justine*, *l'Abécédaire de la e-santé*, *le Glossaire de l'innovation en santé*), la réflexion sur la littérature que nous avons réalisée et les exemples donnés prouvent la persistance d'une confusion terminologique et conceptuelle entre ces termes qui fait écho à l'incertitude liée à l'évolution continue et à la numérisation croissante caractérisant ce domaine. Par ailleurs, force est de constater que ce changement de paradigme des pratiques médicales impose de repenser l'acte thérapeutique, d'inventer de nouveaux métiers et d'intégrer de nouvelles formations multidisciplinaires adaptées à la révolution numérique.

Dans ce travail, nous souhaitons avoir posé les jalons pour l'émergence d'une « conscience terminologique » en matière de télémédecine, dans un souci de créer un dialogue entre les différentes disciplines concernées et ouvrir des pistes de réflexion pour la construction d'un langage partagé. En fait, ce n'est que grâce à une terminologie de qualité, et donc standardisée, que la communication médicale peut être sans équivoque. Par ailleurs, l'emploi d'une langue commune aux différents prestataires du soin peut réduire de manière considérable les erreurs de diagnostic et permettre de travailler sur des données structurées et harmonisées, et donc plus faciles à interpréter.

Finalement, nous avons également signalé que cette même confusion terminologique et conceptuelle caractérisant la langue française est également observée en italien, ce qui ouvre des pistes de recherche dans les différentes langues, notamment dans les langues du Réseau panlatin de terminologie-REALITER.

Références bibliographiques

- Avis de la Commission de l'éthique en science et technologie, *La télésanté clinique au Québec : un regard éthique*, 2014. URL : <https://www.ledevoir.com/documents/pdf/telesante_avis2014.pdf>, consulté le 19-11-2025.
- Béjean M., Dumond J.-P., Habib J., *Petit guide d'exploration au pays de la santé numérique*, Fondation de l'Avenir, 2015. URL : <https://www.fondationdelavenir.org/wp-content/uploads/2015/11/2015_petitguide_sante_numerique.pdf>, consulté le 19-11-2025.
- Bonadonna M.F., *Pour une histoire de la terminologie française de l'énergie électrique*, in « Synergies Espagne », 5, 2012, pp. 65-76.
- Commission générale de terminologie et de néologie (CGTN), *Recommandation sur les équivalents français du préfixe e-*, in « Journal officiel », 22/07/2005. URL : <https://www.legifrance.gouv.fr/download/pdf?id=NFP-zF-1mggwV16GnVpma8AH6Jhdq_FfXc3_EJXu4WE=>>, consulté le 19-11-2025.
- Confédération Suisse, Office Fédéral de la Santé Publique, *Stratégie Cybersanté (eHealth) Suisse*, 2007. URL : <<https://www.bag.admin.ch/bag/fr/home/strategie-und-politik/nationale-gesundheitsstrategien/strategie-ehealth-schweiz.html>>, consulté le 19-11-2025.
- Conseil national de l'Ordre des Médecins, *La Santé connectée. Enjeux et perspectives, Compte rendu du 3 février 2015*. URL : <https://www.conseil-national.medecin.fr/sites/default/files/externalpackage/debat/54mmy9/debatsanteconnecteecnom2015_0.pdf>, consulté le 19-11-2025.
- Dubuc R., *Manuel pratique de terminologie*, Québec, Linguattech éditeur, 2002.
- Dury P., *L'obsolescence terminologique dans le domaine de la pharmacologie*, in « Linx », 82, 2021. URL : <<https://journals.openedition.org/linx/8024>>, consulté le 19-11-2025.
- Faure P., *Des discours de la médecine multiples et variés à la langue médicale unique et universelle*, in « Asp », 58, 2010, pp. 73-86.
- Grimaldi G., *Les traits de la distance et de l'isolement dans le lexique autour de la pandémie*, in « Re-pères-Dorif », 25, 2022, URL : <<http://www.dorif.it/reperes/claudio-grimaldi-les-traits-de-la-distance-et-de-lisolement-dans-le-lexique-autour-de-la-pandemie/>>, consulté le 19-11-2025.
- Humbley J., *La néologie terminologique*, Limoges, Lambert-Lucas, 2018.
- Iazzetta C., *La terminologia della medicina digitale: il caso "problematico" del termine e-health nei dizionari e nei glossari online*, in C. Grimaldi (a cura di), *Ieri e oggi: la terminologia e le sfide delle Digital Humanities*, Milano, EDUCatt, 2024, pp. 53-71.
- ITU/BDT Regional Economic and Financial Forum of Telecommunications/ICTs for Arab States, *eHealth pour le Renforcement des Systèmes de Santé*, Nouakchott, 2017. URL : <<https://www.itu.int/en/ITU/RegionalPresence/ArabStates/Documents/events/2017/EFF17/Pres/S3-Mr.ZOMBRE.pdf>>, consulté le 19-11-2025.
- Kocourek R., *La langue française de la technique et de la science*, Wiesbaden, Oskar Brandstetter, 1991.
- Lasbordes P., *La télésanté : un nouvel atout au service de notre bien-être. Un plan quinquennal éco-responsable pour le déploiement de la télésanté en France*, Rapport remis à Madame Roselyne Bachelot-Narquin, ministre de la Santé et des Sports, 2009. URL : <<https://www.vie-publique.fr/files/rapport/pdf/094000539.pdf>>, consulté le 19-11-2025.
- Légifrance, *Article L6316-1 – Code de la santé publique*, 28/12/23. URL : <https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000048703124>, consulté le 19-11-2025.

- Ministère du travail, de l'emploi et de la santé, *Guide méthodologique pour l'élaboration du programme régional de télémedecine*, 2011. URL : <https://sante.gouv.fr/IMG/pdf/guide_methhodologique_elaboration_programme_regional_telemedecine.pdf>, consulté le 19-11-2025.
- Raus R., *Productivité de cyber et hyper dans le lexique français d'Internet*, in « La Linguistique », 37-2, 2001, pp. 71-88.
- Simon P., Acker D., *La place de la télémedecine dans l'organisation des soins*, Rapport pour le Ministère de la Santé et des Sports et la Direction de l'Hospitalisation et de l'Organisation des Soins, 2008. URL : <https://sante.gouv.fr/IMG/pdf/Rapport_final_Telemedecine.pdf>, consulté le 19-11-2025.
- Simon P., *Télémedecine. Enjeux et pratiques*, Mont-Saint-Guibert, Le Coudrier Édition, 2015.

Ressources lexicographiques (consulté le 19-11-2025)

- Abécédaire de la e-santé. URL : <<https://esante.gouv.fr/abecedaire>>.
- Accademia della Crusca. URL : <<https://accademiadellacrusca.it/it/>>.
- Commission d'enrichissement de la langue française, *Vocabulaire de l'informatique (liste de termes, expressions et définitions adoptées)*, 2021. URL : <<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043228194>>.
- Dictionnaire de l'Académie française (9^e édition). URL : <<https://www.academie-francaise.fr>>.
- FranceTerme. URL : <<https://www.culture.fr/franceterme>>.
- Glossaire de l'innovation en santé. URL : <<https://www.ciusante.org/Glossaire-de-innovation-en-sante>>.
- Glossaire terminologique en télésanté du Chu Sainte-Justine. URL : <https://www.chusj.org/getmedia/52783d1d-ca65-477f-84d2-93cb8e7105df/telesante_glossaire_fr.pdf.aspx>.
- Grand dictionnaire terminologique* (GDT). URL : <<https://vitrinelinguistique.oqlf.gouv.qc.ca/>>.
- Le Robert*. URL : <<https://dictionnaire.lerobert.com/>>.
- TERMIUM Plus®. URL : <<https://www.btb.termiumplus.gc.ca>>.

TermTestQA: evaluación del contenido y de la calidad de la información generada por ChatGPT en el campo de la terminología

CRISTINA VARGA

1. Introducción

El desarrollo y la diversificación de las API basadas en modelos lingüísticos de gran tamaño (LLM) ha abierto un abanico de posibilidades del uso de dichas aplicaciones en varios campos como la traducción, la medicina y las relaciones públicas. Dado el interés suscitado por el uso de la IA en el campo de la traducción y de la traducción especializada, recientemente han empezado a surgir planteamientos sobre su uso en la investigación terminológica (Massion, 2021). Sin embargo, de momento no se conocen datos concretos sobre el alcance y los límites del uso de las API basadas en LLM en el campo de la terminología.

Para responder a la pregunta si los LLM se pueden utilizar de forma efectiva en los aspectos prácticos y teóricos de la investigación terminológica que planteamos en el presente artículo, creemos pertinente emprender un análisis basado en datos empíricos. Consideramos que este enfoque es necesario porque los LLM adquieren cada vez más fluidez en la comunicación lingüística y se les presenta con mayor frecuencia como posibles fuentes de información y conocimiento. Es por esto que los LLM están sometidos a ensayos en distintos ámbitos, en especial en medicina (Goddard, 2023; Siontis *et al.*, 2024), aunque también en informática (Wang *et al.*, 2018), donde los desarrolladores de software quieren optimizar su rendimiento.

Hasta el presente se han publicado distintos modelos basados en cuestionarios que miden el rendimiento de los LLM. Entre ellos mencionamos *SuperGLUE* (Wang *et al.*, 2019) un test que mide la comprensión del lenguaje natural y *MMLU – Massive Multitask Language Understanding* (Hendrycks *et al.*, 2020), que evalúa las LLM en varios campos especializados como las matemáticas, la literatura entre otros. Existen también diversas bases de datos utilizadas para la evaluación del conocimiento del cual disponen los LLM como *Web Questions* (Berant *et al.*, 2013), *TriviaQA* (Joshi *et al.*, 2017), *EfficientQA* (Min *et al.* 2021) y *TruthfulQA* (Lin *et al.*, 2022).

En el campo de la traducción automática se han publicado unos cuantos modelos de evaluación (Chowdhery *et al.*, 2022). Los que más nos han llamado la atención son los modelos que han sido utilizados para la evaluación de la traducción automática realizada por los LLM en varios idiomas entre los cuales el rumano (Bojar *et al.*, 2016; Laskar *et*

al., 2023), ya que la primera fase de la evaluación de los LLM en terminología pretendemos realizarla también en este idioma.

La mayoría de los modelos mencionados tienen como objetivo detectar los puntos débiles de las respuestas generadas por los LLM y analizar las respuestas erróneas de las mismas. Dichas respuestas erróneas se clasifican en varias categorías a partir de meras inadvertencias hasta alucinaciones importantes (Shuster *et al.*, 2021; Zhou *et al.*, 2021; Krishna *et al.*, 2021).

Para evaluar el uso de los LLM en el campo de la terminología opinamos que sería aconsejable seguir el ejemplo de los modelos de evaluación desarrollados previamente en otros campos. Ejemplos que convendría adaptar a las particularidades y a las exigencias del ámbito de la terminología para obtener resultados óptimos. Por lo tanto, en el presente artículo se describirá la elaboración y la configuración de una base de datos construida a partir de preguntas destinadas a evaluar la fiabilidad de los LLM en el ámbito de la terminología en rumano. También se presentarán resultados preliminares del análisis de las respuestas de ChatGPT 3.5 en el ámbito de la terminología y futuras vías de desarrollo.

2. *TermTestQA* – descripción y estructura de la base de datos

En el ámbito de la terminología, la base de datos *TermTestQA* tiene por objeto medir el contenido y la calidad de las respuestas generadas por ChatGPT 3.5, comprobando la conformidad de las mismas con los datos científicos publicados en la literatura. Teniendo en cuenta este hecho, para la elaboración de las preguntas que constituyen la base de datos se han utilizado trabajos de referencia en este campo.

La base de datos *TermTestQA* incluye, en la presente fase de desarrollo, 300 preguntas relativas al ámbito de la terminología, que se ampliarán a un conjunto de 500, un volumen que se acerca más al de los otros modelos de evaluación de los LLM mencionados anteriormente. La elaboración de las preguntas está a cargo de una persona, a saber, la autora del presente artículo, y ninguna de ellas se ha generado automáticamente.

La mayoría de las preguntas consisten en una sola frase. La pregunta la más corta está formada por 14 caracteres mientras que la pregunta más compleja incluye 205 caracteres. En escasas situaciones, con el fin de especificar mejor la pregunta, se han utilizado 2 frases interrogativas como, por ejemplo: *Câte teorii ale terminologiei există în prezent? Care sunt acestea? (¿Cuántas teorías de la terminología existen hoy en día? ¿Cuáles son?).*

En la elaboración de las preguntas se han empleado varias estructuras como frases asertivas, exclamativas y frases interrogativas. Asimismo, se han utilizado distintos estilos y registros lingüísticos empezando por el registro académico y llegando hasta el registro familiar con el fin de observar si se produce alguna variación expresiva a la hora de generar las respuestas. En general, las preguntas formuladas se han expresado de una forma muy parecida a las preguntas que se encuentran en los exámenes de los estudiantes en terminología.

Desde un punto de vista temático, las preguntas pertenecen a varias categorías. Sin atenerse a una distribución uniforme bajo el punto de vista temático, se pueden observar 8 categorías de preguntas con diferentes grados de representatividad en el conjunto de la base de datos. Dichas categorías constituyen aspectos importantes de la disciplina terminológica, tales como: “historia y evolución”, “conceptos específicos”, “perspectivas teoréticas y escuelas”, “métodos de investigación”, “razonamientos y estimaciones”, “herramientas informáticas”, “autores y figuras de autoridad” y “bibliografía”. Para el mejor entendimiento del contenido de la base de datos utilizada para la evaluación de ChatGPT 3.5, a continuación, se describirán con ejemplos las categorías de preguntas mencionadas.

Asimismo, en la primera categoría temática “historia y evolución” que han incluido preguntas sobre los precursores de la terminología moderna, sobre autores de libros fundamentales para este campo de conocimiento, sobre los representantes de las distintas escuelas de terminología. La base de datos incluye también preguntas sobre la periodización del desarrollo de la terminología. La categoría más bien representada se refiere a “conceptos específicos” del ámbito de la terminología y contiene preguntas sobre conceptos como: termino, terminologización, árbol conceptual, ficha terminológica, definición terminológica, glosario, etc. Las preguntas que conciernen las “perspectivas teoréticas y escuelas terminológicas” se refieren a cuestiones sobre las primeras escuelas terminológicas, sobre las contribuciones teóricas fundamentales de las distintas teorías de la terminología, sobre distinciones teóricas entre dichas teorías etc. Las preguntas enfocadas hacia la “metodología de trabajo” incluyen cuestiones sobre como realizar operaciones concretas en el ámbito de la terminología. Por ejemplo *¿cómo estructurar una entrada en un glosario?* *¿Cómo se representa un árbol conceptual?* *¿Cómo se realiza un análisis conceptual?* etc. Preguntas cuya respuesta supone la descripción de un procedimiento específico.

Una categoría que evalúa ChatGPT 3.5 y los LLM de manera diferente es la que se refiere a “razonamientos y estimaciones” ya que se le pregunta al programa sobre la motivación/finalidad/importancia del uso de cierto concepto en el ámbito terminológico. Asimismo, una pregunta como *¿Con que fin se utilizan los sistemas conceptuales en terminología?* Necesita una respuesta elaborada a partir de un razonamiento y no se trata de una mera búsqueda en un corpus de textos.

Los aspectos técnicos de la investigación terminológica están cubiertos por preguntas sobre varias herramientas informáticas como extractores de terminología, concordancias, listas de frecuencia e infografías. La categoría de preguntas “autores y figuras de autoridad” contiene preguntas cuya respuesta debe mencionar el nombre de un autor o de una personalidad del ámbito de la terminología. Las preguntas se refieren a citas de varios autores, a acontecimientos y fechas importantes y a contribuciones de varios terminólogos. Un ejemplo de pregunta perteneciendo a esta categoría es *¿Cuáles son los más importantes terminólogos rumanos?*

La última categoría de preguntas concierne las “referencias bibliográficas” y se refiere a preguntas que deberían mencionar el título de una publicación, artículo o libro en el campo de la terminología. Como, por ejemplo: *¿Qué documento marca la pauta en el trabajo terminológico?*

Para una mejor comprensión de la distribución de las preguntas dentro de las diferentes categorías temáticas, proponemos la siguiente tabla:

Tabla 1 - *Distribución temática de las preguntas*

<i>Categoría</i>	<i>Núm. de preguntas</i>
historia y evolución	13
conceptos específicos	125
perspectivas teóricas y escuelas	24
métodos de investigación	7
razonamientos y estimaciones	83
herramientas informáticas	11
autores y figuras de autoridad	23
Bibliografía	14

El grado de dificultad y de especialización de las preguntas incluidas en la base de datos equivale al de un terminólogo en formación (estudiante de master o de doctorado) y la bibliografía utilizada para la elaboración de las mismas es de referencia y está disponible en acceso abierto en Internet. Concretamente, se ha recurrido a las siguientes obras: ISO 704 (2009; 2022), Cabré (1998), Pavel y Nolet (2001) y Wüster (2010). Ellas constituyen también el punto de referencia para la evaluación del contenido y de la calidad de las respuestas generadas automáticamente por ChatGPT 3.5.

En una primera fase, las preguntas se han elaborado en rumano y posteriormente se han traducido al inglés. Pretendemos seguir desarrollando la base de datos *TermTestQA*, añadiendo más preguntas hasta lograr un conjunto de datos que contenga un mínimo de 500 preguntas. Asimismo, nos proponemos traducir las preguntas hacia otros idiomas romances (FR, ES, CAT) para poder evaluar el rendimiento de ChatGPT en un mayor número de lenguas ya que los LLM están entrenados de manera diferente en función del idioma.

Todas las 300 preguntas se han sometido a un ensayo con ChatGPT 3.5, lo que nos ha permitido observar que este primer conjunto de preguntas se centra principalmente en aspectos teóricos y que *TermTestQA* se puede desarrollar también hacia la evaluación de los LLM sobre los aspectos prácticos de la terminología.

3. Metodología de evaluación de las respuestas generadas automáticamente

Para evaluar la calidad y el contenido de las respuestas de ChatGPT 3.5 entorno a la terminología el programa ha sido sometido a prueba con las 300 preguntas de las que dispone *TermTestQA* en esta fase de desarrollo. Como resultado se ha obtenido un corpus de respuestas generadas automaticamente por el programa. Dichos resultados se han guardado en un fichero Ms Word de 224 de paginas en formato A4 de extensión y 85.783 palabras que han sido analizadas. Dado el hecho de que la base de datos *TermTestQA* se

sigue desarrollando, consideramos los presentes resultados como un primer acercamiento hacia la evaluación de los LLM en el ámbito de la terminología, evaluación que será afinada más adelante a partir de un conjunto más complejo de preguntas.

Para la evaluación de las respuestas se plantea el siguiente protocolo de análisis:

Respuesta ChatGPT	Aspectos lingüísticos formales	La extensión de la respuesta está adaptada a la complejidad de la cuestión	Si/No
		Registro lingüístico	culto/estándar/familiar/jerga
		Problemas de ortografía	Si/No
		Problemas de puntuación	Si/No
	Contenido discursivo	Problemas gramaticales	Si/No
		Respuesta informativa	completa/parcial
		Respuesta redundante	completa/parcial
		Respuesta vacía de contenido	completa/parcial
	Calidad de la información generada	Información correcta	Si/No
		Información incorrecta	Imprecisión
			Omisión
			Alucinación

Las respuestas generadas por ChatGPT 3.5 han sido examinadas y analizadas por un investigador humano, la autora del presente artículo, comprobándose 3 aspectos principales: a) aspectos lingüísticos formales, b) contenido discursivo y c) calidad de la información.

En la categoría “aspectos lingüísticos formales (a)”, se han analizado 5 aspectos de las respuestas generadas por ChatGPT 3.5:

1. la adaptación de la respuesta a la complejidad de la pregunta;
2. el registro empleado en la generación de las respuestas (culto, estándar, familiar, jerga);
3. ortografía correcta;
4. puntuación correcta;
5. morfosintaxis correcta.

Con el fin de evaluar el “contenido discursivo de las respuestas (b)”, se ha examinado la calidad de las mismas y se han identificado 3 categorías diferentes: *respuesta informativa*, *respuesta redundante* y *respuesta vacía de contenido*. En el marco de esta investigación, entendemos por respuesta informativa cualquier respuesta generada por ChatGPT 3.5 que aporte datos científicos relevantes, completos y veraces a una pregunta. En el mismo marco, se consideran como respuestas redundantes las respuestas que presentan un contenido informativo repetitivo, formulado de distintas maneras en diferentes partes de la respuesta generada automáticamente. Asimismo, asumimos que las respuestas vacías de contenido son aquellas respuestas generadas por ChatGPT 3.5, o partes de las mismas, en las que, aún tratándose de un texto articulado de forma completa y correcta, al usuario no se le transmite ninguna información.

En cuanto a la calidad de la información generada, se distinguen dos categorías de respuestas: 1) respuestas correctas y 2) respuestas incorrectas. Se consideran como *respuestas correctas* aquellas que presentan información completa y adecuada sobre el tópico de la pregunta, teniendo como punto de referencia la bibliografía mencionada anteriormente.

En cuanto a las respuestas incorrectas, en función del tipo de error que presentan, se pueden clasificar en 3 categorías distintas: 1) respuestas imprecisas, 2) omisiones y 3) alucinaciones.

En la presente investigación se entiende por *respuestas imprecisas* a aquellas respuestas que, a pesar de presentar información correcta hasta cierto punto, contienen errores. Otra categoría de errores detectada en el corpus de respuestas generadas por ChatGPT 3.5 la constituyen las *omisiones*. Se trata de respuestas en las que la información se presenta correctamente, pero de forma incompleta. Una categoría distinta la constituyen las *alucinaciones de la IA*, término con el que se designan las situaciones en las que los LLM generan respuestas equivocadas que no se basan en datos/informaciones auténticas. Consideramos las alucinaciones como el más grave tipo de respuestas erróneas, ya que pueden manipular datos erróneos a los usuarios que utilizan ChatGPT como medio de información.

A continuación, se presentarán los resultados obtenidos en esta primera fase de evaluación de ChatGPT 3.5 a partir del análisis cuantitativo y cualitativo de las respuestas generadas por el programa en rumano, en el ámbito de la terminología.

4. Resultados de la evaluación de las respuestas de ChatGPT 3.5

4.1 Análisis cuantitativo

Analizando las respuestas de ChatGPT 3.5 en esta primera fase de desarrollo de *TermTestQA* se ha observado que, desde un punto de vista cuantitativo, 100% de las 300 respuestas generadas por el programa no han sido adaptadas a la complejidad de la pregunta. Incluso si la respuesta correcta tenía que constar solo del nombre de un autor, de la denominación de una teoría o del título de un libro, ChatGPT 3.5 ha generado en rumano textos complejos, en general estructurados en tres párrafos. Para obtener una respuesta más reducida, el usuario tiene que pedir explícitamente que dicha respuesta sea expresada en un número exacto de palabras.

En cuanto al registro lingüístico, al formular las respuestas en rumano, en 100% de los casos, sin importar el registro lingüístico utilizado en la pregunta, ChatGPT 3.5 contesta en un registro estándar. Para un cambio de registro, el usuario tiene que mencionar el registro exacto que se tiene que utilizar para una respuesta en concreto. También el cambio de registro afecta generalmente la calidad del contenido de la respuesta generada.

Las respuestas generadas por ChatGPT en rumano presentan problemas de morfosintaxis y de ortografía que, en el estado presente del análisis apreciamos en un 2% del texto generado. Apreciamos que dichos problemas se pueden explicar por el hecho de que el rumano es un idioma con el que los LLM no son muy bien entrenados, hecho mencionado también por otros estudiosos (véase 1. Introducción). No se han observado problemas de puntuación en las respuestas generadas en rumano por ChatGPT 3.5.

En cuanto al contenido de las respuestas, de momento se han obtenido 47,28% de respuestas correctas y 52,72% de respuestas erróneas a las preguntas del *TermTestQA*. Entre las respuestas erróneas, se han registrado 15,45% respuestas imprecisas, 9,09% omisiones y 28,18% alucinaciones de la IA. Lo que nos llama más la atención en este primer análisis no es tanto el porcentaje elevado de respuestas correctas generadas por el programa sino el porcentaje muy elevado de las alucinaciones de la IA. Casi un tercio de las preguntas son formadas completamente o parcialmente por alucinaciones, o sea datos/hechos inventados por el LLM, lo que no recomienda las API basados en dichos modelos lingüísticos como fuente fiable de información especializada.

Tal como se ha mencionado, los presentes datos cuantitativos son provisionales e ilustran el estado presente de desarrollo de la base de datos *TermTestQA*. De esta manera, solo se han analizado las respuestas de ChatGPT 3.5 en rumano. En el futuro, contamos con ampliar la base de datos *TermTestQA* en un 40% para una evaluación más detallada de los LLM. Igualmente, pretendemos transformar la base de datos en una herramienta de evaluación de los LLM en un entorno multilingüe enfocado hacia las lenguas románicas.

4.2 Análisis cualitativo

Desde un punto de vista cualitativo, los aspectos lingüísticos formales, de las respuestas generadas por ChatGPT 3.5 presentan varios tipos de errores ortográficos, uno de los más importantes siendo el uso de los caracteres especiales en rumano. Se ha observado en las respuestas del asistente conversacional el uso alternativo de palabras con y sin caracteres especiales como en el ejemplo: (1) Termenii din nomenclatură sunt selectați și definiți în conformitate cu criterii standardizate.

En otras situaciones, ChatGPT no respeta la normativa del rumano utilizando una forma de sustantivo equivocada “tehnicii” (de genitivo/dativo) en lugar de un adjetivo: (2) Densitatea terminologică se referă la proporția de termeni specializați sau *tehnicii* care apar într-un text sau într-un corpus de texte în raport cu numărul total de cuvinte.

También se ha observado el uso de palabras erróneas en rumano como (3) “laço-vegetarian” en lugar de “lacto-vegetarian”, no resulta claro si se trata de una alucinación de la IA, o sea una palabra inventada por el LLM o si es una unidad léxica que está mal escrita en las fuentes a los que acude ChatGPT para la minería de datos.

Dadas los límites impuestos al presente artículo no citaremos ejemplos de respuestas que no han sido correctamente adaptadas a la complejidad de la respuesta, solo mencionaremos el hecho de que a una pregunta del tipo: (4) *¿Cómo se denomina la herramienta informática que sirve para visualizar todos los contextos en los que aparece un término dado en un texto?* y cuya respuesta más sencilla sería herramienta de concordancias, ChatGPT 3.5 genera en rumano una respuesta de 201 palabras, estructurada en 3 párrafos, explicando que se trata de un “motor de búsqueda”.

Desde el punto de vista del contenido discursivo, la mayoría de las respuestas de ChatGPT 3.5 presentan una estructura heterogénea. Parte de las respuestas son informativas, pero también presentan fragmentos redundantes y textos vacíos de contenido.

Como, por ejemplo, (5) *¿En qué época y dónde se originó la terminología?* La respuesta generada presenta en 156 palabras una parte informativa:

Terminologia ca disciplină distinctă și sistematică a apărut în secolul al XX-lea, odată cu publicarea lucrării lui Eugen Wüster intitulată “Einführung in die allgemeine Terminologielehre”. (La terminología como disciplina independiente y sistemática surgió en el siglo XX con la publicación de la obra de Eugen Wüster “Einführung in die allgemeine Terminologielehre”)

Una parte redundante:

Terminologia ca disciplină distinctă și sistematică [...], terminologia ca disciplină academică distinctă, a început să fie dezvoltată abia în secolul al XX-lea, [...]. Astfel, terminologia a apărut ca disciplină distinctă în secolul al XX-lea [...]. (La terminología como disciplina distinta y sistemática [...], la terminología como disciplina académica distinta, empezó a desarrollarse sólo en el siglo XX, [...]. Así pues, la terminología surgió como disciplina distinta en el siglo XX [...])

Y una parte de contenido vacío ya que se puede utilizar casi en cualquier contexto:

[...] odată cu evoluția științelor lingvistice și a altor domenii de cunoaștere. [...] iar lucrarea sa a reprezentat un punct de cotitură în dezvoltarea acestei discipline, [...]. ([...] con la evolución de las ciencias lingüísticas y otros campos del conocimiento. [...] y su obra supuso un punto de inflexión en el desarrollo de esta disciplina, [...])

En cuanto a la calidad de la información proporcionada por ChatGPT 3.5 una gran parte de las respuestas son acertadas. Como en el ejemplo a la pregunta *¿Las palabras del vocabulario general pueden convertirse en términos? ¿Mediante qué proceso?* Se ha generado una respuesta formada por 124 palabras, completamente correcta de la que presentamos solo una parte:

(6) Da, cuvintele din vocabularul general pot deveni termeni într-un anumit domeniu de cunoaștere prin procesul de specializare terminologică sau terminologizare. (Sí, las palabras del vocabulario general pueden convertirse en términos de un campo de conocimiento concreto mediante el proceso de especialización terminológica o terminologización)

Sin embargo, la mayoría de las respuestas al cuestionario *TermTestQA* pertenecen a la categoría de respuestas erróneas, desde diferentes puntos de vista. Un tipo de error recurrente son las respuestas con un contenido informativo poco preciso como en el caso de la respuesta a la pregunta *¿Cuál es el primer texto que atestigua la actividad terminológica?*:

(7) Atestarea activității terminologice este dificil de stabilit cu precizie, deoarece termenii au fost utilizati și definiți încă din cele mai vechi timpuri. Înșã, unul dintre primele texte care sugerează o preocupare pentru definirea și standardizarea

termenilor într-un domeniu specific este “De Architectura” (“Despre arhitectură”) scris de Marcus Vitruvius Pollio, cunoscut simplu sub numele de Vitruvius, în secolul I î.Hr. (Resulta difícil determinar la naturaleza exacta de la actividad terminológica, ya que los términos se han utilizado y definido desde la antigüedad. Sin embargo, uno de los primeros textos que sugieren una preocupación por definir y normalizar términos en un campo específico es “De Architectura”, escrito por Marco Vitruvio Polio, conocido simplemente como Vitruvio, en el siglo I a.C.)

Otra categoría de respuestas erróneas la constituyen las omisiones. Un ejemplo de omisión se puede observar en la respuesta a la pregunta *¿Cuál es la relación fundamental en la que se basa la terminología?* La respuesta menciona solo la relación concepto-termino omitiendo el tercer elemento de la relación, el objeto:

(8) Relația fundamentală pe care se bazează terminologia este legătura dintre termeni și concepte. (La relación fundamental en la que se basa la terminología es el nexo entre términos y conceptos)

Y, por último, la categoría mejor representada y al mismo tiempo la más problemática desde el punto de vista informativo está constituida por las *alucinaciones de la IA*. Se trata de respuestas cuyo contenido informativo ha sido totalmente inventado por el LLM y que no se basan en ningún hecho real. Para ilustrar esta categoría se ha escogido la respuesta insólita y difícil de elaborar a la pregunta *¿Cuál es el primer glosario terminológico que se conoce? ¿Cómo se titula?*

(9) Primul glosar terminologic cunoscut este “Glosarul lui Esarhaddon”, care datează de la sfârșitul secolului al VII-lea î.Hr. Esarhaddon a fost un rege asirian care a domnit între 681 și 669 î.Hr. (El primer glosario terminológico que se conoce es el “Glosario de Esarhaddon”, que data de finales del siglo VII a.C. Esarhaddon fue un rey asirio que gobernó entre los años 681 y 669 a.C.)

Se puede observar que la respuesta contiene información inventada sobre un glosario inexistente que se pone en relación con una figura histórica real de la antigüedad, el rey asirio Esarhaddon. Si se le pide al ChatGPT de mencionar una fuente bibliográfica que mencione dicho glosario, el LLM inventa una fuente bibliográfica:

(10) Un titlu care menționează “Glosarul lui Esarhaddon” este: “Mesopotamian Glossaries from Esarhaddon’s Library: First-Person Voice and Semantic Agency in Early Second Millennium BCE Akkadian Lexicography”. (Una de las fuentes que menciona el “Glosario de Esarhaddon” es: “Glosarios mesopotámicos de la biblioteca de Esarhaddon: voz en primera persona y agencia semántica en la lexicografía acadia de principios del segundo milenio a.C.”)

Dada la facilidad con la cual se presentan fuentes y datos inventados, consideramos que este tipo de contenidos es el más peligroso por el usuario que piensa utilizar ChatGPT 3.5 como fuente de información especializada.

5. Conclusiones

Con esta contribución hemos querido destacar la importancia de la evaluación de los LLM en ámbitos especializados como la terminología en el contexto del desarrollo y del uso generalizado de los LLM. Consideramos que la base de datos *TermTestQA* es un barómetro que le puede indicar al terminólogo y a los profesionales del campo cual es de nivel de fiabilidad de herramientas informáticas como ChatGPT, cuáles son los puntos fuertes y las debilidades de los LLM.

Con estos primeros resultados procedentes de la aplicación del protocolo de análisis, nos motivamos a seguir desarrollando la base de datos para obtener información más detallada sobre la forma y el contenido de las respuestas generadas por ChatGPT. Opinamos que una vez hayamos obtenido resultados cualitativos y cuantitativos detallados en rumano, estaremos en condiciones de ampliar el protocolo de análisis a otros idiomas como el inglés, el francés, el español y el catalán. De este modo obtendremos conjuntos de datos que podrán compararse posteriormente para observar respuestas contrastadas en distintos idiomas.

Una vez finalizado el desarrollo de *TermTestQA* y demostrada su utilidad, la base de datos de preguntas en los cinco idiomas (EN, FR, ES, CAT, RO) se publicará en acceso abierto para su posterior uso por parte de terminólogos, profesores y estudiantes que deseen evaluar el rendimiento de los LLM en el campo de la terminología.

Actualmente, después de haber puesto a prueba la base de datos, podemos afirmar que, en rumano, ChatGPT 3.5 presenta dificultades a distintos niveles, como por ejemplo la macroestructura discursiva de las respuestas generadas, la ortografía y el léxico utilizado. Respecto al contenido informativo de las respuestas proporcionadas, se puede decir que no sólo es informativo, como cabría esperar de un diálogo especializado, sino que también presenta redundancia y fragmentos textuales vacíos de contenido.

Por lo que respecta a la precisión de las respuestas proporcionadas, ChatGPT no alcanza un nivel suficientemente elevado para constituir una fuente de información en el ámbito de la terminología en rumano, ya que más de la mitad de las respuestas analizadas resultaron ser erróneas. Entre los tipos de errores que se producen al interrogar el asistente conversacional en rumano, los más graves son las alucinaciones de la IA.

Referencias bibliográficas

- Berant J *et al.*, *Semantic parsing on Freebase from question-answer pairs*, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013. URL: <<https://aclanthology.org/D13-1160.pdf>>, último acceso el 28-11-2025.
- Bojar O. *et al.*, *Findings of the 2016 conference on machine translation*, in *Proceedings of the First Conference on Machine Translation*, 2016. URL: <<https://aclanthology.org/W16-2301/>>, último acceso el 28-11-2025.
- Cabré M.T., *Terminology: theory, methods and applications*, Philadelphia, John Benjamins, 1998.
- Chowdhery A. *et al.*, *PaLM: Scaling language modeling with pathways*, 2021. URL: <<https://arxiv.org/abs/2204.02311>>, último acceso el 28-11-2025.

- Goddard J., *Hallucinations in ChatGPT: a Cautionary Tale for Biomedical Researchers*, in «The American Journal of Medicine», 136/11, 2023. URL: <<https://doi.org/10.1016/j.amjmed.2023.06.012>>, último acceso el 28-11-2025.
- Hendrycks D. et al., *Measuring Massive Multitask Language Understanding*, 2020. URL: <<https://arxiv.org/abs/2009.03300>>, último acceso el 28-11-2025.
- ISO 704: 2009, *International Standard, Terminology work – Principles and methods*. URL: <https://ddialliance.org/sites/default/files/shared/2012/DDI%20Moving%20Forward/ISO/ISO704_2009.pdf>, último acceso el 28-11-2025.
- Joshi M. et al., *Triviaqa: a large scale distantly supervised challenge dataset for reading comprehension*, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017. URL: <<https://aclanthology.org/P17-1147/>>, último acceso el 28-11-2025.
- Krishna K. et al., *Hurdles to progress in long-form question answering*, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021. URL: <<https://aclanthology.org/2021.naacl-main.393/>>, último acceso el 28-11-2025.
- Laskar M. et al., *A Systematic Study and Comprehensive Evaluation of ChatGPT on Benchmark Datasets*, in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023. URL: <<https://aclanthology.org/2023.findings-acl.29/>>, último acceso el 28-11-2025.
- Lin S. et al., *TruthfulQA: Measuring How Models Mimic Human Falsehoods*, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022. URL: <<https://aclanthology.org/2022.acl-long.229/>>, último acceso el 28-11-2025.
- Massion F., *Terminology in the Age of Artificial Intelligence*, Berlin, Peter Lang, 2021.
- Min S. et al., *NeurIPS 2020 EfficientQA Competition: Systems, Analyses and Lessons Learned*, in «Neural Information Processing Systems», 2021. URL: <<https://arxiv.org/abs/2101.00133>>, último acceso el 28-11-2025.
- Pavel S., Nolet D., *Handbook of Terminology*, Quebec, Minister of Public Works and Government Services Canada, 2001.
- Shuster K. et al., *Retrieval augmentation reduces hallucination in conversation*, in *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2021. URL: <<https://aclanthology.org/2021.findings-emnlp.320/>>, último acceso el 28-11-2025.
- Siontis K.C. et al., *ChatGPT hallucinating: can it get any more humanlike?*, in «European Heart Journal», 45/5, 2024, pp. 321-323.
- Wang A. et al., *GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding*, in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2018. URL: <<https://aclanthology.org/W18-5446/>>, último acceso el 28-11-2025.
- Wang A. et al., *SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems*, 2019. URL: <<https://arxiv.org/abs/1905.00537>>, último acceso el 28-11-2025.
- Wüster E., *Introducción a la teoría general de la terminología y a la lexicografía terminológica*, Barcelona, Publicacions de l'Institut Universitari de Lingüística Aplicada, 2010.
- Zhou C. et al., *Detecting hallucinated content in conditional neural sequence generation*, in *Findings of the Association for Computational Linguistics: ACL-IJCNLP*, 2021. URL: <<https://arxiv.org/abs/2011.02593>>, último acceso el 28-11-2025.

En droit pénal français, il y a *prévenus* et *prévenus* : un cas de polysémie interdisciplinaire pas comme les autres

ANCA-MARINA VELICU

1. Du caractère interdisciplinaire de notre entreprise

Le caractère interdisciplinaire de la terminologie comme théorie d'une pratique intégratrice est avant toute chose la conséquence des relations très étroites que le travail terminologique entretient avec les domaines d'activité ou du savoir dont on étudie les concepts (et les désignations) – les Sachwissenschaften de Wüster 1993 (« domaines de référence » des linguistes).

Nous traiterons ici de la terminologie du droit pénal, et viserons la migration d'un terme/concept entre deux matières pénales distinctes, qui font l'objet de deux codes différents : la procédure pénale, d'une part, et le droit pénitentiaire, de l'autre. Une réflexion sur la circulation des concepts d'une branche du droit à l'autre, ne participera d'une « interdisciplinarité de proximité » (Jollivet, Legay, 2005 : 185), que si le statut disciplinaire des branches concernées a été dûment établi. Trois types d'arguments peuvent être avancés à ce propos : (i) leur autonomie conceptuelle relative (la procédure pénale est un droit de forme, le droit pénitentiaire, un droit de fond, à l'instar du droit pénal général – cf. Plawski, 1976 : 11 ; 267), (ii) leur présence disjointe dans les curriculums académiques¹ et dans les publications de spécialité² et (iii) leur autonomie normative : CPP_{FR} 1959, CPénit_{FR} 2022³.

Dans son étude ayant eu précisément « pour objectif de démontrer » la constitution de la « nouvelle discipline » juridique, Stanislas Plawski en soulignait, sans le nommer, le caractère en soi interdisciplinaire (discipline de frontière) : l'émergence du droit pénitentiaire a entraîné « la disparition d'une césure entre le droit pénal d'un côté et la criminologie et toutes les sciences criminelles d'un autre côté » (1976 : 11-12).

¹ Formations diplômantes et certifications offertes par les universités dans chacune des deux matières pénales, cours distincts dans les curriculums de licence ou de master en Droit.

² Des bibliographies scientifiques distinctes : <<http://iscrim.labo.univ-poitiers.fr/wp-content/uploads/sites/267/2020/09/Bibliographie-Droit-Penal-et-ProcEDURE-penale-sept.2020.pdf>> (consulté le 19-11-2025) ; <www.enap.justice.fr/sites/default/files/biblio_droit_penitentiaire_2017.pdf> (consulté le 19-11-2025).

³ Pour des arguments plus techniques.

2. Méthodologie de la recherche

Nous entamerons une analyse terminologique-conceptuelle fondée sur le corpus (*vs* guidée par le corpus), afin de tester deux hypothèses complémentaires sur les relations possibles entre l'usage actuel du nom juridique *prévenu* (ainsi que du participe passé passif et de l'adjectif déverbal correspondants), respectivement en procédure pénale et en droit pénitentiaire.

Première hypothèse : polysémie interdisciplinaire⁴, du nom juridique, par transfert notionnel d'un sous-domaine du droit, à l'autre.

Seconde hypothèse : homonymie (deux sous-domaines du droit pénal, deux termes, deux procédés de conversion autonomes à partir d'acceptions distinctes du vocable⁵ verbal *prévenir*).

Si la première hypothèse jouissait finalement de plus d'arguments déductifs (raisonnement linguistique) et inductifs (textuels), le procédé de création terminologique à l'origine de l'émergence de l'usage pénitentiaire relèverait en soi de la dimension interdisciplinaire (emprunt interne dit « transdisciplinaire » dans la norme ISO 704 : 2022).

Le corpus étudié est composé de textes normatifs (codes français : pénal, de procédure pénale et pénitentiaire), mais des textes de jurisprudence seront évoqués au besoin (référéncés alors en notes de bas de page).

Comme déjà dit, nous nous attacherons à tirer au clair les relations entre deux acceptions juridiques du nom *prévenu*, en français contemporain. Pour y arriver, nous ferons également appel à l'argument diachronique⁶, et à l'argument des équivalents interlinguaux (roumains). D'où l'intégration au corpus, des anciens codes français (*Code d'instruction criminelle* de 1808, *Code pénal* de 1810) et des codes roumains pénal et de procédure pénale (les nouveaux codes actuellement en vigueur, mais également les codes « socialistes » et les codes pénal et de procédure pénale de Carol II). Nous allons désigner ces textes normatifs à l'aide des abréviations proposées en bibliographie, l'année étant, sauf spécification contraire, celle de leur première version. La mise à jour des codes en vigueur est prise en compte, mais nous spécifierons l'année d'introduction ou de modification de tel ou tel article cité ou mentionné, lorsqu'elle est nécessaire pour la logique de l'exposé. Les concepts juridiques traités n'étant pas souvent définis par le législateur français, pour la présentation des systèmes notionnels, nous avons en outre consulté des textes de divulgation juridique, officiels (textes d'information citoyenne⁷), et des dictionnaires juridiques explicatifs. Les dictionnaires de langue générale (le *TLFi* –

⁴ Pour une mise en œuvre du concept de polysémie interdisciplinaire dans une étude de terminologie descriptive (polysémie du nom *adaptation* dans les sciences du vivant), cf. Simonet, 2009.

⁵ *Mot-forme*, *lexème*, *lexie* et *vocable* sont entendus ici au sens de Polguère, 2008 : 46-60.

⁶ Cf. Dury, Picton (2009) pour une mise en perspective théorique des enjeux de la recherche diachronique en terminologie contemporaine ; Zanola (2014) et Grimaldi (2017) comme illustrations exemplaires de la méthode diachronique, en terminologie descriptive.

⁷ Notamment des textes publiés sur les sites <www.vie-publique.fr/> et <www.justice.gouv.fr> (consultés le 19-11-2025).

état 1994 ; le *Larousse* et le *Robert* en ligne – dictionnaires mis à jour 2024) nous auront permis de traiter la question étymologique.

3. Deux concepts juridiques différents

En droit pénal français (contemporain), en fonction de la phase procédurale (pendant ou avant le procès pénal), les prévenus sont opposés aux mis en examen, aux témoins assistés et aux suspects. Pendant l'enquête de police judiciaire, on est appelé *suspect* (ou *personne suspectée*, ou encore *personne mise en cause*⁸) ; pendant l'instruction, le suspect deviendra *témoin assisté*, voire (si les choses se corsent) *mis en examen* (synonyme : *personne mise en examen*) ; une fois renvoyé devant un tribunal, l'auteur présumé de l'infraction sera appelé, selon la gravité de l'infraction, *prévenu* (s'il est poursuivi pour un délit ou pour une contravention) et *accusé* (s'il est poursuivi pour un crime). L'opposition *prévenu vs accusé* est interne à une seule et même phase de la procédure pénale : le procès. Les deux espèces ainsi discriminées sont opposées, en bloc, aux mis en examen, aux témoins assistés et aux suspects (statuts distingués en fonction du seul critère de la chronologie procédurale). Nous sommes donc en présence d'un système conceptuel générique à deux niveaux de subordination, même si le genre prochain des prévenus *vs* accusés n'est pas nommé (*prévenu* avait fonctionné, dans le CIC_{FR} 1808, comme terme générique valant des phases procédurales antérieures à la mise en accusation, et apparaît en variation libre avec *inculqué*, dans les articles sur l'instruction, mais n'a jamais remplacé *accusé* dans les contextes relatifs au procès).

Les prévenus sont renvoyés devant un tribunal de police (pour des contraventions) ou bien correctionnel (pour des délits), tandis que les accusés sont jugés par une cour criminelle (peine encourue de 15 à 20 ans)⁹ ou en cour d'assises (peine encourue de plus de 20 ans, et appels). Les prévenus encourent des peines d'amende (pour des contraventions) ou d'emprisonnement (pour les délits), tandis que les accusés encourent des peines de réclusion (pour les crimes de droit commun) et de détention criminelle (pour les crimes politiques). Un prévenu reconnu non coupable fera l'objet d'une relaxe, un accusé, d'un acquittement. C'est dire que l'opposition *prévenu vs accusé* est opérée à la fois pendant les débats et pendant le prononcé du jugement.

Un autre usage juridique du nom *prévenu* est attesté dans un Décret de 1985 : « personne placée en détention provisoire ».

Le plus clair de notre contribution visera justement à préciser les relations entre *prévenu (vs accusé)*, d'une part, et *prévenu – personne en détention provisoire*, de l'autre.

⁸ Synonyme attesté dans le CPP_{FR} 1959/version à jour 2024 : *personne mise en cause* (Art. 77-2, Art. 48-1). Variante de divulgation juridique : *le mis en cause* (<www.vie-publique.fr/fiches/268563-quels-sont-les-differents-intervenants-dune-affaire-penale>, consulté le 19-11-2025).

⁹ Institution de date plus récente, créée par la Loi n° 2019-222 du 23 mars (5 juges, pas de jurés, alors que la cour d'assises compte, elle, seulement 3 juges, mais également 6 jurés).

4. Analyse du corpus

Des 24 occurrences de *prévenu* dans le CP_{FR} 1994, aucune n'est en relation avec la notion de détention provisoire¹⁰, les concepts typiquement associés étant le délit, l'emprisonnement, l'amende (les jours-amende), le travail d'intérêt général, ainsi que des concepts généraux tels que la culpabilité ou l'action civile. Bien que *contravention* y soit attesté 63 fois dans la « Partie législative » (contre 166 occurrences dans la « Partie réglementaire »), aucun des contextes n'instancie le cumul (à distance) avec *prévenu*, preuve de l'association prototypique *prévenu* – *délit*. Il n'y a pas une seule définition du concept désigné, seule l'étude des cooccurents permettant de trancher sa compréhension. Comparer :

(1) Lorsqu'un *délit* est puni d'une peine d'*emprisonnement*, la juridiction peut prononcer une *peine de jours-amende* consistant pour le *condamné* à verser au Trésor une somme dont le montant global résulte de la fixation par le juge d'une contribution quotidienne pendant un certain nombre de jours. Le montant de chaque jour-amende est déterminé en tenant compte des ressources et des charges du *prévenu* [...]. (CP_{FR} 1994, Art. 131-5)

(2) En même temps qu'elle se prononce sur la *culpabilité* du *prévenu*, la juridiction statue, s'il y a lieu, sur *l'action civile*. (CP_{FR} 1994, Art. 132-58)

Du point de vue de la catégorisation syntaxique, c'est le nom qui y est le plus souvent attesté ; le participe passé (passif, peut-être adjectivé) n'y a que 4 occurrences :

(3) L'ajournement avec injonction ne peut intervenir qu'une fois ; il peut être ordonné même si la *personne physique prévenue* ou le représentant de la *personne morale prévenue* n'est pas présent. (CP_{FR} 1994, Art. 132-68¹¹)

Le CPP_{FR} 1959 comporte 331 occurrences de *prévenu*, mais la séquence *personne prévenue* n'y est attestée que 2 fois (Art. 63-3-1 et 63-3-3) : *prévenu* signifie alors simplement « être avisé de quelque chose ». Il s'agit d'une personne gardée à vue¹² qui a le droit de prévenir quelqu'un, et la ou les personnes prévenues sont celles à qui elle aura téléphoné. Ce sont des occurrences triviales sans vocation terminologique.

¹⁰ On n'y parle que de *personne détenue provisoirement* (CP_{FR} 1994, Art. 434-11, en vigueur depuis 2002).

¹¹ Cf. aussi Art. 132-60.

¹² La garde à vue n'est pas à confondre avec la détention provisoire ; la première est une mesure (de courte durée : jours), prise par la police judiciaire pendant l'enquête préliminaire ; la seconde, une mesure privative de liberté proposée par le juge d'instruction et décidée par un juge des libertés et de la détention, et peut durer des semaines voire des mois. La première est exécutée dans les locaux de la police, la seconde, en maison d'arrêt. Le sujet passif d'une garde à vue est un suspect, celui d'une détention provisoire, un mis en examen, un prévenu ou un accusé. Cf. <www.service-public.fr/particuliers/vosdroits/F14837> et <www.service-public.fr/particuliers/vosdroits/F1042> (consultés le 19-11-2025).

Et si, dans ce code-ci, le mot *prévenu* apparaît également à proximité du terme *détention provisoire*, les contextes concernés montrent clairement que ce n'est là qu'une association contingente. Ainsi, en (4) ci-après, doublement virtualisée par *si* conditionnel, et par le verbe modal *pouvoir*, la détention provisoire du prévenu ne relève décidément pas du nécessaire :

(4) [...] *si* les éléments de l'espèce lui paraissent exiger une mesure de *détention provisoire*, le procureur de la République *peut* traduire le *prévenu* devant le *juge des libertés et de la détention* [...]. (CPP_{FR} 1959, Art. 396, italique de notre main)

Dans la « Partie législative » du code, *prévenu* (n.m.) est toujours en usage (emploi référentiel), et tous les contextes où *prévenu* et *détention provisoire* se côtoient, soit virtualisent cette association, par des marques explicites de modalisation¹³, soit en réduisent la portée extensionnelle à coup de relatives restrictives (éventuellement réduites), qui impliquent l'existence de membres de la classe non concernés par la mesure en question :

(5) *Les personnes mises en examen, prévenus ou accusés soumis à la détention provisoire* la subissent dans une maison d'arrêt. (CPP_{FR} 1959, Art. 714)

Dans la « Partie réglementaire », toutefois, *prévenu* apparaît également en mention (emploi métalinguistique), comme désignation d'une personne en détention provisoire (définition par compréhension assortie de l'énumération des catégories d'objets susceptibles de tomber sous le concept) :

(6) Sont désignées dans le présent titre par le mot *détenus*, les personnes faisant l'objet d'une mesure privative de liberté à l'intérieur d'un établissement pénitentiaire. Sont désignés par le mot *condamnés*, uniquement les condamnés ayant fait l'objet d'une décision ayant acquis le caractère définitif. [...]
Sont indistinctement désignés par le mot *prévenus*, tous les détenus qui sont sous le coup de poursuites pénales et n'ont pas fait l'objet d'une condamnation définitive [...], c'est-à-dire aussi bien les *personnes mises en examen*, les *prévenus*, et les *accusés*, que les *condamnés* ayant formé *opposition*, *appel* ou *pourvoi*. (CPP_{FR} 1959, Partie réglementaire. Décrets simples, Art. D50, version modifiée 1998, première version 1985¹⁴)

L'hyperonyme de *prévenu* est alors *détenu*, et son antonyme – *condamné* ou (*personne*) *détenu(e) condamné(e)*. Quant à *prévenu* (*vs accusé*, *vs personne mise en examen*), sans en être un vrai hyponyme, il fonctionne dans cet environnement comme désignation d'une classe d'objets susceptibles d'illustrer également ce concept-ci.

¹³ D'autres exemples de virtualisation : Art. 396 ([le juge des libertés et de la détention] « peut placer le prévenu en détention provisoire jusqu'à sa comparution devant le tribunal ») ; Art. 401-1 (« dans le cas où la personne est placée en détention provisoire », où « la personne » reprend « le prévenu » de la phrase précédente : « si le prévenu [qui n'a pas comparu] est arrêté à la suite d'un mandat d'amener ou d'arrêt »).

¹⁴ Avec *inculpés* pour *personnes mises en examen*.

Si, dans le CPP_{FR} 1959, le nom *prévenu* est employé aussi avec le sens de *personne en détention provisoire*, le CPénit_{FR} 2022 lui-même ne reflète nullement cette évolution. Seules y sont attestées les désignations analytiques *personne prévenue* et *personne (placée) en détention provisoire* (dont l'antonyme sera *personne condamnée* – avec la variante *personne détenue condamnée* – et l'hyperonyme *personne détenue*). La complémentarité relative des deux codes est sous-tendue par des préférences syntaxiques distinctes : le CPP_{FR} 1959 atteste surtout le nom *prévenu*, tandis que le CPénit_{FR} 2022 n'atteste que le participe passé passif, adjectivé *prévenu-e*. Les articles L112-3 et D112-28 du CPénit_{FR} 2022 désignent même, par *personne prévenue*, un *prévenu* (*vs accusé*) ; les corrélations *tribunal judiciaire/personne prévenue*, *cour d'appel/personnes appelantes* et *cour d'assises/personnes accusées* en témoignent :

(7) Une *maison d'arrêt* est située près de chaque *tribunal judiciaire*, de chaque *cour d'appel* et de chaque *cour d'assises*, sauf auprès des tribunaux et des cours qui sont désignés par décret. Dans ce dernier cas, le décret détermine la ou les maisons d'arrêt où sont retenues *les personnes prévenues, appelantes ou accusées* ressortissant à chacune de ces juridictions. (CPénit_{FR} 2022, Art. L112-3)

Il faut remarquer le statut peu terminologique du verbe *prévenir*, en droit pénal français contemporain (notamment en matière procédurale) : l'acception juridique *citer en justice* bien attestée par le passé¹⁵, est à présent obsolète, et ni la rection indirecte (*être prévenu de*), ni la forme active ne sont plus du tout attestées en contexte juridique (*prévenir qqn d'un délit*¹⁶, *prévenir qqn de vol*) ; la forme passive commence à resurgir, dans des textes de jurisprudence, jamais toutefois avec le complément d'objet indirect *de* + [nom d'infraction]¹⁷ : *personne prévenue pour un délit* apparaît à 4 reprises (statistiques Google mai 2024), dans des commentaires de décisions du Conseil constitutionnel de 2020, sur des QPC (questions prioritaires de constitutionnalité) traitant des conditions de détention provisoire. Il s'agit en fait de la même note de bas de page, dans les quatre textes concernés :

(8) En dehors de l'instruction, une *personne prévenue pour un délit* peut également être placée en détention provisoire dans certaines situations, dans l'attente de sa comparution devant le tribunal correctionnel¹⁸.

¹⁵ À titre d'illustration : « [être] prévenu de délits de police, ou de crimes simplement infamants » (CP_{FR} 1810, Art. 238).

¹⁶ Seule interprétation de possible : l'en informer par avance (usage de langue générale, non juridique).

¹⁷ Comparer au verbe *accuser*, qui réalise encore, lui, ce type d'emploi : *être accusé d'un crime*, *être accusé de vol à main armée*, y compris à la forme active (*accuser qqn d'un crime/de vol à main armée*).

¹⁸ Commentaire de la Décision n° 2020-858/859 du Conseil Constitutionnel, relative à une QPC (question prioritaire de constitutionnalité) sur les « conditions d'incarcération des détenus prévenus », note 3. URL : <www.conseil-constitutionnel.fr/sites/default/files/as/root/bank_mm/decisions/2020858_859qpc/2020858_859qpc_ccc.pdf> (consulté le 19-11-2025).

Ce n'est pourtant pas un hapax, la construction *prévenu-es pour* étant même attestée dans un article du CPP_{FR} 1959, en vigueur depuis 2010 (dix ans plus tôt donc) :

(9) Les *détenus* qui sont *prévenus pour* une cause et *condamnés pour* une autre doivent être soumis au même régime que *les condamnés* [...] ils peuvent être détenus dans des *établissements pour peines*. (CPP_{FR} 1959, D52)

L'existence d'antonymes distincts pour les deux domaines thématiques (procédure pénale et droit pénitentiaire respectivement) est un argument classique pour un diagnostic de polysémie (ou plutôt, pour l'existence de sens distincts¹⁹). La traduction (interlinguale) différente des deux types d'occurrences en est un autre.

En effet, le terme français *prévenu* (*vs accusé*) est rendu en roumain par *inculpat* (vectorisation : vers le roumain, degré d'équivalence : plus étendu, puisque *inculpatul* – « l'inculpé » – correspond en droit roumain contemporain à la fois au *prévenu* et à l'*accusé* du droit pénal français²⁰) ; par contre, le *prévenu* (*vs détenu condamné*) sera rendu en roumain par *arestatul preventiv* (litt. « l'arrêté préventif » – équivalence bidirectionnelle et exacte).

5. Première hypothèse : emprunt transdisciplinaire sous extension de sens

L'emprunt transdisciplinaire (procédé consistant à attribuer un terme d'un domaine à un concept nouveau d'un autre domaine) est « souvent²¹ » légitimé par une relation d'analogie entre compréhensions des concepts source et cible (ISO 704 : 2009 (F), §B 3.5). Les glissements de sens métonymiques, généralisants ou particularisants seraient typiquement à l'origine de polysémies intradisciplinaires (*ibid.*, §B 3.4). Mais rien n'interdit le transfert d'un mot d'un domaine à l'autre et le réinvestissement conceptuel non pas analogique mais généralisant, particularisant, en passage du concret à l'abstrait, etc.

Le terme de procédure pénale *prévenu*, qui désigne un statut procédural spécifique en attente du jugement, a l'air d'avoir été étendu, en droit pénitentiaire, à tous les trois statuts possibles de la personne placée en détention provisoire, en deçà de l'incarcération

¹⁹ Car il ne préjuge en rien de l'existence d'un quelconque lien dérivationnel entre les sens ainsi discriminés. Nous partons ici d'une définition intuitive de la polysémie comme « sens différents, mais apparentés » (Kleiber, 1999 : 54).

²⁰ Avec la disparition, en droit pénal roumain contemporain, à la fois des *curți cu juri* (cours à jurés – équivalent interculturel des cours d'assises françaises) et de la distinction entre espèces d'infractions selon leur gravité, la valeur exacte des termes *prévenus* et *accusés* est devenue « incommensurable » en roumain juridique. Au sens de Kuhn (2000), deux théories/langages sont « incommensurables » si leurs structures sont non homologues. Pour commentaire et application à la problématique linguistique, cf. Grimaldi, 2018 : 54-55. La notion d'« anisomorphisme » (Zgusta, 1971 : 294-296) rend compte de plus ou moins les mêmes données, concernant des langues en contact (en lexicographie bilingue).

²¹ Mais pas systématiquement.

(donc y compris aux accusés et aux mis en examen). Sommes-nous en présence d'un élargissement de sens canonique ?

Non, car la dérivation du nouveau sens ne repose pas uniquement sur l'effacement de sèmes dans le terme emprunté²². La dérivation du sens pénitentiaire semble, au contraire, requérir l'introduction d'un genre prochain autre. Le prévenu des maisons d'arrêt est une personne incarcérée non condamnée, tandis que les mis en examen, prévenus et accusés de la procédure pénale sont, en deçà de l'incarcération, des personnes poursuivies non condamnées (plutôt : en attente du jugement : condamnation ou relaxe/acquittement). Tous les prévenus (*vs* accusés) ne sont pas des prévenus (*vs* détenus condamnés²³), ni tous les prévenus (*vs* détenus condamnés) ne sont des prévenus (*vs* accusés).

Seraient-ce les caractères que les deux concepts n'ont pas en commun qui en expliquent les corrélations extensionnelles seulement « contingentes » (*vs* systématiques) ?

- a. Chronologie procédurale : [personne renvoyée devant une juridiction répressive] ∈ prévenu (*vs*²⁴ accusé) ;
- b. Type d'infraction : [pour un délit]/[pour une contravention] (*vs* [pour un crime]) ∈ prévenu (*vs* accusé) ;
- c. Type de juridiction répressive : [tribunal correctionnel]/[tribunal de police] (*vs* [cour criminelle]/[cour d'assises]) ∈ prévenu (*vs* accusé)
- d. Type de local d'incarcération : [maison d'arrêt] (*vs* [prison] *vs* [local de police]) ∈ prévenu (*vs* personne détenue condamnée *vs* personne gardée à vue)

Le fait est que les caractères que les deux concepts ont en commun (intersection sémantique) semblent différents et par leur saillance relative (les caractères posés étant plus saillants cognitivement que ceux qui sont simplement présupposés ou bien impliqués) et par leur statut (caractères distinctifs *vs* non distinctifs, hérités du genre prochain ou, par son intermédiaire, de genres plus hauts dans la hiérarchie), ce qui n'est pas sans obscurcir le sentiment de « parenté » :

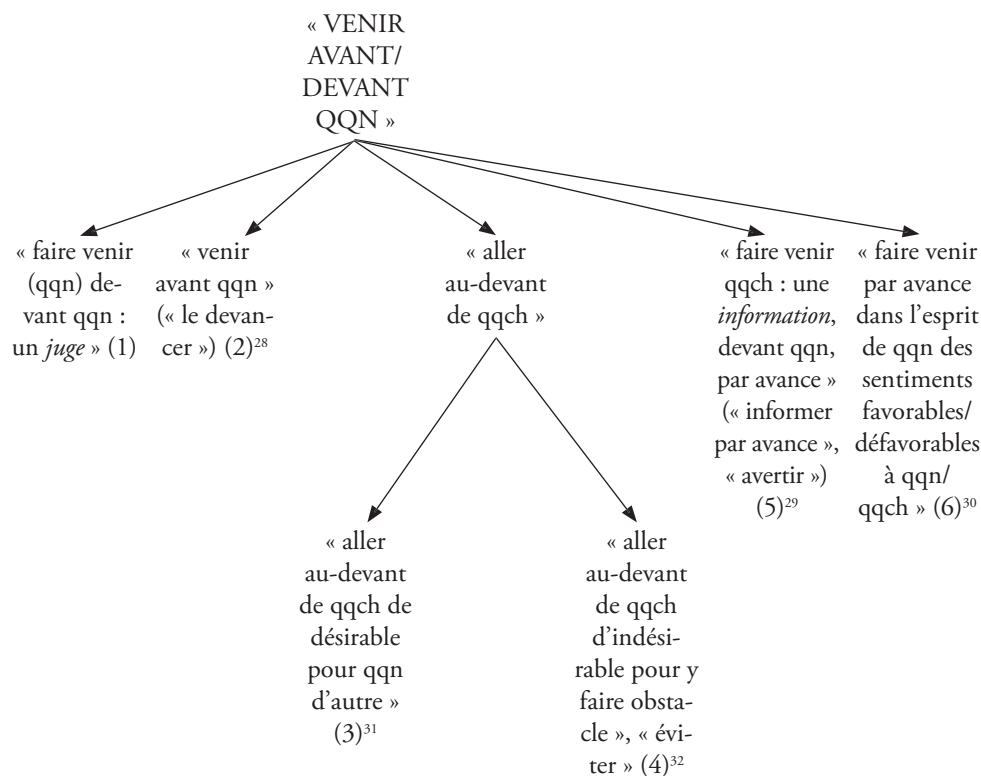
- a. [pas encore jugé] (ou : [pas encore condamné]) : principal caractère distinctif (un caractère posé donc) des deux espèces de *personnes incarcérées* (les *prévenus* et les *condamnés*), mais caractère partagé seulement impliqué pour les *prévenus* (*vs* *accusés*) – le fait d'être ou de ne pas être jugé (ou condamné) à un moment donné d'une procédure judiciaire étant pertinent pour des *personnes poursuivies* mais pas sinon (les citoyens n'ayant pas maille à partir avec la justice n'en sont pas concernés) ;
- b. [personne poursuivie pour une infraction] : caractère partagé hérité d'un genre superordonné (un caractère posé donc), pour les *prévenus* (*vs* *accusés*, *vs* *mis en*

²² Effacement de caractères distinctifs du concept désigné : panier qui en vient à désigner toute corbeille, à force de perdre son sème associatif « contenir du pain ».

²³ Ils peuvent faire l'objet d'un contrôle judiciaire, d'un port de bracelet électronique ou ne faire l'objet d'aucune limitation de libertés tant que présumés innocents.

²⁴ Prévenus et accusés sont alors au même titre des personnes renvoyées devant des juridictions répressives, mais ici nous avons noté « prévenu (*vs* accusé) » le sens pénal (*vs* pénitentiaire) du terme polysémique à l'étude.

Figure 1 - Polysémie du verbe prévenir (1) à (6) par ordre d'attestation (cf. TLFi)



Les objectifs de la détention provisoire³³ (dite par le passé *détention préventive*³⁴) sont tous formulés en termes d'évitement d'une action indésirable pour la suite de la procédure pénale : empêcher des pressions sur les témoins, les victimes ou leurs familles, empêcher toute concertation frauduleuse avec des coauteurs ou des complices, éviter que les preuves ou indices ne soient compromis, etc³⁵. Le nom (ou l'adjectif) *prévenu* (« qui est placé en détention provisoire ») serait donc à ramener au sens « éviter », plutôt qu'au sens « citer en justice ». La distance entre *citer en justice* et *éviter* est trop conséquente à la fois historiquement (dates des premières attestations) et sémantiquement (structure sémique et type syntaxique/sélection d'arguments) pour que subsiste une quelconque in-

²⁸ X est arrivé bien tard, Y l'a prévenu de plus d'une heure [Y est arrivé plus d'une heure auparavant]. Y voulait rendre service à X, mais Z l'a prévenu [Z a rendu ce service à X avant que Y ne le fasse] ».

²⁹ « Prévenir Y de l'arrivée de X ».

³⁰ « Prévenir (qqn) en faveur ou contre l'accusé, être prévenu en faveur contre l'accusé ».

³¹ « Prévenir les désirs/ordres/questions de qqn. ».

³² « Prévenir une catastrophe ».

³³ Vœu pieux de l'infracteur présumé (terme « politiquement correct »).

³⁴ Terme moins hypocrite (point de vue de l'autorité).

³⁵ Cf. <www.service-public.fr/particuliers/vosdroits/F36506/0?idFicheParent=F1042> (consulté le 19-11-2025).

tersection sémique entre *prévenu* cité (« personne renvoyée devant le juge pour un délit ou une contravention »), d'une part, et *prévenu* en maison d'arrêt, de l'autre (« personne placée en détention provisoire pour empêcher ses mauvais agissements éventuels »).

Deux processus autonomes de conversion, à partir d'acceptions du verbe polysémique source non seulement distinctes, mais sans rapport direct aucun devraient aboutir à deux termes distincts³⁶.

7. Retour sur les détenus condamnés et conclusion

Et si, après tout, les détenus condamnés avaient pour pendant non pas des prévenus, tout court, mais des détenus prévenus ?

Nous sommes passée à côté de la vraie portée d'exemples tel (9) *supra* (« les détenus qui sont *prévenus* pour une cause et *condamnés* pour une autre »), et telle la QPC sur les « conditions d'incarcération des *détenus prévenus* », évoquée en note 18 simplement comme référence du verbe-terme (*être*) *prévenu* pour (illustré sous (8)).

L'emploi pénitentiaire du mot *prévenu* se laisserait alors, après tout, analyser selon le patron habituel des polysémies de transfert généralisantes : d'abord, *prévenu* [adj. part. passé (passif)] perdrait des sèmes, au point de n'en garder que « (qui est) poursuivi pour une infraction » et « (qui est) en attente du jugement ». Ensuite, il y aurait composition syntaxique de *détenu* (hyperonyme) et de *prévenu* (adjectif allégé de ses caractères distinctifs quant au type d'infraction et de juridiction), avec perte de saillance relative du trait générique « poursuivi pour une infraction », qui se laisse de toute façon inférer à partir de *détenu* (encyclopédie). Un « trait sémantique capital » dans l'environnement *prévenu* (*pour*) (*vs* *accusé* (*de*)) deviendra « trait subordonné » dans *détenu prévenu*. Résultat de ce processus compositionnel : « personne incarcérée [$\supset$ poursuivie pour une infraction] & qui est en attente du jugement ».

L'extension de sens que subit l'adjectif déverbal *prévenu* dans l'environnement *détenu* relèvera d'un emprunt transdisciplinaire standard, qui repose sur la « circulation de concepts » d'une discipline à l'autre, et sur de fécondes « interférences » (Morin, 1994 ; Velicu, 2017) et aboutit à l'émergence d'un nouveau concept (ici, plus étendu) dans le contexte disciplinaire d'accueil (polysémie interdisciplinaire de l'adjectif déverbal).

Mais le réinvestissement conceptuel du nom « prévenu » en droit pénitentiaire sera redevable d'une méthode distincte de création terminologique. Le nom *prévenu* employé en lieu et place de *détenu prévenu* sera le résultat de processus d'ellipse suffisamment fréquents en discours, pour que se produise l'absorption du sens « personne incarcérée ». *Prévenu* (n.m.) sera alors une « forme raccourcie » caractérisée, au sens de la norme ISO 704 : 2000, « forme réduite » selon la version 2009 de la même norme, « forme courte » dans la version 2022. Les « formes raccourcies » des termes complexes ne coïncident

³⁶ Le *TLFi* traite pourtant *marche de l'escalier* et *ralentir sa marche* comme sens d'un même vocable polysémique, bien que la première provienne de l'emploi transitif indirect (*marcher sur*) et le second, de l'emploi inergatif du verbe base (*marcher vite*).

jamais avec l'hyperonyme : ce sont typiquement des signifiants de caractères distinctifs qui refont surface.

Le point crucial, dans cette analyse, est que la polysémie de transfert (par extension de sens) ne concerne plus directement le nom (terme simple), mais l'adjectif déverbal (ou le verbe passif lui-même), de sorte que la génération d'un nouveau sens du nom substantif ne remette pas en cause le genre prochain : le caractère seulement contingent de l'association entre les « personnes poursuivies pour un délit, en attente du jugement » et les « personnes incarcérées, en attente du jugement » n'a plus à être érigé en mécanisme sémantogène *ad hoc* si le premier concept n'est pas à l'origine du second.

À l'instar du levier proverbial d'Archimède, en linguistique (et en terminologie descriptive), un bon exemple peut renverser une théorie. Le raisonnement aboutissant à l'impasse dans l'analyse de l'extension de sens du terme *prévenu*, en droit pénitentiaire reposait sur l'étude d'exemples où le nom *prévenu* était directement opposé au nom *condamné* ou au syntagme *détenu(s) condamné(s)*. Ces exemples imposaient le changement d'hyperonyme (*personne incarcérée* ou (*personne*) *détenu(e) vs personne poursuivie*) et, partant, emphatisaient les divergences de structuration sémique entre configuration de départ et configuration d'arrivée.

Rien de tel si nous prenons pour « point d'appui » non pas *prévenu vs (détenu) condamné*, mais *détenu (qui est) prévenu (pour) vs détenu (qui est) condamné (pour)*. L'alternance unité phraséologique/terme complexe/terme simple caractérise la langue du droit, en particulier dans le paradigme des sujets passifs de la procédure pénale (*personne mise en cause (dans une affaire pénale/ dans une infraction)/le mis en cause, personne mise en examen (dans l'affaire X)/le mis en examen, personne accusée (d'un crime)/l'accusé, personne prévenue (pour un délit)/le prévenu*). Pourquoi ne serait-elle pas instanciée aussi en droit pénitentiaire (*détenu prévenu (pour un délit)/le prévenu, détenu condamné (pour un délit)/le condamné*) ?

Le grand mérite de l'analyse fondée sur le corpus est qu'elle permet au linguiste/terminologue d'outiller des recherches ciblées correspondant à des « exemples construits » imaginés, « forts », et susceptibles d'étayer (ou de démonter) l'hypothèse à l'étude. À la différence de l'analyse guidée par le corpus, qui enregistre les données fournies et en tire des conclusions éventuellement généralisables à d'autres occurrences, l'analyse basée sur le corpus fédère à la même enseigne fantôme de l'exemple construit et recours à l'intuition de sujet parlant du linguiste, d'une part, et objectivité des données attestées, de l'autre, se prêtant à la fois à une approche déductive et à une approche inductive, à une approche onomasiologique et à une approche sémasiologique. En tant que telle, elle répond, croyons-nous, à la définition même de l'interdisciplinarité comme partage de méthodes plutôt que seulement de concepts (Nicolesco, 1996 : XXVII ; 2011 : 96). Nous avons essayé ici, à propos des prévenus en tous genres du droit pénal français, d'en explorer les limites (et d'en illustrer, nous l'espérons, le potentiel).

Références bibliographiques

- Dury P., Picton A., *Terminologie et diachronie : vers une réconciliation théorique et méthodologique ?*, in « Revue Française de Linguistique Appliquée », XIV-2, 2009, pp. 31-41.
- Grimaldi C., *Discours et terminologie dans la presse scientifique française (1699-1740). La construction des lexiques de la botanique et de la chimie*, Oxford, Peter Lang, 2017.
- Grimaldi C., *Explorer la langue en diachronie : lexicographie et grammaire dans l'enseignement des langues de spécialité*, in « ELA. Études de linguistique appliquée », 189, 2018/1, pp. 49-61.
- Jollivet M., Legay J.-M., *Canevas pour une réflexion sur une interdisciplinarité entre sciences de la nature et sciences sociales*, in « Natures Sciences Sociétés », 13/2, 2005, pp. 184-188.
- Kleiber G., *Chapitre II. Une "polysémie" entrée en matière*, in G. Kleiber, *Problèmes de sémantique*, Villeneuve d'Ascq, Presses universitaires du Septentrion, 1999. URL : <<https://doi.org/10.4000/books.septentrion.116728>>, consulté le 19-11-2025.
- Kuhn T., *The Road Since Structure*, J. Conant, J. Haugeland (eds.), Chicago, University of Chicago Press, 2000.
- Morin E., *Sur l'interdisciplinarité*, in « Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires », 2, 1994. URL : <<https://ciret-transdisciplinarity.org/bulletin/b2c2.php>>, consulté le 19-11-2025.
- Nicolesco B., *La Transdisciplinarité. Manifeste*, Perpignan, Éditions du Rocher, 1996.
- Nicolesco B., *De l'interdisciplinarité à la transdisciplinarité : fondation méthodologique du dialogue entre les sciences humaines et les sciences exactes*, in « Nouvelles perspectives en sciences sociales », 7(1), 2011, pp. 89-103.
- Plawski S., *Droit pénitentiaire*, Lille, Presses Universitaires du Septentrion, 1976.
- Polguère A., *Lexicologie et sémantique lexicale. Notions fondamentales*, Montréal, Presses de l'Université de Montréal, 2008.
- Simonet G., *Le concept d'adaptation : polysémie interdisciplinaire et implication pour les changements climatiques*, in « Natures Sciences Sociétés », 17-4, 2009, pp. 392-401.
- Van Campenhoudt M., *Réseau notionnel, intelligence artificielle et équivalence en terminologie multilingue : essai de modélisation*, in A. Clas et al. (éds.), *Lexicomatique et dictionnaires*, Montréal, AUPÉLF-UREF et Beyrouth, F.M.A., 1996, pp. 281-306.
- Velicu A., *Les infractions à la française et à la roumaine – réaménagements conceptuels et normalisation terminologique*, in M.T. Cabré, M. de Blas (éds.), *Terminologie pour la normalisation et terminologie pour l'internationalisation*, Barcelona, Institut d'Estudis Catalans TERMCAT/ Universitat Politècnica de Catalunya, 2017, pp. 41-49.
- Wüster E., *Einführung in die allgemeine Terminologielehre und terminologische Lexicographie*, Copenhagen, 1985 (Wien, Springer, 1979). URL : <<https://digitalia.lib.pt.ehu/hu/wuster-eugen-einfuehrung-in-die-allgemeine-terminologielehre-1-fzh-kopenhagen-1985-4333#page/28/mode/1up>>, consulté le 19-11-2025.
- Wüster E., *Die Allgemeine Terminologielehre – ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften*, in H. Picht, C. Laurén (eds.), *Ausgewählte Texte zur Terminologie*, Wien, Termnet, 1993, pp. 331-376.
- Zanola M.T., *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*, Paris, L'Harmattan, 2014.
- Zgusta L., *Manual of Lexicography*, Prague, Academia ; The Hague, Mouton, 1971.

Corpus (consulté le 19-11-2025)

- Code pénal de la République française*. URL : <www.legifrance.gouv.fr/Codes/texte_lc/LEGITEXT000006070719/2024-01-15/> [CP_{FR} 1994]
- Code de procédure pénale de la République française*. URL : <www.legifrance.gouv.fr/Codes/texte_lc/LEGITEXT000006071154/> [CPP_{FR} 1959].
- Code pénitentiaire de la République française* (2023), Partie législative. URL : <www.legifrance.gouv.fr/Codes/section_lc/LEGITEXT000045476241/LEGISCTA000045478145/2023-12-01/#LEGISCTA000045480638> [CPénit_{FR} 2022].
- Code d'instruction criminelle de 1808*. URL : <https://ledroitcriminel.fr/la_legislation_criminelle/anciens_textes/Code_instruction_criminelle_1808.htm> [CIC_{FR} 1808³⁷].
- Code d'instruction criminelle de 1808 (Texte intégral de la version en vigueur en 1929)*. URL : <https://ledroitcriminel.fr/la_legislation_criminelle/anciens_textes/Code_instruction_criminelle_1929/Code_1808.htm> [CIC_{FR} 1808/1929].
- Code pénal*. Paris, Imprimerie impériale, 1810. URL : <www.ledroitcriminel.fr/la_legislation_criminelle/anciens_textes/Code_penal_1810/Code_penal_1810_1.htm> [CP_{FR} 1810].
- Codul Penal din 1936 [Codul Penal "Carol al II-lea"]*. URL : <<https://lege5.ro/App/Document/heztqnzu/codul-penal-din-1936>> [CP_{RO} 1936].
- Codul de Procedură Penală din 17 martie 1936*. URL : <<https://legislatie.just.ro/Public/DetaliiDocumentAfis/51252>> [CPP_{RO} 1936].
- Codul Penal din 21 iunie 1968 (*republicat*)*. *Monitorul Oficial*, nr. 65 din 16 aprilie 1997. URL : <<https://legislatie.just.ro/Public/DetaliiDocument/137244>> [CP_{RO} 1968/1997].
- Codul de Procedură Penală din 1 ianuarie 1968 (**republicat**)*. *Monitorul Oficial*, nr. 78 din 30 aprilie 1997. URL : <<https://legislatie.just.ro/Public/DetaliiDocumentAfis/193>> [CPP_{RO} 1968/1997].
- Codul Penal din 17 iulie 2009*. URL : <<https://legislatie.just.ro/public/detaliiidocument/109855>> [CP_{RO} 2009].
- Noul Cod de Procedură Penală din 2010*. URL : <<https://legislatie.just.ro/Public/DetaliiDocument/120611>> [CPP_{RO} 2010].
- Ordonanța nr. 2/2001 privind regimul juridic al contravențiilor*. URL : <<https://legislatie.just.ro/Public/DetaliiDocumentAfis/254014>> [O.G. 2/2001].

³⁷ Année d'entrée en vigueur.

Revue internationale qui s'intéresse aux terminologies spécialisées dans leurs relations avec la dimension communicative et discursive (monolingue et plurilingue), *TermCD – TERMinologie, Communication et Discours* suit une perspective diachronique et synchronique et vise à constituer un lieu de rencontre et de débat autour de la terminologie, qui représente une voie d'accès privilégiée aux savoirs spécialisés et permet d'en reconstruire l'histoire et leur évolution dans le temps.

Rivista internazionale i cui interessi riguardano le terminologie specialistiche e le loro relazioni con la dimensione comunicativa e discorsiva (monolingue e plurilingue), *TermCD – TERMinologie, Communication et Discours* adotta una prospettiva di studio tanto diacronica quanto sincronica, intendendo porsi quale luogo di confronto e dibattito sulla terminologia, che rappresenta una via di accesso privilegiata ai saperi tecnico-scientifici e permette di ricostruirne la storia e l'evoluzione nel tempo.

International journal whose interests concern specialized terminologies and their relationship with the communicative and discursive dimension (monolingual and multilingual), *TermCD – TERMinologie, Communication et Discours* follows a diachronic and synchronic perspective and aims to constitute a place of meeting and debate around terminology, which represents a privileged access to specialized knowledge and makes it possible to reconstruct its history and its evolution over time.

EDUCatt - Ente per il Diritto allo Studio Universitario dell'Università Cattolica

Largo Gemelli 1, 20123 Milano - tel. 02.72342235 - fax 02.80.53.215

Largo Gemelli 1, 20123 Milano | tel. 02.7234.2235 | fax 02.80.53.215

e-mail: editoriale.dsu@educatt.it (produzione); librario.dsu@educatt.it (distribuzione)

web: libri.educatt.online

ISBN: 979-12-5535-581-6

ISSN digitale: 3034-8668



euro 13,00