

li290

Linguistic Insights

Studies in Language and Communication

Federica Vezzani

**Terminologie numérique :
conception, représentation
et gestion**

Peter Lang

Cet ouvrage se consacre à la notion de terminologie numérique considérée comme une approche de la discipline impliquant la représentation numérique d'informations conceptuelles et linguistiques d'un domaine spécifique. L'objectif est l'illustration des étapes de conception et d'implémentation de base de données terminologiques multilingues permettant le respect des meilleures pratiques dans la gestion des données terminologiques du numérique. Pour ce faire, l'ouvrage met en exergue les nouvelles compétences du terminologue à l'ère numérique. Celles-ci trouvent leur véritable essence dans l'esprit interdisciplinaire et collaboratif de la recherche.

Federica Vezzani est chercheuse à l'Université de Padoue. Ses intérêts scientifiques concernent la terminographie et la mise en place des bases de données terminologiques multilingues conformément aux principes FAIR (Findability, Accessibility, Interoperability, Reusability) de la science ouverte.

Terminologie numérique :
conception, représentation et gestion



Linguistic Insights

Studies in Language and Communication

Edited by Maurizio Gotti,
University of Bergamo

Volume 290

ADVISORY BOARD

Vijay Bhatia (Hong Kong)
David Crystal (Bangor)
Konrad Ehlich (Berlin / München)
Jan Engberg (Aarhus)
Norman Fairclough (Lancaster)
John Flowerdew (Hong Kong)
Ken Hyland (East Anglia)
Roger Lass (Cape Town)
Françoise Salager-Meyer (Mérida, Venezuela)
Srikant Sarangi (Cardiff)
Susan Šarčević (Rijeka)
Lawrence Solan (New York)



PETER LANG

Bern · Berlin · Bruxelles · New York · Oxford

Federica Vezzani

Terminologie numérique : conception, représentation et gestion



PETER LANG

Bern · Berlin · Bruxelles · New York · Oxford

Information bibliographique publiée par « Die Deutsche Nationalbibliothek » « Die Deutsche Nationalbibliothek » répertorie cette publication dans la « Deutsche Nationalbibliografie » ; les données bibliographiques détaillées sont disponibles sur Internet sous <http://dnb.d-nb.de>.

Cette publication a été entièrement financée par le projet ExaMode, dans le cadre du programme H2020 de l'Union européenne, sous la convention de subvention n° 825292.

ISSN 1424-8689

ISBN 978-3-0343-4394-7 (Print)

E-ISBN 978-3-0343-4264-3 (E-PDF)

E-ISBN 978-3-0343-4431-9 (EPUB)

DOI 10.3726/b19407

© Federica Vezzani, 2022

bern@peterlang.com, www.peterlang.com

PETER LANG



Open Access: Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0. Pour consulter une copie de cette licence, visitez le site internet <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Table des matières

Liste des figures	9
Liste des tableaux	11
Préface	13
Introduction	17
1 Numérisation de la terminologie	23
1.1 Pour une métaterminologie	24
1.1.1 Terminologie, terminographie et terminotique	24
1.1.2 Terminologie computationnelle	27
1.1.3 Terminologie numérique	32
1.2 Curation des données numériques	34
1.2.1 Principes FAIR	35
1.2.2 Principes des données liées	38
1.2.3 Données linguistiques liées ouvertes	40
1.3 Conclusion	42
2 Fondements pour une terminologie numérique	45
2.1 Théories et approches en terminologie	45
2.2 Travail terminologique numérique	54
2.2.1 Dimension conceptuelle	55
2.2.2 Dimension linguistique	57
2.3 Conclusion	58
3 Conception d'une ressource terminologique	61
3.1 Principes FAIR en terminologie	62
3.1.1 Terminologie trouvable	63
3.1.2 Terminologie accessible	64
3.1.3 Terminologie interopérable	66
3.1.4 Terminologie réutilisable	67

3.2	Paradigme de la « terminologie FAIR »	68
3.3	Modélisation conceptuelle de la ressource	69
3.3.1	Schéma entité-association	70
3.3.2	Métamodèle structurel <i>Terminological Markup Framework</i>	73
3.3.3	Catégories de données	78
3.4	Conclusion	82
4	Vers l'implémentation d'une ressource terminologique	83
4.1	Représentation des données terminologiques	84
4.1.1	Format <i>TermBase eXchange</i>	86
4.1.1.1	Dialectes TBX	87
4.1.1.2	Conception d'un dialecte	88
4.2	Actualisation des données terminologiques	93
4.3	Publication des données terminologiques	96
4.4	Conclusion	98
5	Étude de cas : la ressource TriMED	99
5.1	Langue médicale	100
5.1.1	Perspective intra-linguistique	101
5.1.2	Perspective inter-linguistique	104
5.2	Analyse des exigences	108
5.3	Modèle structurel de TriMED	110
5.4	Fiche terminologique	113
5.4.1	Catégories de type terminologique	115
5.4.2	Catégories de type conceptuel	120
5.4.3	Catégories de type administratif	123
5.5	Répertoire des catégories de TriMED	125
5.6	Mise en œuvre TBX	128
5.6.1	Formulation du dialecte TBX-TriMED	129
5.6.2	Position des catégories de données	138
5.7	Application Web TriMED	141
5.7.1	Utilisateurs	144
5.7.1.1	Le patient	145
5.7.1.2	Le traducteur	146
5.7.1.3	Le médecin	149

5.8 Application Web RCD	151
5.9 Collection terminologique	153
5.10 Conclusion	158
6 <i>Is it FAIR enough?</i>	159
6.1 <i>Resource Description Framework – RDF</i>	161
6.2 De TBX à RDF	163
6.3 Conclusion	165
Annexe A	167
Annexe B	187
Bibliographie	191
Index	211

Liste des figures

Image 1.1	Nuage LLOD, Juillet 2020	41
Image 2.1	Le double triangle sémiotique	53
Image 3.1	Exemple schéma entité-association	72
Image 3.2	Le métamodèle TMF	76
Image 3.3	Exemple de catégorie de données sur DatCatInfo	80
Image 4.1	Exemple de TBXMD du module <i>Multilang</i>	90
Image 5.1	Schéma entité-association de TriMED	112
Image 5.2	Identifiants dans TriMED	124
Image 5.3	Répertoire des catégories de TriMED	126
Image 5.4	Formalisme partiel du module <i>Trimed</i>	131
Image 5.5	Position des catégories dans TriMED	139
Image 5.6	Page d'accueil TriMED	142
Image 5.7	À propos de TriMED	143
Image 5.8	Interface « Patient »	145
Image 5.9	Interface « Traducteur »	147
Image 5.10	Panneau « Usage » fiche bilingue <i>Scintigraphy</i> – <i>Scintigrafia</i>	148
Image 5.11	Affichage de la fiche en format TBX	149
Image 5.12	Interface « Médecin »	151
Image 5.13	Interface RCD : catégorie de données /analyse sémique/	152
Image 6.1	Exemple de graphe RDF	162
Image 6.2	Formalisme complet du module <i>Trimed</i>	187

Liste des tableaux

Tableau 5.1	Catégories de type terminologique	115
Tableau 5.2	Catégories de type conceptuel	120
Tableau 5.3	Catégories de type administratif	123
Tableau 5.4	Exemples de termes dans TriMED classés par sphères conceptuelles	155
Tableau 6.1	Spécifications des catégories de données en TriMED	167

Préface

Les normes au service de la terminologie numérique

Ce livre intitulé « Terminologie numérique : conception, représentation et gestion » est innovateur dans sa conception et dans son but. Cela le rend indispensable pour tout terminologue provenant des sciences sociales et humaines dont l'objectif est de s'aventurer dans l'une des facettes du travail terminologique consistant à créer, à gérer et à mettre à jour des ressources terminologiques. Ces dernières sont des conteneurs de données terminologiques collectées de façon systématique d'après des principes théoriques et méthodologiques de la science de la terminologie.

Le terminologue d'aujourd'hui opère nécessairement dans le numérique, d'où l'importance qu'assume son activité dès lors qu'il s'agit de partager les données sur lesquelles il travaille. Ces données doivent être interchangeables et interopérables pour correspondre aux politiques scientifiques qui visent à accélérer le développement vers la Science Ouverte. C'est dans cet esprit que Federica Vezzani se propose d'implémenter les principes FAIR (**F**aciles à trouver, **A**ccessibles, **I**nteropérables et **R**éutilisables) à la conception et la construction des ressources terminologiques selon trois normes élaborées au sein de l'ISO SC3, CT 37 :

- i. ISO-16642 : 2017 – Applications informatiques en terminologie – Plate-forme pour le balisage de terminologies informatisées ;
- ii. ISO-12620 : 2019 – Gestion des ressources terminologiques – Spécifications des catégories de données ;
- iii. ISO-30042 : 2019 – Gestion des ressources terminologiques – TermBase eXchange (TBX).

Ces trois documents fournissent les lignes directrices et les exigences nécessaires pour structurer et représenter les données afin de permettre

l'interchangeabilité et leur réutilisation. Le défi majeur du travail terminologique est bel et bien la possibilité de réutiliser les données qui se doivent accompagnées de descriptions standardisées.

À ce stade Federica Vezzani, rejoignant la tendance de la Science Ouverte, propose un nouveau paradigme qu'elle désigne de « terminologie FAIR », où elle dessine une méthodologie innovatrice qui associe les principes FAIR aux principes de gestion de la terminologie représentés dans les normes ISO supra mentionnées.

Finalement, l'auteure de l'ouvrage applique son paradigme à une ressource terminologique médicale polyvalente – TriMED – qui se dirige à des publics qui ont des besoins de communication spécifiques dans des contextes médicaux – patients, médecins, traducteurs. Du point de vue de sa structure, la fiche terminologique de TriMED correspond aux requis normalisateurs. Du point de vue du contenu, on peut y trouver une quantité diverse d'informations concernant le comportement linguistique des termes et du vocabulaire spécialisé. Enfin, une application Web est construite qui permet la consultation, mais aussi la réutilisation des données terminologiques multilingues.

La perspective théorique et l'approche méthodologique de l'auteure de l'ouvrage sont ancrées dans une terminologie à dimension conceptuelle et linguistique dont les principes se reflètent dans l'usage qu'elle fait des normes et dans le dessin de TriMED. L'approche de Federica Vezzani se veut multidisciplinaire et appelle au besoin de constituer des équipes mixtes pour mener à bon terme un travail terminologique qui actuellement est nécessairement numérique. Ce livre a le grand mérite d'assumer un point de vue qui l'accompagne l'auteure au fil des pages.

Dans le chapitre premier « Numérisation de la terminologie », l'auteure fait une approche historique au concept de « numérique » dans le contexte du travail terminologique, ce qui l'amène à reprendre des termes comme « terminographie », « terminotique » et « terminologie computationnelle » en faisant une incursion aux spécificités de chacune de ces approches. Elle passe ensuite aux questions actuelles concernant le numérique en s'attardant sur l'avènement du Web sémantique e du *Linked Open Data* (LOD).

Avec le deuxième chapitre intitulé « Fondements pour une terminologie numérique », Federica Vezzani fait le tour des théories, selon elle

les plus proéminentes, ainsi que de diverses approches appliquées au travail terminologique où elle déclare sa préférence pour une approche à la terminologie qui est bidimensionnelle et qui prend tout sens, vue la place accordée aux normes ISO dans son étude.

Le troisième chapitre est consacré à la « Conception d'une ressource terminologique » où l'auteure élabore la proposition d'une « terminologie FAIR », ce qui l'induit à décrire deux des normes, à savoir, l'ISO-16642 : 2017 qui permet la modélisation et l'organisation des entités « concept », « terme » et « langue » (au sens que la norme leurs donne) et l'ISO-12620 : 2019 qui, à son tour, permet de représenter et d'harmoniser les propriétés (catégories de données) choisies pour chaque entité.

La troisième norme ISO évoquée supra est traitée dans le quatrième chapitre intitulé « Vers l'implémentation d'une ressource terminologique ». Ici, Federica Vezzani se concentre sur le format de représentation TermBase eXchange (TBX) (ISO-30042, 2019) qu'elle considère être le modèle à suivre pour l'adoption d'une approche FAIR à la terminologie. Elle termine ce chapitre, en proposant des pistes de réflexion sur la mise à jour et la publication des données terminologiques contenues dans une ressource FAIR.

L'« Étude de cas : la ressource TriMED » est décrite dans le cinquième chapitre et concerne la conception et la mise en œuvre de la base de données TriMED. C'est dans ce chapitre que l'auteure explique toutes les spécificités détaillées de la ressource TriMED dont le but est de démontrer que la méthodologie utilisée est répliquable à tout autre domaine. Le détail de spécificité du dessin de la ressource fait de ce chapitre un texte très didactique qui pourra très facilement se transformer en un guide pour tout novice de la terminologie numérique.

Le sixième chapitre dont le titre est « Is it FAIR enough? » nous renvoie à une sorte de palimpseste construit sur un jeu de mots provocateur. Ce chapitre est une réflexion critique aux aspects de la FAIRisation. L'auteur considère que la démarche dessinée est un bon point de départ pour aller plus loin, car la méthodologie proposée doit être mise à jour et constitue le levier pour d'autres envols tenant en compte d'autres normes tels que RDF ou Ontolex-Lemon (une norme *de facto*) – rien que pour en nommer quelques-unes – pour avancer vers

l'interopérabilité entendue comme la capacité que les systèmes existants ou à créer ont à s'adapter et à interagir entre eux.

Rute Costa

Faculdade de Ciências Sociais e Humanas
Universidade NOVA de Lisboa
NOVA CLUNL, Lisbonne

Introduction

Cet ouvrage se consacre à la question de l'évolution de la discipline de la terminologie à l'ère numérique. La terminologie est la science qui traite de l'étude et de la représentation des connaissances spécialisées et, à ce titre, englobe une double dimension d'analyse, conceptuelle et linguistique, qui caractérise le spectre d'investigation du chercheur. La complémentarité des deux dimensions permet, en effet, d'examiner, d'une part, les conceptualisations d'un domaine de travail donné, d'autre part, les représentations linguistiques (termes) qui concrétisent verbalement les concepts spécialisés étudiés. Le travail du terminologue vise donc à maximiser et à mettre en corrélation les deux dimensions de la terminologie à la fois dans une perspective monolingue et multilingue.

La création de ressources terminologiques contenant la terminologie d'un secteur spécialisé est l'une des principales activités de cette discipline. Ces outils, s'ils sont partagés à grande échelle, constituent un patrimoine inestimable pour l'avancement de la recherche, non seulement dans le domaine strictement linguistique, mais aussi pour des secteurs plus ou moins voisins tels que l'informatique, le traitement automatique des langues, la recherche de l'information. Or, la mise à disposition d'une ressource terminologique numérique sur le Web est une activité qui nécessite une étude de conception préalable approfondie. Le processus de numérisation du matériel terminologique implique donc de repenser la structure des données, des métadonnées et des collections que l'on veut représenter afin qu'elles soient organisées et gérées de manière optimale sur le Web. Faire de la terminologie à l'ère du numérique signifie prendre en considération tous les aspects qui permettent au chercheur de conformer et de modéliser son produit par rapport aux exigences actuelles en termes de bonne gestion des données numériques.

Les raisons qui sous-tendent cet ouvrage s'inscrivent dans ce contexte. En particulier, notre premier objectif est d'illustrer les étapes de conception et de mise en œuvre d'une base de données terminologiques numérisée. Reprenant les lignes directrices promues au niveau

européen pour soutenir la trouvabilité, l'accessibilité, l'interopérabilité et la réutilisabilité (*FAIR principles*) des données de la recherche, nous proposons le paradigme de la « terminologie FAIR » comme une succession de processus afin de modéliser et de rendre disponible une ressource terminologique structurée conformément aux normes ISO actuelles en matière de gestion de la terminologie.

En parallèle, le deuxième objectif de ce travail porte sur la volonté de valoriser les compétences du terminologue numérique. Dans cette perspective, nous entendons souligner la forte composante interdisciplinaire qui caractérise cette figure. Cependant, il faut noter que l'interdisciplinarité évoquée n'est pas conçue comme la caractéristique de savoir maîtriser à la perfection différentes disciplines, mais plutôt comme la capacité de savoir dialoguer et collaborer en synergie avec des experts d'autres disciplines. Et c'est précisément à partir de l'ouverture vers d'autres champs de recherche que la terminologie prend une nouvelle forme et entraîne le développement de compétences de pensée numérique afin de modéliser conceptuellement et structurer correctement le résultat de la recherche terminologique.

Cet ouvrage s'adresse à tous les chercheurs en terminologie qui souhaitent en savoir plus sur la création d'une base de données terminologiques numérique. Sans privilégier un domaine de recherche particulier, les orientations présentées peuvent s'appliquer à tout domaine de travail spécialisé et s'adapter à toute prémisse théorique adoptée. Cependant, il est important de souligner que nous nous concentrons davantage sur l'aspect de modélisation conceptuelle de la ressource. Par conséquent, certaines phases d'implémentation plus techniques ne seront pas envisagées, car elles dépasseraient le cadre de cette étude. Des indications bibliographiques sur ces aspects seront de toute façon indiquées dans les chapitres correspondants.

D'autre part, ce travail s'adresse également à ceux qui n'ont pas de connaissances en terminologie et appartiennent au côté plus informatique de la recherche scientifique. Comme nous l'avons évoqué précédemment, l'aspect collaboratif est essentiel pour la réalisation d'un travail de terminologie numérique. Par conséquent, les notions théoriques fournies peuvent être utiles aux informaticiens afin de comprendre les exigences sous-jacentes à la mise en œuvre effective d'une ressource terminologique.

Cet ouvrage est organisé en six chapitres. Le premier chapitre est destiné à introduire le processus de numérisation de la terminologie. Dans une vocation méta-terminologique, nous retraçons les acceptions classiquement attribuées à la discipline et à toutes ses déclinaisons (terminographie, terminotique et terminologie computationnelle). Ensuite, nous présentons la notion de « terminologie numérique » comme une approche visant à la représentation numérique d'informations terminologiques collectées au sein de ressources conçues à cet effet. Dans cette perspective, nous nous concentrons ensuite sur les principes existant actuellement pour une organisation efficace des données et des métadonnées de la recherche en général et, finalement, des données linguistiques en particulier.

Le deuxième chapitre vise à présenter les fondements pour une terminologie numérique. En ce sens, nous retraçons les théories et les approches qui se sont succédé et/ou juxtaposées depuis la naissance de la discipline : la théorie générale de la terminologie, la théorie communicative de la terminologie, la socioterminologie, la terminologie sociocognitive, la terminologie culturelle, la terminologie textuelle et l'ontoterminologie. Ce panorama global permet, par la suite, d'approfondir les étapes du travail terminologique et, en particulier, d'explorer à la fois les dimensions conceptuelles et linguistiques de la terminologie d'un secteur de spécialité donné.

Une fois les méthodologies de collecte et d'analyse des concepts et des termes présentées, le chapitre 3 ouvre une réflexion évolutive sur la manière de structurer ces données au sein d'une ressource terminologique à concevoir. En particulier, nous présentons, dans un premier temps, une adaptation des principes FAIR susmentionnés à la discipline de la terminologie et nous formulons le paradigme de la « terminologie FAIR » aboutissant à l'adoption de trois normes ISO pour la gestion des terminologies, à savoir la norme ISO-16642 : 2017, la norme ISO-12620 : 2019, et la norme ISO-30042 : 2019, afin de rendre les (méta)données terminologiques trouvables, accessibles, interopérables et réutilisables. La deuxième partie du chapitre permet d'entrer au cœur de la phase de modélisation conceptuelle de la ressource. À partir de la définition du schéma entité-association, nous définissons les objets d'intérêt, leurs propriétés et les associations qui les lient. Ensuite, nous explorons les deux premières normes mentionnées ci-dessus en tant

que la première permet de modéliser les entités « concept », « terme » et « langue » et de définir clairement leurs relations, tandis que la seconde permet de représenter et d’harmoniser les propriétés (catégories de données) choisies pour chaque entité.

Le chapitre 4 vise à faire avancer la réflexion de la phase de conception à la phase d’implémentation de la ressource terminologique. Le format *TermBase eXchange* promu dans le cadre de la troisième norme mentionnée ci-dessus est adopté comme modèle de représentation des données terminologiques, car il est spécifiquement conçu pour l’échange et la réutilisation des données. Par la suite, nous articulons quelques réflexions concernant les phases d’actualisation – à savoir l’insertion, la modification et la suppression – et de publication de données terminologiques, proposant au lecteur une illustration critique des problèmes qui peuvent survenir et quelles solutions peuvent être adoptées pour mieux gérer une collection terminologique.

À titre d’exemple, le chapitre 5 illustre une démonstration de l’application du paradigme FAIR en présentant une étude de cas d’une ressource terminologique pour le domaine médical. La ressource TriMED décrite fait partie d’un projet doctoral visant à fournir un outil multilingue et polyvalent qui prend en charge les besoins d’information de différentes catégories d’utilisateurs – patients, traducteurs et médecins – impliqués dans le processus de communication médicale. En particulier, nous décrivons l’analyse des exigences menée à la fois en termes de spécificités de la langue de spécialité médicale et en termes de structuration de la ressource. Nous présentons donc un nouveau modèle de fiche terminologique structurée et standardisée qui présente un vaste panorama d’informations afin de cadrer le comportement morphosyntaxique, sémantique et phraséologique des termes étudiés. Enfin, nous illustrons l’application Web conçue *ad hoc* pour la consultation et la réutilisation à grande échelle de la collection terminologique multilingue.

En guise de conclusion, le chapitre 6 illustre les possibilités qui s’offrent au terminologue numérique une fois le processus de mise en œuvre au format standard TBX achevé. Cette dernière section vise à démontrer que le processus de FAIRisation des données terminologiques est une activité dynamique qui évolue dans le temps et est sujette à des transformations continues afin de tirer les bénéfices découlant,

par exemple, de l'adoption d'une structuration des données conforme aux principes des données liés qui sous-tendent le Web sémantique.

La réalisation de cet ouvrage n'aurait pas été possible sans l'encadrement et l'appui de mes deux mentors, Professeure Geneviève Marie Henrot et Professeur Giorgio Maria Di Nunzio qui, avec une grande générosité, m'ont enseigné et transmis la passion pour la recherche scientifique. Je tiens aussi à exprimer ma reconnaissance à la Professeure Rute Costa pour tous ses conseils, le temps consacré et ses précieux enseignements tant sur le plan professionnel que personnel.

1 Numérisation de la terminologie

*Hermès doit troquer la plume, le stylet, le papyrus, le papier pour
l'écran d'ordinateur, le disque dur, la tablette.*

(Meunier, 2014)

Numériser signifie représenter des informations sous forme numérique. Le processus de numérisation et la transformation que cela implique se reflètent dans tous les aspects de notre vie. Dans le domaine de la recherche en terminologie, ce processus se traduit, comme nous le verrons au cours de cet ouvrage, par une refonte épistémologique de la discipline même et des savoirs et savoir-faire scientifiques associés. En ce sens, la citation d'ouverture de ce chapitre nous permet de nous positionner en vue de revisiter les disciplines des sciences humaines et sociales en clé numérique et computationnelle. Loin d'enquêter sur des questions inter-linguistiques portant sur l'équivalence traductive entre l'expression anglaise « *digital humanities* » et l'expression française « humanités numériques »¹, Gefen (2015) affirme que :

les humanités numériques sont favorisées et favorisent en retour l'émergence d'une épistémologie empirique et décentrée, indissociable d'une approche pragmatique, interdisciplinaire, décloisonnée des faits culturels.

Dans cette perspective, ce premier chapitre est consacré à la présentation de l'évolution de la discipline de la terminologie en tant que science humaine et sociale, à la réflexion sur les défis épistémologiques auxquels le chercheur est confronté et à l'investigation des besoins qu'il se trouve à devoir satisfaire à l'ère numérique actuelle.

1 À ce propos, voir, par exemple, l'introduction de l'article de Berra (2012).

1.1 Pour une métaterminologie

Mais qu'est-ce que la terminologie numérique ? Et quelles compétences analytiques interdisciplinaires implique-t-elle pour le déroulement optimal du travail terminologique numérique ?

Afin d'introduire ces aspects, cette première section vise à passer brièvement en revue, du point de vue métaterminologique, l'ensemble des acceptions classiquement retenues pour définir les disciplines de la « terminologie », de la « terminographie », de la « terminotique » et de la « terminologie computationnelle ». À partir de ce dernier domaine de recherche appliquée, nous tracerons les frontières – plus ou moins marquées – entre les concepts de « computation » et de « numérisation » appliqués à la terminologie dans le but d'introduire la notion de « terminologie numérique ».

1.1.1 Terminologie, terminographie et terminotique

« Terminologie » est un terme polysémique renvoyant à la fois à la science qui étudie les termes pertinents à une langue de spécialité donnée, et à l'ensemble de ces mêmes termes pour un domaine de spécialisation donné (Sager, 1990). D'un point de vue métaterminologique, il est intéressant de noter l'évolution de la norme ISO 1087, visant à systématiser le vocabulaire de la discipline, et de comparer sa version obsolète publiée en 2000 et intitulée « Travaux terminologiques – Vocabulaire » (ISO 1087-1, 2000) et sa version actuellement en vigueur de 2019 intitulée « Travail terminologique et science de la terminologie – Vocabulaire » (ISO 1087-1, 2019). En effet, alors que dans la première le terme « terminologie » conserve cette double signification, dans la seconde on distingue entre 1) « terminologie » comme l'« ensemble des désignations et des concepts appartenant à un domaine ou à un sujet », et 2) « science de la terminologie » en tant que « science étudiant les terminologies, les aspects du travail terminologique, et les ressources terminologiques et données terminologiques qui en découlent ».

En gardant à l'esprit cette distinction, dans cet ouvrage on se référera – pour de pures raisons conventionnelles – à la « terminologie »

en tant que discipline savante indépendante (Santos et Costa, 2015) dont les objets d'étude sont les « termes », à savoir les désignations des concepts d'un domaine de travail spécialisé². Par conséquent, comme son champ d'investigation tombe sur les langues de spécialité (Lerat, 1995), c'est-à-dire les langues relevant d'un domaine spécialisé de l'activité humaine (comme la médecine, l'économie, le droit), la terminologie diffère de la discipline de la « lexicologie » portant plutôt sur l'étude des « mots » de la langue générale qui ne véhiculent pas des connaissances spécialisées pour un domaine spécifique :

thus the difference between terminology and lexicology is that the former is confined to the lexicon of subject fields (fields of special knowledge), and the latter is concerned with the general lexicon of a language (Warburton, 2021, p. 4).

L'opposition entre « terminologie » et « lexicologie » se reflète également sur le volet pratique des deux disciplines. Selon la norme ISO 1087-1 (2019), la « terminographie » est définie en tant que « travail terminologique visant à créer et à tenir à jour des ressources terminologiques ». La terminographie envisage donc l'ensemble des activités menées par le terminologue (ou terminographe ?)³ pour la conception, l'implémentation et la préservation des ressources terminologiques, structurées comme des collections de données terminologiques relatives à une langue de spécialité donnée. Au contraire, la « lexicographie » porte plutôt sur la création et le maintien de ressources lexicographiques relevant de la langue générale⁴. Comme l'affirme L'Homme (2004), les côtés applicatif et théorique de la discipline sont loin d'être mutuellement exclusifs :

2 Il faut également préciser que la discipline prend des nuances différentes selon le modèle théorique adopté (Costa, 2006). La section 2.1 du chapitre suivant sera consacrée à la description des différents courants de pensée, des approches et des méthodes qui se sont succédé et/ou juxtaposés au fil du temps.

3 Il est intéressant de noter que, dans la norme en question, il n'y a pas une définition pour la figure du « terminologue » et/ou « terminographe ».

4 L'article de Bergenholtz et Kaufmann (1997) propose, dans cette perspective, une enquête critique contrastive des notions de « terminographie » et « lexicographie ».

la pratique fait souvent évoluer une théorie dans une direction qu'elle n'avait pas empruntée jusque-là. Les modèles théoriques, quant à eux, permettent aux praticiens d'examiner un objet sous un éclairage particulier.

Cette dichotomie est fondamentale pour les aspects de numérisation de la terminologie. Comme nous le verrons au fil des pages suivantes, le volet théorique (conception ou modélisation conceptuelle) et le volet pratique (implémentation) et de la terminologie numérique sont des éléments indissociables et mutuels qui caractérisent le travail du terminologue numérique.

Par ailleurs, à la fin des années 1980, la notion de « terminotique » commence à prendre forme dans l'espace de recherche francophone (Gile, 1987 ; Auger, 1989 ; Gouadec, 1987). Le terme – né de la fusion de « terminologie » et « informatique » – désigne une discipline appliquée à mi-chemin entre la linguistique computationnelle, l'ingénierie linguistique et la terminologie, et relève de l'ensemble des activités menées par le terminologue impliquant l'utilisation des outils informatiques. L'informatique commençait donc à assumer un rôle prépondérant dans la réalisation des étapes du travail terminologique. À titre d'exemple, l'étude de (Marzá, 2009, p. 159–160) résume une sélection de tâches terminographiques pour lesquelles la terminotique joue un rôle crucial :

- sélection automatique du corpus composé de textes numériques spécialisés ;
- extraction automatique des termes pertinents pour le domaine de travail ;
- élaboration automatique des définitions de termes ;
- formulation automatique de la structure conceptuelle du domaine de travail ;
- etc.

Aujourd'hui, l'utilisation d'applications informatiques pour effectuer des travaux terminologiques se configure comme une pratique couramment consolidée. La multiplication massive de ce type d'outils, leur convivialité et le besoin croissant de travailler avec de grandes quantités de données font donc de cette pratique une partie intégrante du

travail quotidien du terminologue. Comme l'affirmait déjà L'Homme en 2004 :

actuellement, toute recherche portant sur des termes fait appel à une forme ou une autre de traitement informatique, si bien que la distinction entre terminotique et terminographie ne se justifie que dans un contexte pédagogique (L'Homme, 2004).

En effet, il existe actuellement de nombreux enseignements universitaires de « terminotique » qui sont généralement dispensés et intégrés dans des programmes de formation qui s'adressent aux futurs professionnels de la traduction spécialisée, dans le but d'alimenter les connaissances sur l'utilisation efficace des outils d'extraction automatique de la terminologie et de gestion des données terminologiques multilingues.

1.1.2 Terminologie computationnelle

Aujourd'hui, l'expression « terminologie computationnelle » est de plus en plus courante dans la littérature scientifique. Cette désignation fait référence à un domaine de recherche appliquée ayant pour objet le développement d'outils de manipulation de données textuelles à partir de fonctions mathématiques, de modèles statistiques et de processus de calcul.

La terminologie computationnelle constitue un domaine de recherche du traitement automatique de la langue dont les acteurs – informaticiens, spécialistes des sciences de l'information, linguistes et, bien entendu, terminologues – mettent au point divers traitements automatiques appliqués aux textes spécialisés (L'Homme, 2004).

La « terminologie computationnelle » et la « terminotique », décrite dans la section précédente, sont donc deux disciplines aux objectifs différents : la première porte sur le développement de systèmes, par exemple, pour l'extraction automatique de termes à partir de corpus spécialisés, la deuxième implique plutôt l'utilisation et l'intégration de ces systèmes pour favoriser et soutenir le déroulement du travail terminologique.

Du point de vue historique, le site du sixième atelier de terminologie computationnelle de 2020 offre un aperçu du contexte dans lequel

la discipline est née⁵. La recherche en terminologie computationnelle s'est consolidée au début des années 1990, lorsqu'une série d'articles sur l'extraction monolingue et multilingue de termes à partir de corpus textuels est apparue dans les principaux colloques de linguistique computationnelle, tels que l'*Association for Computational Linguistics* (ACL) et l'*International Conference on Computational Linguistics* (COLING)⁶. Même avant cette période, au cours des années 1980, la nécessité d'un forum commun pour échanger des informations entre les chercheurs en terminologie était bien reconnue. En 1988, la première séance d'un colloque important dans le domaine de la terminologie, à savoir *Terminology and Knowledge Engineering* (TKE), s'est tenue à Trèves, en Allemagne. TKE était et est davantage orienté vers l'organisation des concepts, ce qui reflète une caractéristique importante de la gestion de termes en terminologie. Par ailleurs, en 1995, un groupe de chercheurs français en intelligence artificielle et traitement du langage naturel a organisé le premier colloque en « Terminologie et Intelligence Artificielle » (TIA) à Villetaneuse (France) qui a rapidement acquis le rôle du forum international pour la recherche terminologique. Ensuite, en 1997, un groupe de recherche japonais a développé une tâche commune de reconnaissance automatique des termes, dont le résultat a été rapporté lors de l'atelier NTCIR TMREC (*Workshop on Research in Japanese Text Retrieval and Term Recognition*) qui a eu lieu en 1999⁷.

Dans ce contexte, la communauté scientifique voit naître le domaine de recherche de la terminologie computationnelle. Les vingt dernières années sont en effet la démonstration de la consolidation de la discipline à travers des ateliers spécifiquement conçus à cet égard. Sa première apparition a eu lieu en 1998 lors du colloque international de linguistique computationnelle COLING-ACL'98. À cette occasion, le *First Workshop on Computational Terminology* a été organisé avec l'objectif de rassembler des chercheurs de différentes communautés scientifiques (linguistique computationnelle, terminologie, traduction

5 <https://sites.google.com/view/computerm2020/background>.

6 Le site *ACL Anthology* offre un aperçu de tous les actes publiés dans les colloques les plus importants de linguistique computationnelle : <https://www.aclweb.org/anthology/>.

7 <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings/index.html>.

automatique, recherche d'information et lexicographie) qui partageaient un même intérêt pour les aspects informatiques du traitement terminologique. Les travaux présentés à l'atelier ont ensuite été publiés dans l'ouvrage édité par Bourigault et al. (2001), *Recent advances in Computational Terminology* (John Benjamins Publishing Company). À partir de ce moment-là, d'autres événements ont contribué à promouvoir cette nouvelle discipline et à susciter de nouveaux intérêts sur le terrain⁸. En 2002 le *Second Workshop on Computational Terminology* a conduit à la publication d'un numéro spécial sur la revue internationale *Terminology* édité par Kageura et al. (2004). En 2004 et 2014, les troisième et quatrième ateliers internationaux ont été respectivement organisés toujours lors de l'*International Conference on Computational Linguistics* (COLING). Les deux derniers événements les plus récents ont eu lieu respectivement en 2016 et 2020. Le *5th International Workshop on Computational Terminology* du 2016 a conduit à la publication du récent numéro spécial édité par Drouin et al. (2018) sur la revue *Terminology*. Le *6th International Workshop on Computational Terminology* du 2020 a été intégré sous forme d'atelier d'une journée à l'occasion du colloque *Language Resources and Evaluation Conference* (LREC 2020, Marseille-France)⁹.

La terminologie computationnelle rassemble l'intérêt des chercheurs qui ont ressenti le besoin d'améliorer la communication, ou d'accéder aux informations liées à un domaine de travail spécifique. De plus, les besoins scientifiques dans des domaines à croissance rapide (comme la biomédecine, la chimie et l'écologie) et la quantité écrasante de données textuelles publiées quotidiennement exigent que la terminologie soit acquise et gérée de manière systématique et automatique. À ce propos, différents types de ressources terminologiques, tels que des dictionnaires spécialisés, des bases de données terminologiques et des glossaires ont été mis en œuvre afin de répondre à ces besoins. Dans ce contexte, plusieurs techniques empruntées au traitement du langage naturel, à la recherche d'information, à la linguistique de corpus ou à l'intelligence artificielle permettent d'extraire et de représenter

8 <https://www.aclweb.org/anthology/venues/computerm/>.

9 Les actes de l'atelier sont disponibles à l'adresse suivante : <http://www.lrec-conf.org/proceedings/lrec2020/workshops/COMPUTERM2020/index.html>.

des connaissances spécialisées. Grâce à ces approches, une énorme quantité de données peut être rapidement traitée et filtrée.

D'un point de vue théorique et pratique, les chercheurs en terminologie computationnelle se concentrent d'abord sur l'identification et sur l'extraction automatique des unités terminologiques. Dans la littérature, de nombreuses études sont principalement axées sur l'extraction automatique de termes à partir d'un corpus de documents spécialisés au moyen d'approches 1) linguistiques, 2) statistiques et 3) hybrides : voir, entre autres, Nakagawa (2001) ; Vu et al. (2008) ; Foo (2012) ; Amjadian et al. (2018) ; Sandoval et al. (2019) ; Simon et Kešelj (2018). L'acquisition de termes liés et pertinents à un domaine spécifique de l'activité humaine est effectuée automatiquement à l'aide d'approches computationnelles importées du domaine de la recherche d'information : *Term frequency-Inverse Document Frequency* (TF-IDF) (Salton et Yang, 1973), *Mutual Information* (Church et Hanks, 1990), *T-Score* (Church et al., 1991), *C/NC value* (Frantzi et al., 1998). En outre, des ressources spécifiquement conçues pour cette tâche ont été élaborées afin d'augmenter les performances d'extraction : entre autres, Termostat (Drouin, 2003), BiTermEx (Planas, 2012), TermExtractor (Sclano et Velardi, 2007), TBXTools (Oliver et Vázquez, 2015) et l'outil TermEvaluator (Inkpen et al., 2016) conçu plutôt pour évaluer la qualité du processus d'extraction automatique¹⁰.

L'importance d'une résolution efficace de cette tâche se reflète enfin dans de nombreux domaines de recherche. L'extraction automatique de termes permet d'effectuer des tâches liées à la recherche d'information (comme le repérage de documents pertinents pour une requête donnée), à la fouille de textes (*text mining*), à la construction de ressources terminologiques, à la traduction automatique, etc. Le point de départ de tous ces travaux concerne l'identification des candidats termes et, par conséquent, le filtrage entre, d'une part, les mots d'ordre général et, d'autre part, les termes spécifiques d'un domaine donné. En effet, toutes les études précédemment citées portent (plus ou moins explicitement) sur le concept de « poids » des termes dans une collection de documents

10 Toujours dans ce contexte, l'étude de Costa et al. (2016) présente la description détaillée de différents outils d'extraction terminologique évaluant leur utilité pour les professionnels de la traduction.

afin d'indiquer les différents degrés de pertinence à un domaine. Ce concept a été exprimé, au fil du temps, à travers différentes dénominations. Sparck-Jones (1972) définissait la notion de « spécificité » des termes comme une valeur calculable en fonction de la fréquence d'apparition des termes dans une collection des documents :

the specificity of an individual term is the level of detail at which a given concept is represented [...] terms should be weighted according to collection frequency, so that matches on less frequent, more specific, terms are of greater value than matches on frequent terms (Sparck-Jones, 1972).

En outre, Kageura et Umino (1996) introduisaient le concept de *termhood* (« termicité » ou « termitude » (Humbley, 2016)) afin d'indiquer le degré de relation d'une unité linguistique à des concepts spécifiques pour un domaine, une valeur qui peut également être calculée à l'aide d'approches statistiques. Dans ce sens, le degré de termicité d'un terme est donc une valeur déterminante pour la tâche d'extraction automatique des termes et repose, en général, sur la fréquence d'apparition d'un terme candidat dans le corpus analysé.

Une fois les termes extraits, les variantes terminologiques, telles que des synonymes ou des variantes morpho-syntaxiques, peuvent être regroupées afin de fournir un aperçu précis du contenu d'un document spécialisé (voir, à ce propos, les numéros spéciaux de la revue *Terminology* édités par Kageura et al. (2004) ; Drouin et al. (2015)). Une autre question importante dans ce domaine de recherche est l'extraction de termes à partir de corpus bilingues pour le processus de traduction automatique. De nombreuses études visent donc à identifier des techniques pour l'extraction et l'alignement automatiques de paires d'équivalents (voir, entre autres, Loginova Clouet (2014) ; Lefever et al. (2009) ; Macken et al. (2013)). En outre, les corpus utilisés pour extraire les unités terminologiques peuvent être étudiés plus en détail pour identifier les informations de nature conceptuelle. Dans ce sens, les corpus peuvent conduire à la construction de hiérarchies ou de réseaux conceptuels dans lesquels les termes sont interreliés selon des relations hyperonymiques/hyponymiques, méronymiques/holonymiques ou synonymiques/quasi-synonymiques (Gábor et al., 2018). Les hiérarchies conceptuelles peuvent ensuite être mises en œuvre dans des bases de connaissances terminologiques (Wang et al., 2018).

Enfin, dans l'introduction du numéro spécial édité par Drouin et al. (2018), on dénote la tendance de recherche impliquant la tâche de filtrage de l'information terminologique :

as the sheer volume of data being handled by terminology extraction tools is exploding, we believe that filtering techniques become one of the crucial or parts of computational terminology processing (Drouin et al., 2018, p. 2).

Traditionnellement, les résultats de l'extraction automatique des termes utilisés par les terminologues étaient principalement filtrés manuellement par des spécialistes de la langue ou du domaine. Aujourd'hui, afin de favoriser cette opération chronophage, l'intérêt de la recherche est orienté vers la proposition de méthodes automatiques de filtrage de l'information et le raffinement du produit de l'extraction automatique à l'aide de techniques d'apprentissage profond (*deep learning*) et, en particulier, du modèle vectoriel *word embedding* (Bengio et al., 2003 ; Mikolov et al., 2013 ; Yadav et Bethard, 2018).

1.1.3 Terminologie numérique

L'évolution historique et les intérêts de recherche de la terminologie computationnelle nous permettent d'avancer vers la notion de « terminologie numérique ». Pour tracer une ligne de partage entre les deux disciplines, il faut tout d'abord s'attarder sur les notions de « computationnel » et de « numérique », sous forme d'adjectifs substantivés. Comme nous l'avons vu précédemment, le « computationnel » fait référence à une propriété formelle de nature mathématique liée au concept de calculabilité (Meunier, 2014)¹¹. Appliqué à la discipline de la terminologie, le « computationnel » permet donc d'exploiter des fonctions mathématiques et des processus de calcul pour l'exécution automatisée de tâches terminologiques.

11 En outre, le lecteur est renvoyé à l'article cité ici pour une analyse approfondie des termes « computation » et « ordinateur » : « la computation est un terme qui désigne en mathématiques une propriété formelle d'une fonction mathématique, alors que le terme ordinateur désigne la machine physique particulière parmi plusieurs autres possibles peut effectuer concrètement le calcul d'une fonction » (Meunier, 2014, p. 7).

Pour sa part, le « numérique » concerne plutôt la représentation sous forme de nombres d'une information donnée (par exemple, texte et images). En ce sens, nous définissons la terminologie numérique (*digital terminology*) comme une approche à la discipline classique impliquant la représentation numérique d'informations de type terminologique au sein de ressources spécifiquement conçues dans ce contexte. Ce type de représentation nécessite un processus de modélisation conceptuelle de la structure des données, des métadonnées et de la collection terminologique préalable aux phases proprement dites d'implémentation et de peuplement de la ressource. La terminologie numérique comprend donc l'ensemble des activités pour la conception et l'implémentation de ressources numérisées contenant les informations terminologiques pour un domaine de travail donné¹².

Le développement d'un tel type d'outils n'est en aucun cas un phénomène nouveau dans le monde de la terminologie. En effet, le terminologue/terminographe s'est toujours occupé, comme nous l'avons vu précédemment, de « créer » et « tenir à jour » les ressources terminologiques (ISO 1087-1, 2019). Cependant, comme nous le verrons dans les sections suivantes, les exigences de l'ère numérique actuelle et l'avènement du Web sémantique impliquent que cette tâche soit réalisée en répondant à des critères et des principes spécifiques pour l'organisation et la gestion optimale des données. À ce propos, nous citons un extrait de l'introduction du numéro spécial *Terminological resources in the digital age* publié par Roche et al. (2019) sur la revue *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* :

in a globalised society, terminological dictionaries – including resources such as knowledge and terminological databases, ontologies, wordnets, “traditional” dictionaries, etc. – should comply with both human and machine needs.

Lors de la réalisation d'une ressource, le terminologue de l'ère numérique doit donc prendre en compte non seulement les besoins d'information de l'utilisateur, mais aussi les exigences structurelles de

12 Les chapitres suivants (2–3–4) seront entièrement consacrés à la définition et à la description de ces activités.

la ressource elle-même afin qu'elle puisse être intégrée dans le Web sémantique :

changes regarding information and language processing brought forward by the evolution of society have led to a series of consequences in: (i) the design of terminological resources; (ii) the way data and knowledge are represented; (iii) the way data are interrelated, both within and between resources; (iv) the way users access data; and (v) users' expectations (Roche et al., 2019).

Tous les aspects énumérés ci-dessus constituent donc de nouveaux éléments de réflexion sur lesquels le terminologue de l'ère numérique actuelle doit se pencher. En termes de complémentarité, l'évolution de cette figure implique donc le développement de compétences de pensée numérique accompagnant les réflexions conceptuelles et linguistiques traditionnelles qui caractérisent la discipline. Conformément aux principes qui seront présentés dans la section suivante, le terminologue numérique est donc celui qui considère le produit de son travail comme une ressource « numérisable » qui doit être conçue et structurée de manière optimale afin de préserver et de partager sa valeur à long terme. Cette figure veille donc, non seulement à obtenir une donnée terminologique à partir de l'étude d'une langue de spécialité donnée, mais aussi à la modéliser structurellement de la manière la plus efficace, par exemple, pour en assurer son accessibilité et sa réutilisation pour d'autres études et/ou applications.

1.2 Curation des données numériques

La terminologie numérique, telle qu'elle vient d'être définie, implique donc des compétences de gestion optimale des données visant à préserver la valeur, favoriser le partage et assurer la réutilisation du produit terminologique sur le Web. En général :

gérer les données, c'est s'assurer que celles-ci sont correctement sélectionnées, décrites, préservées et rendues accessibles pour un traitement et/ou une réutilisation, et ce, bien au-delà du projet de recherche qui les a fait naître et les a exploitées au premier chef.

C'est ainsi que les auteurs Calderan et Millet (2015) définissent l'ensemble de tâches relevant de l'activité de « curation de données » (de l'anglais *data curation*). Cette notion identifie donc l'ensemble de bonnes pratiques pour la gestion et la mise à la disposition des données de la recherche (Palmer et al., 2013), dont la responsabilité incombe inévitablement au chercheur qui produit ces données (McLure et al., 2014 ; Corti et al., 2019). La bonne gestion des données n'est pas un objectif en soi, mais plutôt le principal moyen menant à la découverte et à l'innovation des connaissances, ainsi qu'à l'intégration et à la réutilisation des données par la communauté scientifique.

L'ampleur et l'importance de la curation des données à l'ère numérique ont fait émerger de nombreuses initiatives visant à fournir des lignes directrices pour en assurer leur partage et leur réutilisation sur le Web. Dans les sections suivantes, nous nous concentrons sur la description de ces principes qui sous-tendent l'activité de curation des données de la part du terminologue numérique.

1.2.1 Principes FAIR

Les exigences de préservation, de partage et de réutilisation ne s'appliquent pas uniquement au domaine de recherche de la terminologie, mais se configurent comme des enjeux pertinents pour toutes les disciplines scientifiques au sens large.

Dans ce contexte, un ensemble de principes a été publié par Wilkinson et al. (2016), sous le nom de *FAIR Guiding Principles*¹³, afin de promouvoir l'organisation et la gestion optimales des données de la recherche. Ces lignes directrices soulignent la nécessité de mettre à la disposition de la communauté scientifique des données de la recherche trouvables, accessibles, interopérables et réutilisables (de l'anglais *Findable, Accessible, Interoperable and Reusable*)¹⁴. Les principes FAIR

13 <https://www.go-fair.org/fair-principles/>.

14 Une note de nature inter-linguistique s'impose à cet égard. Dans le contexte francophone, l'expression anglaise *FAIR data* trouve comme équivalent français l'expression « données FAIR » dont l'acronyme renvoie à différentes formes

se réfèrent à trois types d'entités : les données (ou tout objet numérique), les métadonnées (informations sur cet objet numérique) et les infrastructures. Toutes les composantes du processus de recherche bénéficient donc de l'application de ces principes, afin d'en garantir la transparence, la reproductibilité et la réutilisabilité. En particulier :

- La « trouvabilité » implique que les données et les métadonnées soient identifiées de manière unique et persistante grâce à l'attribution de PID (*Persistent Identifier* – identifiant pérenne). De plus, les données doivent être décrites de manière exhaustive grâce à l'utilisation de métadonnées qui doivent, à leur tour, indiquer explicitement à quelles données elles se réfèrent. Enfin, les deux éléments doivent être indexés sur des ressources trouvables.
- L'« accessibilité » implique que les données et métadonnées puissent être récupérées grâce à leur identifiant en utilisant des protocoles de communication standards, ouverts, gratuits, universellement applicables et, si nécessaire, prévoyant des procédures

développées selon les sources consultées. Si l'on regarde, par exemple, le « Plan de données de la recherche du CNRS - Novembre 2020 » (<https://bit.ly/3Hy6jnq>), on peut remarquer que la traduction française utilisée correspond à « données **F**aciles à trouver, **A**ccessibles, **I**ntéropérables et **R**éutilisables », permettant ainsi de conserver l'acronyme d'origine. Le même avantage est apporté par la solution proposée sur le site FranceTerme (<http://www.culture.fr/franceterme/terme/INFO962>, date d'accès : 28/02/2022) dans lequel l'expression « données **F**acilement **A**ccessibles, **I**nteropérables et **R**éutilisables » est illustrée comme traduisante. Dans ce dernier cas, il convient cependant de noter que le choix retenu entraîne, à notre avis, la perte de la composante de « Trouvabilité » (*Findability*) qui se retrouve incorporée dans la formule « Facilement Accessible ». En effet, le fait d'être trouvable et le fait d'être accessible désignent deux propriétés différentes selon les caractéristiques illustrées dans ces pages. Ainsi, une ressource linguistique peut être trouvable (ou facilement trouvable) sur le réseau Internet par un utilisateur, mais en même temps inaccessible car, par exemple, elle n'inclut pas de protocoles de communication ouverts. L'accessibilité est donc une propriété qui se manifeste secondairement à la trouvabilité des données, des métadonnées ou de l'infrastructure. Pour ces raisons, nous utiliserons dans cet ouvrage l'acronyme FAIR retraçant le sens qui promet une distinction claire entre les deux composants de *Findability* et d'*Accessibility* qui seront traduits individuellement par « Trouvabilité » et « Accessibilité ».

- d'authentification/d'autorisation. Enfin, les métadonnées doivent être accessibles même lorsque les données ne sont plus disponibles.
- L'« interopérabilité » implique que les données et métadonnées utilisent un langage de représentation des connaissances formel, accessible et partagé. De plus, ces deux éléments doivent utiliser des vocabulaires respectant les principes FAIR et inclure, le cas échéant, des références à d'autres données et métadonnées.
 - La « réutilisabilité » implique que les données et métadonnées soient décrites de manière exhaustive avec des attributs pertinents et précis, qu'elles soient délivrées avec des licences d'utilisation et qu'elles aient clairement indiqué les détails concernant leur origine. Enfin, les deux éléments doivent suivre les normes de référence pour le domaine d'intérêt.

Une architecture de données FAIR constitue donc le pilier pour l'évolution de la science ouverte. Les travaux de la plateforme *European Open Science Cloud* (EOSC) sont basés sur ces lignes directrices avec l'ambition ultime de développer un « Web de données et de services FAIR » pour la science en Europe sur lequel un large éventail de services à valeur ajoutée peut être construit¹⁵.

Dans le cadre de la recherche en linguistique, l'application de ces principes acquiert un rôle fondamental pour préserver la valeur des ressources produites constituant un héritage culturel inestimable qui, par sa nature, se prête à une réutilisation fructueuse à de nombreuses fins. Toutefois, dans ce contexte, les données de la recherche sont encore loin d'être FAIR. En effet :

language resources (dictionaries, terminologies, corpora, etc.) developed in the fields of corpus linguistics, computational linguistics and natural language processing (NLP) are often encoded in heterogeneous formats and developed in isolation from one another (Cimiano et al., 2020).

Cela rend leur découverte, leur réutilisation et leur intégration, une tâche difficile et lourde. Dans ce contexte, il faut noter les efforts menés par l'infrastructure de recherche européenne *Common Language*

15 <https://bit.ly/3FfDMmm>.

Resources and Technology Infrastructure (CLARIN)¹⁶ qui permet aux chercheurs en sciences humaines d'accéder aux ressources et technologies linguistiques disponibles au niveau européen, et vise à promouvoir une architecture des données conforme aux principes FAIR¹⁷ (De Jong et al., 2018). Dans cette perspective, cet ouvrage entend contribuer à l'application des principes FAIR dans l'espace de recherche en terminologie. Dans le chapitre 3, nous présenterons la formulation du paradigme de la « terminologie FAIR » afin de fournir des lignes directrices pour la conception et l'implémentation d'une ressource terminologique numérisée conforme aux lignes directrices décrites ci-dessus.

1.2.2 Principes des données liées

Dans le cadre de la curation des données numériques, les principes des données liées jouent un rôle clé en tant qu'ensemble de bonnes pratiques pour partager et lier les données sur le Web sur la base des normes internationales promues par le *World Wide Web Consortium* (W3C) (Bizer et al., 2009).

Introduit par Berners-Lee (2006), le terme *Linked Data* (données liées) désigne les pratiques nécessaires à la mise en œuvre de l'architecture du Web sémantique, entendu comme une extension du Web actuel dans lequel :

information is given well-defined meaning better enabling computers and people to work in cooperation (Berners-Lee et al., 2001).

Les principes des données liées sont donc appliqués pour permettre un passage massif d'un « Web de documents », orienté principalement vers les humains, à un « Web de données », destiné plutôt aux machines afin de favoriser un leur dialogue plus efficace¹⁸. En particulier, les quatre principes des données liées prévoient :

16 <https://www.clarin.eu>.

17 <https://www.clarin.eu/fair>.

18 Pour une description approfondie du Web sémantique, le lecteur est également renvoyé à la récente publication de Hendler et al. (2020).

1. l'utilisation d'URI (*Uniform Resource Identifier* – identifiant uniforme de ressource)¹⁹ pour nommer de manière unique toute entité, concrète ou abstraite, présente sur le Web (par exemple, toutes les entrées des termes dans une base de données terminologiques) ;
2. l'emploi d'URI HTTP (*Hypertext Transfer Protocol* – protocole de transfert hypertexte)²⁰ afin que tout agent (ordinateur ou humain) puisse consulter et récupérer ces noms via des protocoles de communication standard ;
3. l'enrichissement d'informations utiles sur l'URI consulté à travers l'utilisation de standards Web tels que le *Resource Description Framework* (RDF)²¹ pour la représentation des données et SPARQL²² pour l'interrogation de données liées ;
4. l'inclusion de liens vers d'autres URI afin d'établir un réseau de ressources liées et de faciliter la découverte de nouvelles entités.

Grâce à l'architecture établie avec l'adoption de ces principes, tout – des objets du monde réel aux concepts abstraits – peut désormais être identifié et relié sur le Web, aidant en ce sens à soutenir également la réalisation des principes FAIR susmentionnés pour les données numériques (Cimiano et al., 2020).

Un autre élément clé dans le cadre des données liées est la nécessité de rendre les mêmes données disponibles via des licences ouvertes, conduisant ainsi à la notion de données ouvertes liées (*Linked Open Data* – LOD). Les données ouvertes liées reposent sur un système

19 Un URI est une chaîne de caractères identifiant une ressource (physique ou abstraite) sur le Web. Un URI est une version généralisée d'une URL (*Uniform Resource Locator* – localisateur uniforme de ressource), et une URL est un type d'URI sous forme de chaîne de caractères qui permet d'identifier une ressource du Web par son emplacement et de préciser le protocole Internet pour la récupérer : <https://www.ietf.org/rfc/rfc3986.txt>.

20 <https://datatracker.ietf.org/doc/html/rfc7540>.

21 Le RDF est un modèle de données commun qui permet de modéliser formellement les informations sur le Web sur la base de la notion de triplet en tant qu'association constituée d'un sujet, d'un prédicat et d'un objet : <https://www.w3.org/RDF/>. Le modèle RDF sera décrit plus en détail dans la section 6.1 du chapitre 6.

22 <https://www.w3.org/TR/rdf-sparql-query/>.

d'évaluation à 5 étoiles²³ selon lequel les données à publier doivent répondre cumulativement à un certain nombre de critères :

1. la première étoile prévoit que les données sont disponibles sur le Web, quel que soit le format, sous licence ouverte ;
2. la deuxième étoile présume que les données sont disponibles sous forme de données structurées (par exemple au format Excel) et lisibles par machine ;
3. la troisième étoile envisage à la fois les critères précédents avec, en plus, l'utilisation d'un format non-propriétaire (par exemple CSV au lieu d'Excel) ;
4. la quatrième étoile comprend tout ce qui précède plus l'utilisation de normes ouvertes du W3C, telles que RDF et SPARQL ;
5. la cinquième étoile englobe tout ce qui précède avec l'ajout du lien vers des données appartenant à d'autres personnes afin de fournir du contexte supplémentaire.

Aujourd'hui, de nombreuses données ouvertes liées ont été publiées et reliées entre elles au sein de ce que l'on appelle le nuage de données liées ouvertes (*Linked Open Data Cloud*) qui a connu une croissance exponentielle ces dernières années²⁴.

1.2.3 Données linguistiques liées ouvertes

Le potentiel des données ouvertes liées a permis l'émergence, depuis une dizaine d'années, d'une communauté de recherche dont les efforts sont orientés vers l'application de ces principes aux données de nature linguistique (Cimiano et al., 2020). Dans ce panorama, l'initiative de données linguistiques liées ouvertes (*Linguistic Linked Open Data – LLOD*) (Chiaros et al., 2012) prend forme rassemblant de nombreux chercheurs – en linguistique, linguistique computationnelle et traitement automatique des langues – actifs dans la promotion des meilleures

23 <https://www.w3.org/DesignIssues/LinkedData.html>.

24 À cet égard, le site Web suivant montre l'évolution du diagramme des données ouvertes liées au fil du temps : <https://lod-cloud.net>.

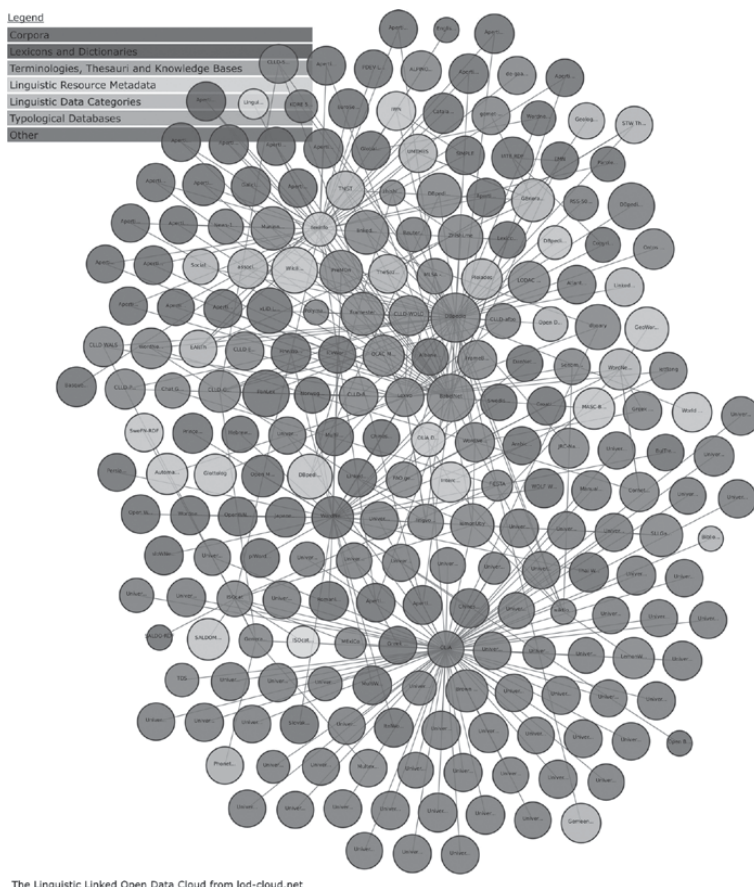


Image 1.1 – Nuage LLOD, Juillet 2020

pratiques et outils pour favoriser l’accessibilité et la réutilisation des ressources linguistiques et langagières sur le Web.

Les efforts de la communauté ont conduit au développement du nuage de données linguistiques liées ouvertes (*Linguistic Linked Open Data cloud*)²⁵ qui indexe l’écosystème croissant d’ensembles de données

25 <http://linguistic-lod.org/lod-cloud>.

linguistiques liés sémantiquement sur le Web. L'image 1.1²⁶ montre les ensembles de données linguistiques actuels (corpus, terminologies, thésaurus, etc.) publiés en tant que LLOD et leurs connexions associées. Le nuage est maintenu par l'*Open Linguistics Working Group* (OWLG)²⁷ de l'*Open Knowledge Foundation* dont l'objectif principal est de promouvoir et d'atteindre l'ouverture des données en linguistique (McCrae et al., 2016). De plus, en ce qui concerne les différentes initiatives visant l'application des LLOD, on peut citer, parmi les autres, le développement du récent projet international Prêt-à-LLOD²⁸, et l'organisation d'événements à fort impact scientifique, comme l'atelier *Linked Data in Linguistics* (LDL)²⁹ et le colloque *Language, Data and Knowledge* (LDK)³⁰.

En ce sens, ces occasions offrent le scénario idéal pour sensibiliser à l'importance de la curation des données de nature linguistique afin de se conformer aux besoins actuels de l'ère numérique et du Web sémantique.

1.3 Conclusion

Ce chapitre avait pour but d'introduire à la notion de numérisation de la terminologie. Dans une perspective méta-terminologique, nous avons retracé les notions de « terminologie », de « terminographie », de « terminotique » et de « terminologie computationnelle » en décrivant, d'une manière générale, les champs d'investigation et les facettes de ces disciplines.

26 Source : "Open Linguistics Working Group (2012), The Linguistic Linked Open Data cloud diagram (draft), version of, <http://linguistic-lod.org/lod-cloud>" (2020).

27 <http://linguistics.okfn.org>.

28 <http://www.pret-a-lod.eu/>.

29 Le lien suivant renvoie au site de la dernière édition de l'atelier de 2020 : <http://ldl2020.linguistic-lod.org>.

30 Le lien suivant renvoie au site de la dernière édition du colloque de 2021 : <http://2021.ldk-conf.org>.

Par ailleurs, une attention particulière a été accordée à la notion de « terminologie numérique » en tant que champ de recherche visant à la représentation de l'information terminologique numérique, et consacré à l'ensemble des activités pour la conception et l'implémentation optimales des ressources terminologiques numériques. Comme nous l'avons annoncé à plusieurs reprises, le terminologue numérique est celui qui conçoit le produit de son travail comme un patrimoine à préserver et à rendre disponible pour une réutilisation future. Cela implique des défis épistémologiques pour le chercheur qui se trouve impliqué dans la tâche de modélisation conceptuelle des données en fonction des besoins actuels de l'ère numérique et de l'avènement du Web sémantique. En ce sens, nous avons défini les principes FAIR, les principes des données liées et les initiatives qui tournent autour de ces bonnes pratiques pour la curation des données linguistiques numériques.

Le chapitre suivant sera consacré à la définition des fondements pour une terminologie numérique afin de décrire, dans un premier temps, les approches théoriques en terminologie et, ensuite, les étapes du travail terminologique préalable aux phases de conception et de mise en œuvre d'une ressource numérique.

2 Fondements pour une terminologie numérique

*The final objective of any theory must be to describe real data [...].
The theory of terminology is no exception.*

(Cabré Castellví, 2003)

Avant d'entrer dans le vif du travail du terminologue numérique, il est nécessaire de retracer les fondements théoriques de la terminologie qui ont contribué à affirmer ce qui est, à ce jour, une discipline indépendante (Costa, 2006, 2013 ; Santos et Costa, 2015) et, de plus en plus fréquemment, objet d'enseignement académique.

Les sections suivantes sont consacrées à l'évolution historique de la terminologie et à la description des différentes approches qui se sont succédé et/ou juxtaposées au fil du temps. Ce panorama nous permettra, par la suite, d'aborder les étapes du travail terminologique et les méthodologies de son élaboration afin d'identifier les éléments sous-jacents à la constitution d'une ressource terminologique numérique conforme aux bonnes pratiques de curation des données.

2.1 Théories et approches en terminologie

L'histoire de la terminologie est étroitement liée aux personnalités qui, par leur théorisation, ont contribué à sa définition formelle.

L'ère pionnière des études terminologiques est liée à la figure de l'ingénieur Eugen Wüster (1898–1977) considéré à l'unanimité comme le fondateur de la terminologie moderne. En effet, la discipline commence à acquérir un statut indépendant depuis la publication de sa thèse de doctorat *Internationale Sprachnormung in der Technik, besonders*

in der Elektrotechnik en 1931 (Felber, 1981). Son travail a permis la fondation de l'« école de terminologie de Vienne »³¹, caractérisée par une conception de la discipline centrée sur le concept et orientée vers la normalisation des termes afin de favoriser une communication claire entre les experts d'un secteur donné. Cependant, la principale étude associée à la figure multiforme de Wüster est le travail publié à titre posthume par son collègue et disciple Helmut Felber (1925–2005) en 1979 : *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Considéré par Picht (1982) comme une ressource pédagogique, l'ouvrage constitue le cadre théorique de ce qu'on appelle aujourd'hui la « Théorie générale de la terminologie » ou « Théorie classique de la terminologie ».

Les principes sur lesquels repose la théorie wüsterienne visent à désambigüiser la communication dans les secteurs de spécialité. Pour ce faire, la théorie générale de la terminologie privilégie une relation biunivoque entre les concepts et les désignations de concepts (termes) : « à une forme correspond un seul concept et un concept est exprimé par une seule forme » (L'Homme, 2004). En adoptant une démarche de type onomasiologique³², le terminologue se trouve engagé d'abord dans l'exploration de la dimension et de la structure conceptuelle d'un domaine de travail donné et, ensuite, dans l'identification de sa réalisation linguistique, c'est-à-dire des termes désignant ces concepts. De plus, selon le principe de biunivocité, les phénomènes linguistiques tels que la synonymie et la polysémie devraient être minimisés, car ils pourraient constituer des obstacles à une communication claire et précise.

La composante de normalisation de la langue est donc au cœur de cette théorie. À cet égard, il faut rappeler que Wüster a travaillé activement dans le domaine de la normalisation internationale en coopérant avec l'Organisation internationale de normalisation – *International Organization for Standardization* (ISO)³³ – et en participant à

31 Pour une description des différentes « écoles » de terminologie, voir, entre autres, le travail de Humbley (2014).

32 La démarche onomasiologique (du concept vers le terme) s'oppose à la démarche sémasiologique généralement utilisée en lexicographie qui prévoit comme point de départ les mots dans leur environnement naturel (les textes) et, par la suite, l'identification de leur sens.

33 <https://www.iso.org/>.

la création du Comité technique 37 chargé de l'élaboration de normes en « Langage et Terminologie »³⁴ (Felber, 1980 ; Galinski, 1982). La vision de la terminologie wüstérienne reflète donc l'union entre la nécessité d'une communication claire et précise entre les experts d'un secteur spécialisé, à une époque où les échanges internationaux dans les domaines de l'industrie et de la technologie étaient de plus en plus fréquents, et sa conception interdisciplinaire de la terminologie comme discipline à la frontière entre la linguistique, l'ontologie et l'informatique (Zanola, 2018).

L'empreinte laissée par le chercheur autrichien a jeté les bases pour des réflexions critiques par la communauté scientifique sur différents aspects de la discipline. En reprenant (Warburton, 2021, p. 12–13), la principale critique adressée est que la théorie générale de la terminologie ne prend pas en compte la langue authentique dans son usage dynamique et que les concepts sont étudiés en dehors des contextes de communication spécifiques. Ces aspects ont conduit à remettre en cause certains des principes de base de la terminologie classique. Selon l'étude de Campo (2013), on peut tracer cinq courants de pensée qui se sont formalisés en tant que théories et/ou approches alternatives à la théorie générale :

1. la théorie communicative de la terminologie (Cabré Castellví, 1993, 2003) ;
2. la socioterminologie (Boulanger, 1995 ; Gaudin, 1993, 2003) ;
3. la terminologie sociocognitive (Temmerman, 2000) ;
4. la terminologie culturelle (Diki-Kidiri, 2000) ;
5. la terminologie textuelle (Auger et L'Homme, 1994 ; Bourigault et Slodzian, 1999 ; Condamines, 2005).

Sans prétendre les analyser tous en détail, nous voyons dans les pages suivantes quels sont les principes généraux sur lesquels reposent ces formulations.

Vers la fin des années 90, M. Teresa Cabré Castellví rassemble une série d'articles qui seront ensuite présentés comme le fondement de la

34 Le lien pour consulter les normes élaborées au sein de ce comité technique est le suivant : <https://www.iso.org/fr/committee/48104/x/catalogue/>.

« Théorie communicative de la terminologie » (Cabré Castellví, 1993, 1999a,b). En mettant l'accent sur le côté communicatif de la discipline, l'objet d'étude de cette approche est l'« unité terminologique » (UT) qui est caractérisée par trois dimensions d'analyse : cognitive, linguistique et communicative. Une extension de cette approche a été formalisée par des publications postérieures (Cabré Castellví, 2003) sous le nom de *theory of doors* proposant un modèle qui :

attempts to represent the plural, but not simultaneous, access to the object; and in such a way that, whether starting from the concept or the term or the situations, the central object or terminological unit, is directly addressed.

Les UT sont donc considérées comme des unités lexicales activées par leurs conditions pragmatiques et adaptées à un type de communication donné. Pour reprendre la terminologie employée par Cabré Castellví (1999a), l'approche théorique communicative est donc orientée vers l'étude de la représentation de la connaissance *in vivo*, c'est-à-dire de la production linguistique spontanée, plutôt que de la représentation de la connaissance *in vitro* issue de la langue normalisée.

Dans la même période historique, une approche orientée vers la dimension sociale de la terminologie commence à prendre forme. L'approche formalisée sous le nom de « Socioterminologie » (Boulangier, 1995 ; Gaudin, 1993, 2003) est la première à remettre en cause et à rejeter la « doctrine wüstérienne » (Humbley, 2004). En particulier, en s'opposant à l'idée de normalisation, les socioterminologues concentrent leur attention sur les aspects sociaux qui interviennent dans la langue et qui influencent les variations terminologiques (Gaudin, 2003). En ce sens, le terme n'est pas perçu comme une étiquette fixe d'un concept, mais il représente plutôt une unité variable selon des facteurs sociolinguistiques. De la dimension sociale, il s'ensuit que les concepts sont des entités dynamiques qui évoluent, car les disciplines auxquelles elles appartiennent se développent au fil du temps. En outre, en contestant le principe d'univocité, les socioterminologues affirment que la polysémie et la synonymie appartiennent à la nature de la langue et, donc se produisent inévitablement dans le discours spécialisé. À partir de cette approche, une pléthore de publications sur la variation des termes a émergé dans le panorama de la recherche en terminologie : parmi celles-ci, on peut citer les études de Freixa (2006) sur les

causes de la variation dénomminative et le volume de Drouin et al. (2017) qui recueille un riche panorama sur les perspectives multiples de la variation terminologique.

La dimension sociale des termes est relancée et élargie même par l'approche sociocognitive à la terminologie qui a ses origines dans le *Centrum voor Vaktaal en Communicatie* (CVC), groupe de recherche coordonné par Rita Temmerman et appartenant au Département de linguistique appliquée de la *Vrije Universiteit Brussel* (VUB)³⁵. Temmerman (2000) développe un « modèle sociocognitif » de la terminologie basé sur la sémantique cognitive et la théorie des prototypes. Contestant l'universalité des concepts présupposée dans la théorie classique de la terminologie, l'auteure soutient que si certains concepts de domaines spécifiques peuvent être délimités à travers une liste de caractéristiques finies, d'autres concepts sont beaucoup plus difficiles à circonscrire. En fournissant quelques exemples du domaine des sciences de la vie, Temmerman démontre que certains concepts ne possèdent pas de prototype universel et, par conséquent, qu'ils sont mieux caractérisés comme des « unités de compréhension » (angl. *unit of understanding*) définies par un processus cognitif d'acquisition de connaissances dans le domaine de travail donné. Selon cette vision, le point de départ n'est pas donc le concept, mais plutôt le terme que l'on trouve dans les textes écrits par des spécialistes du domaine (Temmerman, 1997). Parmi les principes fondamentaux de cette approche, une importance particulière est réservée également aux phénomènes de la synonymie et de la polysémie, qui jouent un rôle clé dans les langues de spécialités, et à la dimension diachronique de la langue considérée comme un axe d'analyse inévitable (Temmerman, 2000). Enfin, une attention particulière est également consacrée à l'étude des métaphores, l'implémentation de ressources terminologiques multilingues et l'analyse de la variation terminologique.

Une approche terminologique entièrement tournée vers l'aspect culturel est plutôt proposée par Diki-Kidiri (2000) et inspirée de la situation spécifique des langues et sociétés africaines. La « terminologie culturelle » :

35 <https://bit.ly/3amNGEF>.

a pour objectif principal l'appropriation de nouveaux savoirs et savoir-faire qui arrivent dans une société donnée. Elle permet à cette société de trouver le mot juste pour exprimer chaque concept nouveau en puisant ses ressources linguistiques dans sa propre culture et selon sa propre perception du réel (Diki-Kidiri, 2007).

En ce sens, l'approche propose que la vision du monde détermine la façon dont les individus classent, ordonnent, nomment et catégorisent tout ce qui est perçu ou conçu. La terminologie culturelle met donc l'accent sur la diversité culturelle et la nécessité de préserver les identités culturelles par l'appropriation et la diffusion d'une terminologie spécialisée.

La dernière approche évoquée par l'étude susmentionnée de Campo (2013) est identifiée sous le nom de « terminologie textuelle » (Auger et L'Homme, 1994 ; Bourigault et Slodzian, 1999 ; Condamines, 2005). En particulier, pour la terminologie textuelle, le texte constitue le point de départ de l'analyse et de la description du terme qui est considéré comme un « construit » résultant de l'analyse faite par le terminographe (Bourigault et Slodzian, 1999). Cette approche s'est développée à la fin du XX^e siècle dans une période historique où l'évolution de la technologie et de la science a entraîné des changements considérables dans la dimension théorique et, surtout, pratique de la discipline de la terminologie. La possibilité de disposer de grandes quantités de données sous forme électronique et la nécessité de les gérer de manière optimale ont contribué au développement d'approches d'analyse terminologique basées sur les corpus (Condamines, 2007). Comme l'affirment Bourigault et Slodzian (1999) :

devant la masse des données à analyser et les délais imposés, la tâche d'analyse de corpus ne peut être envisagée qu'avec l'utilisation des outils de la terminologie textuelle (concordanciers, extracteurs de candidats termes, extracteurs de relations candidats, classifieurs, etc.).

Dans ce sens, les corpus électroniques lisibles par la machine offrent aux terminologues une grande quantité de matériel pour leur travail terminographique. Les avantages sont axés sur une réduction du travail manuel humain, une meilleure gestion de la terminologie et la possibilité d'adapter les applications terminologiques selon les besoins

spécifiques des utilisateurs. En outre, cette méthodologie facilite également les analyses contrastives, par exemple focalisées sur le changement de signification des termes en fonction du contexte ou sur leur variation diachronique.

À ce propos, L'Homme (2006) explique qu'au Canada, deux groupes de recherche placent la terminologie dans un paradigme basé sur le corpus. L'approche conceptuelle basée sur le corpus de l'Université d'Ottawa et l'approche lexico-sémantique basée sur le corpus établie par l'Observatoire de linguistique Sens-Texte (OLST)³⁶ de l'Université de Montréal profitent de la dimension informatique (outils et applications) et de la disponibilité de corpus électroniques pour mener des recherches en terminologie. En particulier, l'approche lexico-sémantique proposée par L'Homme (2006) se concentre sur les propriétés linguistiques du terme, abandonnant la perspective conceptuelle au profit de la dimension sémasiologique promue en lexicographie. Dans l'introduction de sa récente publication, l'auteure explique les raisons de l'utilité d'adopter une approche lexico-sémantique pour l'étude de la terminologie, en soulignant que :

lexical semantics and terminology have much in common, since both disciplines aim to answer questions about the nature of words, word content (i.e. the nature of meaning), the relationship between word content and our construal of reality, and relations between words and word meanings (L'Homme, 2020, p. 1).

Une attention particulière est également portée sur la Sémantique des cadres (Fillmore, 1976) en tant que modèle théorique permettant de répondre à certaines questions concernant les propriétés linguistiques des termes (L'Homme, 2020, p. 43–50). En outre, sa pertinence dans le contexte de la terminologie a conduit à la proposition d'une approche cognitive de la terminologie appelée *Frame-based Terminology* – mettant en œuvre les principes de la Sémantique des cadres – développée par Pamela Faber et ses collègues à l'Université de Grenade (Faber, 2012 ; Faber, 2015).

La dernière approche à la terminologie envisagée dans cet aperçu découle des efforts menés par l'équipe Condillac³⁷, groupe de recherche

36 <http://olst.ling.umontreal.ca>.

37 <http://new.condillac.org>

pluridisciplinaire qui rassemble des chercheurs de différentes institutions, et qui centre ses études sur la combinaison entre ontologie et terminologie³⁸. Le paradigme, défini par Christophe Roche sous le nom d'« ontoterminologie », naît de la volonté de réconcilier et, en même temps, distinguer la double dimension conceptuelle et linguistique de la terminologie (Costa, 2013 ; Roche, 2012, 2015). L'ontoterminologie est une “*terminology whose conceptual system is a formal ontology relying on epistemological principles*” (Roche et al., 2009) et reconnaît la conceptualisation d'un domaine comme le point de départ de tout projet terminologique. Comme la théorie générale de la terminologie, l'ontoterminologie est basée sur une approche onomasiologique : “*concepts in expert knowledge became the starting point in terminological analysis*” (Roche, 2012). En ce sens, l'approche se distingue d'autres méthodes orientées vers l'implémentation des ontologies, comme la « termonotographie » – issue de l'approche sociocognitive de la terminologie (Temmerman, 2000) – qui reste de nature sémasiologique : “*terms (linguistic expressions) in texts became the starting point in terminological analysis*” (Temmerman et Kerremans, 2003).

Le niveau conceptuel de la terminologie est donc séparé de sa dimension linguistique. La première dimension est représentée par l'ontologie formelle, c'est-à-dire la représentation informatique du système conceptuel d'un domaine élaboré grâce à des outils d'ingénierie informatique. La dimension linguistique est plutôt composée de termes liés entre eux par des relations linguistiques-sémantiques, telles que l'hyponymie, l'hyperonymie, la synonymie. En conséquence, les définitions ontologiques (Roche, 2015), qui renvoient à des spécifications logiques du concept, doivent être distinguées des définitions terminologiques qui sont des explications linguistiques. Comme on peut le voir

38 Le terme « ontologie » surgit en référence à une branche de la philosophie qui se concentre sur l'étude de l'être. À partir de la fin du XX^e siècle, le terme acquiert un nouveau rôle dans les domaines de l'informatique, de l'intelligence artificielle et de l'ingénierie des connaissances. Lorsqu'elle est utilisée dans le contexte de la recherche en terminologie, l'ontologie acquiert le sens de “*formal explicit specification of a shared conceptualization*”, telle que définie par Studer et al. (1998) pour le domaine de l'ingénierie des connaissances.

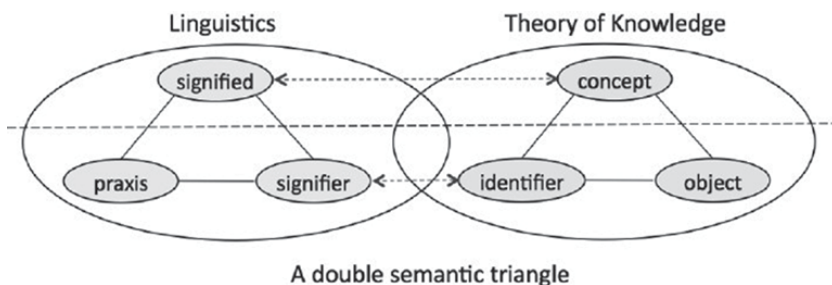


Image 2.1 – Le double triangle sémiotique

dans l'image 2.1³⁹, l'ontoterminologie propose donc un nouveau paradigme qui distingue et relie les deux systèmes sémiotiques (linguistique et conceptuel) qui composent chaque terminologie tout en conservant leurs identités et leurs différences fondamentales. L'approche insiste sur l'importance des principes épistémologiques qui régissent la conceptualisation d'un domaine, ainsi que sur l'importance d'une approche scientifique de la terminologie dans laquelle l'expert joue un rôle fondamental. Une meilleure connaissance du système conceptuel favorise une approche onomasiologique, dans laquelle la connaissance et la compréhension des concepts parviennent à identifier les termes les plus appropriés pour les désigner et, par conséquent, à établir la terminologie la plus efficace et la plus rigoureuse d'un domaine spécialisé :

like classical terminology, ontoterminology enables standardisation of language. But unlike classical terminology ontoterminology preserves the diversity of language between different communities of practice since they share the same domain and standardised conceptualisation (Roche, 2012).

L'ontoterminologie traduit donc la nécessité de replacer le concept et sa désignation au centre de la terminologie, en préservant également sa dimension sociolinguistique.

39 L'image est tirée de l'article Roche (2012) disponible en ligne au lien suivant : http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf.

2.2 Travail terminologique numérique

L'éventail des théories et des approches existant actuellement en terminologie constitue les bases de connaissances pour pouvoir entreprendre le travail terminologique numérique. Dans cet ouvrage, nous adoptons le postulat théorique, décrit à la fin de la section précédente, selon lequel la Terminologie est une discipline ayant une double dimension d'analyse, linguistique et conceptuelle, indispensable pour le travail terminologique (Costa, 2013 ; Roche, 2015 ; Santos et Costa, 2015). En reprenant la norme ISO 1087-1 (2019), le « travail terminologique » est :

l'activité portant sur la systématisation de la collecte, de la description, du traitement et de la présentation des concepts et de leurs désignations.

Dans une note du même article, la norme explique également que « le travail terminologique vise souvent à créer et à tenir à jour des ressources terminologiques » conduisant ainsi à la notion de « terminographie » introduite au chapitre 1. Dans cette perspective, par « travail terminologique numérique », nous entendons l'ensemble des activités visant à la représentation numérique de la dimension conceptuelle et linguistique de la terminologie d'un secteur donné de connaissances spécialisées au sein d'une ressource terminologique.

Selon la prémisse théorique susmentionnée, le terminologue numérique doit lier et maximiser à la fois la perspective conceptuelle et la perspective linguistique pour représenter efficacement la connaissance d'un domaine spécifique au sein d'une ressource terminologique numérique. Pour ce faire, le « concept » et le « terme » doivent être considérés comme deux entités autonomes qui sont pourtant interdépendantes dans le travail terminologique (Silva, 2014) : le premier permet de conceptualiser le monde, alors que le dernier nous permet de parler de cette conceptualisation et de la partager avec d'autres êtres humains. La double dimension implique que tant les conceptualisations des experts sur un sujet donné que leurs discours doivent être considérés et interreliés dans le travail terminologique (Carvalho et al., 2016). En ce sens, le travail terminologique numérique repose sur le bénéfice mutuel tiré de l'adoption d'une approche complémentaire combinant la

perspective onomasiologique (du concept au terme) et sémasiologique (du terme au concept) de la terminologie :

the semasiological and onomasiological approaches are two complementary terminological methodologies that should be used in the construction of knowledge representation tools (Santos et Costa, 2015, p. 153).

Une fois établie la nécessité d’explorer les deux dimensions d’analyse, la question se pose de savoir quelle approche choisir comme point de départ. Comme nous l’avons vu dans la section précédente, le point de départ, qu’il soit conceptuel ou linguistique, est fortement débattu et dépendant du modèle théorique adopté pour l’analyse terminologique. Toutefois, l’étude de Santos et Costa (2015, p. 177) montre que :

for knowledge representation purposes, it would be therefore convenient to depart from an onomasiological approach, followed by the semasiological approach to confirm the representation of conceptualizations.

Pour la construction de ressources terminologiques pour la représentation des connaissances, le terminologue numérique devrait donc adopter comme point de départ une approche basée sur les concepts (onomasiologique), suivie d’une approche basée sur le corpus (sémasiologique). En tout état de cause, quel que soit le point de départ choisi qui peut varier selon la finalité sous-jacente à la réalisation de la ressource, il est nécessaire d’explorer et d’intégrer les deux dimensions d’analyse (conceptuelle et linguistique) afin de maximiser le bénéfice découlant de leur interrelation. Selon l’étude de Costa (2013), en effet, si l’on choisit d’analyser le concept sans prendre en compte le terme, le domaine scientifique sous-jacent à l’analyse n’est plus de la Terminologie, mais de l’Ingénierie des connaissances. À son tour, se concentrer sur le terme tout en faisant abstraction du concept n’est pas, encore une fois, de la Terminologie, mais de la Lexicographie spécialisée.

2.2.1 Dimension conceptuelle

Que signifie analyser la dimension conceptuelle de la terminologie ? Comme mentionné précédemment, la méthodologie d’analyse onomasiologique impose au terminologue d’identifier, de délimiter et de définir

les concepts d'un secteur donné de connaissances spécialisées afin de constituer un système conceptuel dans lequel les concepts sont organisés et classés sur la base de leurs relations. Le système conceptuel se présente sous la forme d'une carte (ou schéma) conceptuelle obtenue à l'aide de supports de représentation graphique tels que CmapTools⁴⁰, un logiciel de cartographie conceptuelle gratuit développé par le *Florida Institute for Human and Machine Cognition* (IHMC). Le système conceptuel montre donc comment les concepts sont liés les uns aux autres. En particulier :

the most frequently used conceptual relations are is_a and part_of, corresponding to generic and partitive hierarchical relations. The part_of relation is also known as a part-whole relation and is classified as a mereology relation based on axioms (Santos et Costa, 2015, p. 164).

Ces types de relations permettent donc de représenter la structure hiérarchique d'un système conceptuel. En outre, un schéma conceptuel peut également inclure des indications sur les caractéristiques essentielles et distinctives d'un concept⁴¹. La production d'un tel schéma nécessite une analyse minutieuse des propriétés des concepts et comment ceux-ci sont liés les uns aux autres. Le terminologue est donc amené à se familiariser (et souvent à acquérir) avec les connaissances spécialisées du domaine de travail en consultant des sources fiables qui contribuent au processus de cartographie. Il est également important de souligner le rôle que jouent les experts du domaine dans l'analyse de la dimension conceptuelle. En effet, *they are the entity that possesses domain knowledge* (Santos et Costa, 2015) : la collaboration entre les experts du domaine et le terminologue est configurée comme un scénario optimal pour valider la qualité des données conceptuelles obtenues dans ce processus d'analyse (Costa et al., 2012).

40 <https://cmap.ihmc.us>.

41 Selon la norme ISO 1087-1 (2019), la caractéristique essentielle est une « caractéristique d'un concept qui est indispensable pour comprendre ce concept », et la caractéristique distinctive est un type de « caractéristique essentielle utilisée pour distinguer un concept d'autres concepts associés ».

2.2.2 Dimension linguistique

La dimension linguistique vise plutôt l'analyse des textes spécialisés considérés comme représentatifs de la *knowledge in action* (Santos et Costa, 2015). En termes de complémentarité, si d'une part, on analyse la conceptualisation des savoirs spécialisés, d'autre part, on se concentre sur la représentation linguistique de la même conceptualisation. Les textes spécialisés sont l'environnement naturel pour l'identification des termes utilisés par les experts pour désigner les concepts objet d'analyse du domaine de travail. En effet, les textes apparaissent comme le principal véhicule pour représenter, structurer et diffuser les connaissances :

the text is the most efficient means for experts to communicate with the members of their professional community. The text is the place of discussion and the place of organization of ideas; it is the place of construction and deconstruction, but also the place of the uncertain and the polemic (Santos et Costa, 2015, p. 158).

Les textes constituent donc une ressource précieuse pour le terminologue et le terrain où il est possible d'identifier les réseaux lexicaux qui devraient refléter autant que possible l'organisation conceptuelle du domaine. En ce sens, le travail du terminologue est orienté vers la collection de textes spécialisés, constituant un corpus de travail (monolingue ou multilingue selon les finalités du projet) représentatif du domaine analysé⁴². Par la suite, à l'aide d'outils automatiques ou semi-automatiques d'extraction de candidats termes à partir d'un corpus électronique⁴³, les termes pertinents pour le domaine de travail donné sont identifiés. Dans le but de représenter la dimension linguistique de la connaissance spécialisée, le terminologue est donc chargé de la constitution de réseaux lexicaux où les termes sont liés entre eux par des relations sémantiques de type hiérarchique, associatif et/ou d'équivalence (Clarke, 2001).

À ce stade, les efforts du terminologue sont orientés vers l'interrelation entre les deux systèmes sémiotiques. Cette tâche consiste en

42 Sur la notion de représentativité du corpus, le lecteur est renvoyé à la publication suivante : Costa et Silva (2008).

43 Voir par exemple les outils décrits dans la section 1.1.2 du chapitre précédent.

la cartographie et l'intégration des informations obtenues à la fois à partir de la représentation conceptuelle et linguistique de la connaissance spécialisée. Le cadre conceptuel et les réseaux lexicaux doivent donc être analysés afin de relier les termes à leurs concepts respectifs. Cette cartographie permet d'intégrer mutuellement les informations obtenues à partir des deux dimensions d'analyse afin d'assurer la systématisation de la représentation et de l'organisation des connaissances du domaine de travail. Dans le cas d'une représentation multilingue, tous les réseaux lexicaux produits pour les langues de travail seront soumis à la cartographie conceptuelle qui, en ce sens, est considérée comme une entité indépendante de la langue. L'objectif, bien que souhaitable, n'est donc pas d'atteindre un chevauchement parfait entre les deux dimensions (conceptuelle et linguistique), mais plutôt de tirer le bénéfice de leur intégration en termes de représentation exhaustive des connaissances d'un secteur de spécialité.

2.3 Conclusion

Ce chapitre a été consacré à retracer les différentes théories et approches de la discipline terminologique. Il est important de noter que, à notre avis, quels que soient le modèle théorique adopté et le point de départ choisi, les notions présentées ici servent de fondements pour pouvoir représenter la terminologie d'un secteur de spécialité donné. En particulier, nous avons décrit – sans privilégier un domaine particulier – les étapes du travail terminologique préalables aux phases de conception et d'implémentation d'une ressource terminologique. Dans ce contexte, l'exploration de la relation entre la dimension conceptuelle et linguistique de la terminologie constitue l'essence du travail terminologique :

it is in the relationship between what is designated and the designation that we can find the essence of terminological work (Santos et Costa, 2015, p. 159).

À ce stade, une fois ce travail réalisé, le terminologue numérique se charge de collecter et de représenter les données de type conceptuel et

linguistique obtenues au sein d'une ressource numérique. Le chapitre suivant abordera les principes de conception d'une base de données terminologique représentant les deux dimensions de la terminologie et répondant aux besoins définis au chapitre 1 pour la curation optimale des données terminologiques numériques.

3 Conception d'une ressource terminologique

*C'était comme un monde nouveau ouvert à moi,
le monde de la science, que je pouvais enfin connaître en toute liberté.*

Marie Skłodowska-Curie (1867–1934)

La première étape de la conception d'une ressource terminologique consiste à analyser une série d'exigences essentielles à satisfaire à différents niveaux. En particulier, l'analyse des exigences envisage 1) ce que nous voulons représenter, 2) à qui nous voulons nous adresser et 3) quelle structure nous voulons attribuer à nos données.

Le premier point a été schématiquement identifié dans le chapitre précédent. La conception d'une ressource, qui prend la forme d'une base de données terminologiques, ayant le but de représenter les connaissances spécialisées d'un domaine de travail spécifique, implique un travail terminologique préalable visant l'identification et la détermination des deux dimensions d'analyse (conceptuelle et linguistique) de la terminologie. Ce travail permet d'identifier les éléments d'intérêt à représenter sous forme numérique dans la base de données (concepts et termes) et leurs relations exprimées sous forme de schéma conceptuel et de réseaux lexicaux monolingues ou multilingues selon la finalité du projet. La deuxième question à se poser concerne l'analyse des besoins d'information que l'on veut satisfaire pour une ou plusieurs catégories d'utilisateurs de la ressource. Puisque le but du présent ouvrage est de fournir des lignes directrices qui peuvent s'adapter à n'importe quel domaine de travail, nous n'abordons pas – du moins pour le moment – cet aspect qui est fortement lié aux objectifs individuels de chaque projet terminologique. Cependant, le chapitre 5 sera consacré à la description d'une étude de cas visant la conception et l'implémentation d'une ressource terminologique pour le domaine médical qui prend le nom de TriMED et, à cette occasion, nous décrivons l'analyse des besoins

d'information des utilisateurs auxquels nous nous adressons (à savoir les patients, les traducteurs et les médecins) afin d'identifier leurs exigences informatives et d'y répondre de manière suffisamment exhaustive.

Le troisième niveau d'analyse envisage une réflexion fondamentale sur la structure des données et les métadonnées terminologiques. Comme nous l'avons vu au chapitre 1 (section 1.2), l'organisation optimale des données de la recherche est configurée comme une étape indispensable au progrès de la science ouverte. Dans la lignée de l'idée que le terminologue numérique est celui qui vise à préserver la valeur de son travail et à assurer sa réutilisation secondaire au profit de la communauté scientifique, la prochaine section sera consacrée à décrire l'application des principes FAIR dans le domaine de recherche de la terminologie. Partant de cette adaptation, nous proposons le paradigme de la « terminologie FAIR » qui se matérialise dans l'application des directives contenues dans les trois dernières normes publiées par l'ISO TC/37 SC 3 pour la gestion des ressources terminologiques (à savoir la norme ISO-16642 : 2017⁴⁴, la norme ISO-12620 : 2019⁴⁵, et la norme ISO-30042 : 2019⁴⁶) afin de rendre les (méta)données terminologiques trouvables, accessibles, interopérables et réutilisables. Enfin, nous décrivons les phases de conception d'une ressource terminologique sur la base des principes précités en nous concentrant, en particulier, sur le métamodèle structurel à adopter et sur la description des catégories de données à représenter.

3.1 Principes FAIR en terminologie

Les principes FAIR voient donc le jour dans le but de favoriser l'adoption d'un ensemble de bonnes pratiques pour l'organisation et la gestion optimale des données de la recherche (Wilkinson et al., 2016). Publiés en 2016 et devenus le cœur et le moteur de la science ouverte⁴⁷, ces

44 <https://www.iso.org/fr/standard/56063.html?browse=tc>.

45 <https://www.iso.org/fr/standard/69550.html?browse=tc>.

46 <https://www.iso.org/fr/standard/62510.html?browse=tc>.

47 <https://bit.ly/3FfDMmm>.

principes se réfèrent aux résultats scientifiques, y compris les données, les métadonnées et les flux de travail qui les entourent, et visent à combler le manque de pratiques amplement partagées, clairement articulées et largement applicables concernant la publication des données générées par la recherche scientifique *lato sensu* (Da Silva Santos et al., 2016). En reprenant la description détaillée de chaque élément fournie sur le site de l'initiative *GO FAIR*⁴⁸, dans les sections suivantes nous retraçons tous les composants nécessaires pour réaliser la FAIRisation des données et nous en proposons l'application à la discipline de la terminologie.

3.1.1 Terminologie trouvable

La première étape pour l'élaboration optimale de la terminologie numérique est de produire de données et de métadonnées terminologiques qui soient facilement trouvables, tant pour les humains que pour les ordinateurs. La « trouvabilité » est en effet un élément essentiel lorsque nous souhaitons un outil disponible, en l'occurrence une ressource terminologique, qui vise à être utilisé et partagé par une grande variété d'utilisateurs. Afin de répondre à ce principe, les (méta)données terminologiques doivent :

- T.1 être associées un identifiant unique et persistant ;
- T.2 être accompagnées d'une riche description ;
- T.3 inclure des références explicites ;
- T.4 être indexées dans une ressource consultable.

Le principe T.1 est le plus important, car les identifiants uniques et persistants à l'échelle mondiale suppriment toute ambiguïté dans la signification de données publiées, en attribuant une étiquette exclusive à chaque élément des métadonnées et à chaque concept de l'ensemble de données. Dans ce contexte, les identifiants prennent la forme de liens Internet : par exemple, un *Uniform Resource Identifier* (URI) qui aboutit souvent à une *Uniform Resource Locator* (URL) d'une page Web. Il existe des répertoires de données qui génèrent automatiquement

48 <https://www.go-fair.org/fair-principles/>.

des identifiants uniques et persistants pour les ensembles de données déposés. Les identifiants permettent aux ordinateurs d'interpréter les données de manière significative dans les étapes de recherche et/ou intégration automatique. Dans ce sens, les identifiants sont essentiels à l'interopérabilité personne-machine et machine-machine, ce qui est la clé de la vision de la science ouverte.

Lors de l'implémentation de ressources terminologiques numérique FAIR, les métadonnées peuvent (et devraient) être accompagnées d'une riche description (T.2), y compris des informations descriptives sur le contexte, la qualité, l'état et les caractéristiques des données terminologiques. La conformité à T.2 aide les utilisateurs à désambigüiser la signification des données, à les localiser et à augmenter leur réutilisation. Dans ce contexte, il est important aussi que les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent (T.3). L'association entre un fichier de métadonnées et l'ensemble de données doit être explicite, en mentionnant l'identifiant unique d'un ensemble de données dans les métadonnées. Enfin, il est essentiel que les données de recherche terminologique et les métadonnées associées soient disponibles et indexées sur une ressource consultable en ligne (T.4).

3.1.2 Terminologie accessible

La condition préalable de « trouvabilité » est suivie de la nécessité de mettre à disposition une terminologie « accessible ». Une fois trouvées les données terminologiques en ligne, l'utilisateur doit savoir également comment y accéder même à travers des systèmes d'authentification et d'autorisation. Ce principe établit que :

- A.1 les (méta)données sont récupérables par leur identifiant à l'aide d'un protocole de communication normalisé ;
- A.1.1 le protocole est ouvert, gratuit et universellement implémentable ;
- A.1.2 si nécessaire, le protocole permet une procédure d'authentification et d'autorisation ;
- A.2 les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.

Les données terminologiques et les métadonnées doivent être récupérables et accessibles via des protocoles de communication standard, tels que l'*Hypertext Transfer Protocol* (HTTP), et elles ne doivent pas être mises à disposition uniquement via un logiciel propriétaire (A.1). En ce sens, des ressources terminologiques comme, pour ne citer que quelques exemples, DiCoInfo⁴⁹ et DicoEnviro⁵⁰ implémentées par le groupe de recherche canadien Observatoire de linguistique Sens-Texte (OLST)⁵¹ de l'Université de Montréal répondent au principe d'« accessibilité » parce qu'elles sont disponibles sur des pages Web qui supportent le protocole HTTP.

Pour maximiser la réutilisation des données, le protocole doit être gratuit et ouvert, et donc, implémentable pour faciliter la récupération des données (A.1.1). En outre, comme clairement mentionné sur le site de l'initiative précitée *GO FAIR*, le principe A.1.2 est un élément clé, mais souvent mal compris, des principes FAIR. Le « A » de l'acronyme FAIR ne signifie pas nécessairement ouvert ou gratuit. Cela implique plutôt de fournir les conditions exactes dans lesquelles les données sont accessibles. Par conséquent, même des données protégées et privées peuvent être FAIR. L'implémenteur peut en effet décider de consentir l'accès à sa propre base de données terminologique par l'intermédiaire de systèmes d'authentification ou d'autorisation afin de définir éventuellement les droits spécifiques de l'utilisateur.

Enfin, il faut également considérer que la maintenance et la gestion des données d'une ressource peuvent avoir un coût élevé qui ne permet pas de les stocker indéfiniment. Toutefois, le stockage des métadonnées est généralement beaucoup plus facile et moins cher. Par conséquent, le principe A.2 stipule que les métadonnées devraient persister même lorsque les données ne sont plus conservées.

49 <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>

50 http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi.

51 <http://olst.ling.umontreal.ca>.

3.1.3 Terminologie interopérable

La mise en œuvre et la maintenance des outils terminologiques sont des tâches qui nécessitent beaucoup de ressources, d'énergie et de temps disponible (Warburton, 2015). Au fil du temps, il se peut que les données terminologiques doivent être intégrées à d'autres données ou interagir avec différentes applications afin d'en assurer l'analyse, le stockage et le traitement. Dans ce contexte, la notion d'« interopérabilité » terminologique acquiert un rôle fondamental. Pour que les (méta) données terminologiques soient interopérables, elles doivent :

- I.1 être exprimées dans un langage formel, accessible et partagé ;
- I.2 s'appuyer sur un vocabulaire contrôlé ;
- I.3 inclure des références croisées.

L'exigence exprimée avec le principe I.1 se traduit par le besoin d'avoir à disposition des données terminologiques échangeables, interprétables et lisibles tant par les personnes, en utilisant un langage partagé, que par les machines sans avoir besoin d'algorithmes spécialisés ou conçus *ad hoc*. L'interopérabilité signifie généralement que chaque système informatique connaît au moins les formats d'échange de données de l'autre système. Pour que cela se produise et pour garantir la recherche et l'interopérabilité automatiques des ensembles de données, il est essentiel d'utiliser 1) des vocabulaires contrôlés, des ontologies, des thésaurus (ayant des identifiants uniques et persistants) (I.2), et (2) un métamodèle de données, c'est-à-dire un cadre bien défini pour décrire et structurer les données terminologiques. Enfin, les (méta)données doivent inclure des références croisées à d'autres (méta)données (I.3). Le but est donc de formuler des liens significatifs entre les (méta)ressources de données pour enrichir les connaissances contextuelles sur les données. Concrètement, il faut spécifier si 1) un ensemble de données s'appuie sur un autre ensemble de données, 2) des ensembles de données, supplémentaires sont nécessaires pour compléter les données, ou 3) des informations complémentaires sont stockées dans un autre ensemble de données.

3.1.4 Terminologie réutilisable

L'objectif ultime des principes FAIR est d'optimiser la « réutilisation » des données. Dans le contexte terminologique, cette notion est étroitement liée au besoin d'échange (importation et/ou exportation) de données terminologiques. Afin de garantir la disponibilité d'une terminologie réutilisable, les métadonnées et les données doivent être :

- R.1 décrites avec une pluralité d'attributs précis et pertinents ;
- R.1.1 publiées avec une licence d'utilisation claire et accessible ;
- R.1.2 associées à la description de leur provenance ;
- R.1.3 organisées selon les normes internationales.

Dans ce contexte, le terminographe qui souhaite mettre à disposition des données terminologiques réutilisables doit fournir non seulement des métadonnées qui permettent leur trouvabilité, mais également des métadonnées qui décrivent de manière détaillée le contexte dans lequel les données ont été générées (R.1). Cela peut inclure, par exemple, la description 1) de l'objectif pour lequel les données (ou plus généralement la ressource terminologique) ont été générées ou collectées, 2) de toute particularité ou limitation des données dont les autres utilisateurs devraient être conscients, 3) de la date et l'auteur de génération/collecte des données, etc. En outre, le principe R.1.1 concerne les droits d'utilisation que l'implémenteur accorde aux données terminologiques. Ce critère devrait être décrit clairement afin d'éviter que l'ambiguïté sur ce sujet puisse limiter considérablement la réutilisation des données par n'importe quel organisme ou individu. La description de la provenance des données est un autre critère qui en garantit la réutilisation (R.1.2). Le terminographe doit donc répondre aux questions suivantes : qui a généré ou collecté les données ? Comment ont-elles été traitées ? La ressource contient-elle des données extraites d'autres sources ? Enfin, pour assurer la réutilisation (et pas seulement) des données terminologiques, le critère R.1.3 déclare expressément la nécessité d'organiser les données selon les normes existantes en la matière. Cette exigence, sur laquelle repose la modélisation conceptuelle proposée aux sections suivantes, découle de l'idée que des données ayant une structure similaire ou identique peuvent « collaborer » pour différentes applications et donc être gérées de manière optimale.

3.2 Paradigme de la « terminologie FAIR »

Les principes FAIR décrits jusqu'à présent et dont nous avons proposé une adaptation en terminologie n'ont pas le statut de normes officielles (Mons et al., 2017). Cependant, leur mise en œuvre dans le cadre de la terminologie numérique peut être favorisée par le biais de l'adoption de normes ISO pour la conception et la gestion de ressources terminologiques.

Nous introduisons donc le paradigme de la « terminologie FAIR » qui se traduit par la mise au point d'un flux de travail terminologique numérique afin de fournir une structure de données qui respecte les besoins de trouvabilité, d'accessibilité, d'interopérabilité et de réutilisabilité. Le paradigme se matérialise dans l'application des directives contenues dans les trois dernières normes publiées par l'ISO TC/37 SC 3 pour la gestion des ressources terminologiques. En particulier, nous nous référons à :

1. la norme ISO-16642 : 2017⁵² qui définit le métamodèle structurel abstrait *Terminological Markup Framework* (TMF) pour la représentation de ressources terminologiques afin de favoriser leur interopérabilité ;
2. la norme ISO-12620 : 2019⁵³ qui définit les propriétés des catégories de données et leur documentation dans un répertoire ouvert qui garantit leur trouvabilité et leur accessibilité ;
3. la norme ISO-30042 : 2019⁵⁴ qui définit le format de représentation *TermBase eXchange* (TBX) spécifiquement conçu pour l'échange et, par conséquent, la réutilisation des données terminologiques.

La combinaison de ces normes constitue la réalisation du paradigme de la « terminologie FAIR ». L'approche proposée s'adresse à tout/terminologue/terminographe expert/e qui veut concevoir une ressource

52 <https://www.iso.org/fr/standard/56063.html?browse=tc>.

53 <https://www.iso.org/fr/standard/69550.html?browse=tc>.

54 <https://www.iso.org/fr/standard/62510.html?browse=tc>.

terminologique numérique, indépendamment du domaine de travail : cette modélisation est donc applicable à tout projet terminologique visant à mettre à la disposition de la communauté des données terminologiques numériques FAIR. Dans les sections suivantes de ce chapitre, nous traitons de la description des deux premières normes, car elles nous permettent de poser les bases de la conception d'une ressource terminologique correspondant au paradigme susmentionné. La première section du chapitre 4 vise plutôt la description de la dernière norme citée en tant que section de l'ouvrage consacrée aux formats de représentation pour l'implémentation de la ressource⁵⁵.

3.3 Modélisation conceptuelle de la ressource

Une fois l'analyse préliminaire des exigences réalisée, la conception d'une base de données terminologique commence dès la phase de modélisation conceptuelle. La « modélisation conceptuelle » d'une base de données est un type de représentation de haut niveau des données qui permet de définir les objets d'intérêt, leurs propriétés et les relations entre ces objets⁵⁶. La modélisation conceptuelle se positionne donc au niveau de la conception abstraite et est indépendante de toute implémentation spécifique :

55 À cet égard, nous précisons que les descriptions fournies dans les pages suivantes ne visent pas à se substituer aux trois normes en question. Notre objectif est de fournir une vue d'ensemble afin d'argumenter la proposition du paradigme de la « terminologie FAIR ». En ce sens, pour tous ceux qui souhaitent concevoir et implémenter une ressource numérique selon les prémisses développées, nous renvoyons à la consultation des trois versions officielles des normes précitées.

56 En ce sens, il est important de préciser que le terme « modélisation conceptuelle » utilisé en informatique ne fait référence ni à la dimension conceptuelle ni à la carte conceptuelle précédemment décrites en référence à la discipline de la terminologie. Nous utilisons donc ce terme dans son acception généralement acceptée dans le cadre de la conception d'une base de données.

The purpose [...] is to represent the informal requirements of an application in terms of a formal and complete description, but independent of the criteria for representation used in database management systems (Atzeni et al., 1999, p. 160).

Afin de modéliser conceptuellement une base de données, il existe différents schémas dans la littérature qui permettent de représenter graphiquement les objets d'intérêt de la ressource⁵⁷. Dans ce chapitre, nous nous référons au schéma entité-association en tant que modèle largement utilisé pour la représentation de base de données.

3.3.1 Schéma entité-association

Le modèle entité-association est un modèle conceptuel utilisé pour la conception d'applications de bases de données relationnelles conçu par Chen (1976). Ce modèle propose un ensemble d'éléments qui permettent de décrire les objets d'intérêt de la ressource. Cette description passe par la réalisation d'un schéma – le schéma entité-association – dont les éléments de base sont :

1. les entités ;
2. les propriétés ;
3. les associations.

L'élément « entité » représente l'ensemble des objets d'intérêt qui ont des « propriétés » (ou attributs) communes. L'« association » représente plutôt un lien logique entre plusieurs entités.

Dans le cas d'un projet de terminologie numérique visant la représentation conceptuelle et linguistique de la connaissance spécialisée d'un domaine de travail donné, nous avons vu, dans la section 2.2, que les entités de base d'intérêt sont donc 1) les concepts, 2) les termes qui les désignent et 3) les langues de travail envisagées. Les associations sont plutôt les relations préalablement identifiées entre l'entité-concept et l'entité-terme au moyen d'une langue de travail donnée. Enfin, les propriétés sont les caractéristiques propres à chaque entité dont les valeurs varient selon l'objet spécifique observé.

57 <https://www.sciencedirect.com/topics/computer-science/conceptual-modelling>.

Dans l'image 3.1, nous illustrons un exemple de schéma entité-association pour une ressource terminologique. Le schéma est composé de trois entités – représentées par des rectangles – reliées entre elles par des associations sous forme graphique de losanges. Le couple de nombres – appelé cardinalité – présent sur chaque segment qui lie les entités aux associations représente respectivement le nombre minimum et le nombre maximum d'objets pouvant être liés entre les entités impliquées. Chaque entité a des propriétés : l'entité-concept a la propriété « identifiant » qui représente une valeur unique qui distingue chaque objet de cette entité ; l'entité-langue a la propriété « code ISO »⁵⁸ qui identifie les objets de cette entité et une propriété « nom » qui dénomme la langue ; l'entité-terme a une propriété « identifiant » qui, comme pour l'entité-concept et l'entité-langue, représente une valeur unique et une propriété « désignation » qui représente la séquence de caractères qui fait référence au terme. Comme on peut le voir sur l'image, les propriétés « identifiant » des trois entités sont marquées d'un cercle plein faisant référence aux valeurs uniques et distinctives de chaque élément de l'entité. Alors que les propriétés marquées d'un cercle vide – « nom » pour l'entité-langue et « désignation » pour l'entité-terme – indiquent tous les attributs n'ayant pas pour objectif d'identifier de manière unique l'objet de l'entité impliquée⁵⁹.

58 On fait référence ici à la norme ISO 639 : <https://www.iso.org/iso-639-language-codes.html>.

59 Considérons, par exemple, le cas de deux termes ayant la même séquence graphique de caractères, mais des identifiants différents, et désignant deux concepts différents.

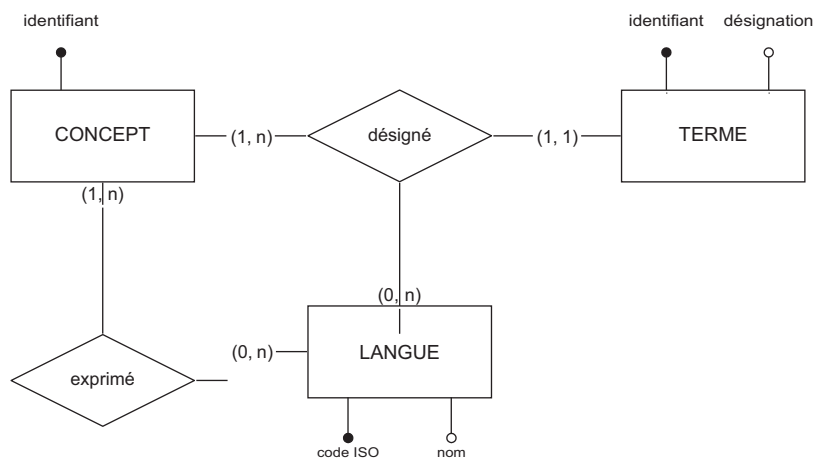


Image 3.1 – Exemple schéma entité-association

L'exemple en question montre un type de schéma ayant des cardinalités selon lesquelles :

- chaque objet de l'entité-concept est exprimé par au moins une langue et au plus un nombre générique n de langues. Le même objet est désigné par au moins un couple terme-langue et au plus n couples terme-langue ;
- chaque objet de l'entité-langue peut exprimer de 0 à n concepts et peut être lié à un nombre de couples terme-concept allant d'un minimum de 0 à un maximum de n ;
- chaque objet de l'entité-terme ne désigne qu'un concept pour une langue donnée.

Cet exemple représente une situation possible dans laquelle le terminologue qui conçoit la base de données veut exprimer le fait que : étant donné un concept, celui-ci ne peut exister que si 1) il est exprimé par une langue et 2) il a aussi au moins un terme qui le désigne. En même temps, une langue existe indépendamment du fait qu'elle exprime un concept : en ce sens, le terminologue vise à sauvegarder potentiellement toutes les langues connues et identifiées par un code ISO dans la base de données, sans nécessairement avoir de concepts liés. Pour la même raison, la langue qui n'exprime aucun concept dans la base de données n'a pas aucun terme associé. Enfin, étant donné un terme identifié de manière univoque, un

seul couple concept-langue associé doit exister. Cette structure permet d'insérer plusieurs termes avec la même séquence de caractères mais qui, étant identifiés différemment (en d'autres termes, ayant des attributs « identifiant » différents), peuvent renvoyer à des concepts différents.

Ceci, comme nous l'avons dit, peut être un exemple possible d'un monde d'intérêt à représenter. En changeant les exigences de la base de données que nous voulons modéliser, nous pourrions construire un monde d'intérêt différent à partir du même schéma mais avec des cardinalités différentes. Par exemple, on pourrait dire qu'une langue dans la base de données n'existe que si 1) elle exprime au moins un concept (1, n vers l'association « exprimé ») et 2) elle est liée à au moins un terme (1, n vers l'association « désigné »). Encore une fois comme exemple alternatif, on pourrait choisir une prémisse théorique selon laquelle un concept existe indépendamment du fait qu'il peut être exprimé par une langue et peut être désigné par un terme, et choisir donc une cardinalité 0, n vers l'association « exprimé » et une cardinalité 0, n vers l'association « désigné ».

En guise de conclusion, la réflexion sur cet exemple sert à faire comprendre au lecteur que la modélisation conceptuelle et le schéma entité-association ne visent pas à représenter toutes les combinaisons possibles de tous les mondes possibles. Ces modèles servent plutôt à représenter un monde d'intérêt particulier avec un certain ensemble d'exigences préalablement identifiées sur la base des prémisses théoriques adoptées. En outre, il est important de souligner que la phase de conception décrite ci-dessus peut être adaptée à la mise en œuvre de tout type de base de données. Or, il semble légitime de se demander ce dont nous disposons en terminologie du point de vue de la modélisation conceptuelle. Dans la section suivante, nous illustrons la norme ISO-16642 : 2017 qui décrit un métamodèle structurel spécifiquement conçu pour la représentation d'objets d'intérêt au sein d'une base de données de type terminologique.

3.3.2 Métamodèle structurel *Terminological Markup Framework*

La norme ISO-16642 : 2017⁶⁰ définit un métamodèle structurel abstrait – à savoir le *Terminological Markup Framework* (TMF) – pour la

60 <https://www.iso.org/fr/standard/56063.html?browse=tc>.

représentation de collections de données terminologiques. Cette norme acquiert un rôle important dans la réalisation du travail du terminologue numérique pour deux raisons principales :

1. la première est que les lignes directrices fournies permettent de modéliser conceptuellement les entités, les propriétés et les associations à représenter au sein d'une ressource terminologique ;
2. la seconde est que l'adoption répandue de cette norme pour la conception des ressources permet d'assurer leur interopérabilité en privilégiant, en ce sens, la mise en œuvre d'un des critères préalablement établis à travers la formulation du paradigme de la « terminologie FAIR ».

En général, l'interopérabilité entre ressources terminologiques pourrait être définie comme une sorte de cohérence structurelle, afin que chaque système ne puisse entraîner d'incohérences internes ni d'incompatibilité avec d'autres systèmes (Lee et Romary, 2010). Appliquée à la conception et l'implémentation des ressources linguistiques, l'interopérabilité se traduit donc par une exigence de base au niveau structurel des ressources : pour qu'ils « collaborent », ces outils doivent partager le même métamodèle structurel. Par exemple, la structure de base d'une ressource lexicographique prévoit qu'une entrée lexicographique contienne un lemme (la forme base d'une unité lexicale) et une ou plusieurs définitions (représentant différentes significations) dans une ou plusieurs langues. En revanche, une ressource terminologique exige qu'une entrée terminologique désigne un concept dans une langue donnée et, dans le cas d'entrées terminologiques multilingues, à un ou plusieurs concepts totalement ou partiellement équivalents dans une ou plusieurs langues. Pour ces deux types de représentations formelles, il y a en fait deux normes différentes : d'une part, la norme ISO 24613-1 : 2019 qui définit le cadre de balisage lexical *Lexical Markup Framework* (LMF)⁶¹ pour la conception de ressources de type lexicographique ; d'autre part, la norme ISO 16642 : 2017 – objet d'intérêt de cette section – qui définit le cadre de balisage *Terminological Markup*

61 <https://www.iso.org/fr/standard/68516.html?browse=tc>.

Framework (TMF) pour la conception de ressources terminologiques. Comme l'affirment Romary et al. (2006, p. 3) :

né d'une volonté de définir une plate-forme unifiée de spécification et de représentation de données terminologiques multilingues, le *Terminological Markup Framework* (TMF, [ISO 16642]) permet ainsi de décrire des méta-contraintes pour le marquage terminologique, c'est-à-dire des contraintes structurelles minimales auxquelles doit répondre tout langage de représentation de données terminographiques.

Le TMF définit donc des mécanismes décrivant les contraintes inhérentes à une représentation donnée, faisant abstraction du choix explicite d'implémentation de cette structure. Il s'ensuit que l'ensemble des *Terminological Markup Languages* (TML) exprimés en langage XML⁶² – c'est-à-dire l'ensemble des formats compatibles avec la plate-forme TMF – peut donc être considéré comme une famille dont les conditions d'interopérabilité peuvent être définies de façon rigoureuse (Romary, 2001b, p. 83).

Comme décrit dans la norme ISO-16642 : 2017, le cadre de balisage terminologique TMF comprend deux niveaux d'abstraction :

1. le premier niveau est le plus abstrait et concerne le métamodèle structurel, indépendant de toute implémentation spécifique, qui doit être partagé par toutes les collections de données terminologiques (CDT) conformes à la norme ;
2. le deuxième niveau concerne le modèle de données spécifique pour chaque collection qui est utilisé pour représenter les catégories de données envisagées.

À ce dernier propos, la norme renvoie à un répertoire de catégories de données, *DatCatInfo*⁶³, qui est une base de données en ligne contenant les informations sur les types de données pouvant être incluses dans les CDT et d'autres ressources linguistiques⁶⁴. Pour le moment, nous nous

62 *eXtensible Markup Language*.

63 www.datcatinfo.net.

64 Plus en détail, si aucune catégorie de données appropriée n'est disponible dans ce répertoire, les implémenteurs d'un nouveau TML doivent proposer la compilation d'un nouveau répertoire de données représentatif de leur ressource.

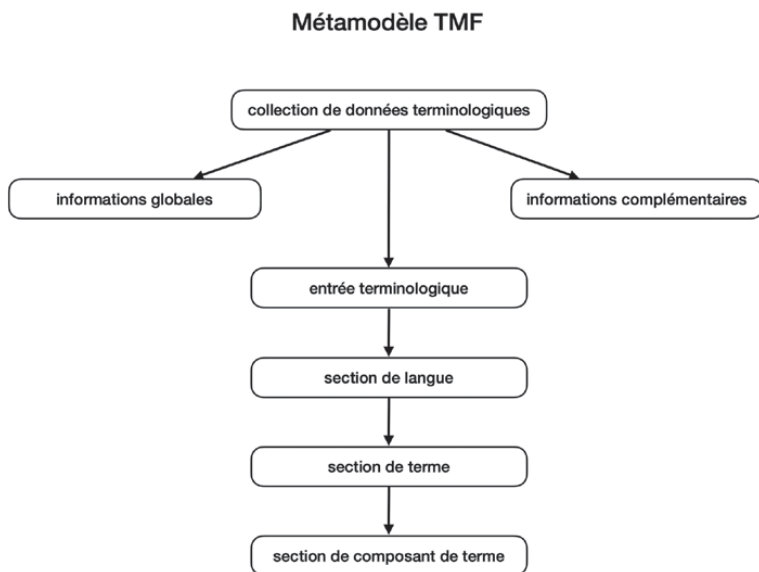


Image 3.2 – Le métamodèle TMF

concentrons sur l'illustration du premier niveau d'abstraction et nous renvoyons le lecteur à la section suivante (3.3.3) pour la description de la représentation et de la gestion des catégories de données.

Dans l'image 3.2 nous illustrons les sept instances qui composent le médamodèle hiérarchique TMF selon la norme de référence :

1. Collection de données terminologiques : cette section représente le niveau le plus haut du métamodèle en tant que macro-conteneur de toutes les informations sur la même collection.
2. Informations globales : cette section contient les informations sur la collection dans son ensemble (ex. nom de la collection et date de la dernière mise à jour).
3. Informations complémentaires : cette section contient généralement des informations de type additionnel (ex. références bibliographiques ou administratives).
4. Entrée terminologique : cette section contient les informations descriptives relatives à un seul concept.

5. Section de langue : cette section sert de conteneur pour la section de terme qui, dans la langue de travail identifiée, désigne le concept ci-dessus. Il existe autant de sections de langue que de langues de travail de la collection de données terminologiques avec lesquelles l'on souhaite exprimer les concepts.
6. Section de terme : cette section contient le terme qui, pour la langue ci-dessus, désigne le concept exprimé dans l'entrée terminologique. Il y a autant de sections de termes que de termes (synonymes entre eux) qui désignent le concept dans cette langue. De plus, la section peut contenir des informations relatives au terme lui-même (ex. partie du discours, et/ou genre et nombre grammatical).
7. Section de composant de terme : dans le cas de termes complexes, par exemple, cette section fournit des informations sur chaque composante du terme en question.

Laissant un instant de côté les deux instances informationnelles (globale et complémentaire) et la dernière instance renvoyant à un niveau de granularité d'analyse qui dépasse le cadre de cet ouvrage, on peut résumer que : une collection de données terminologiques est une ressource contenant un nombre générique n d'entrées terminologiques. Chaque entrée terminologique fait référence à un et un seul concept. Pour chaque entrée terminologique, il y a n sections de langue qui permettent d'exprimer le concept. Pour chaque section de langue, il y a n sections de terme qui, par le biais de cette langue, désignent le concept caractérisant l'entrée terminologique. Comme nous l'avons vu dans la section précédente, les trois entités (concept, langue, terme) et leurs associations sont donc modélisées de manière abstraite et hiérarchique par le standard TMF.

Or, comme nous l'avons dit précédemment, le cadre de balisage TMF comprend non seulement le métamodèle abstrait, mais également le modèle de données qui décrit l'ensemble des catégories de données terminologiques et leurs spécifications pour une ressource terminologique donnée. Les catégories de données et leurs spécifications concernent ce que nous avons précédemment – dans la description du schéma entité-association – défini comme « propriété » des entités impliquées. Dans ce sens, afin qu'une base de données puisse être considérée comme conforme à la norme ISO-16642, elle doit être basée sur :

1. le métamodèle ;
2. la spécification de catégories de données, disponibles à partir d'un répertoire de données, tel que DatCatInfo.

La spécification de catégorie de données contient les restrictions sur les valeurs autorisées d'une catégorie de données, telles que « chaîne de caractères » pour la catégorie /définition/⁶⁵ ou « liste déroulante » pour la catégorie /partie du discours/. Les contraintes incluent également les restrictions sur la position d'une catégorie de données dans la structure hiérarchique TMF. Par exemple, la catégorie de données /partie du discours/ peut être intégrée uniquement – et raisonnablement – au niveau de la section de terme. En termes d'interopérabilité entre deux TML (transfert bidirectionnel d'informations entre deux systèmes informatiques), il faut donc comparer deux aspects : la combinaison de conformité est nécessaire 1) au métamodèle et 2) aux spécifications de catégories de données. Les TML structurés conformément au métamodèle décrit dans cette norme et qui utilisent les mêmes spécifications de catégories de données sont dits interopérables. En d'autres termes, il sera possible de convertir des données d'un TML vers un autre TML sans perte d'informations.

Une fois que les données terminologiques sont structurées selon le métamodèle TMF, nous nous concentrons sur les spécifications des catégories de données qui sont nécessaires afin de répondre aux conditions de conformité décrites précédemment. En termes de complémentarité, l'ISO-16642 : 2017 fait référence à une autre norme, à savoir l'ISO-12620 : 2019, pour la modélisation des catégories de données et de leurs spécifications en tant que propriétés des objets d'intérêt, ce qui fait l'objet de discussion de la section suivante.

3.3.3 Catégories de données

La norme ISO-12620 : 2019⁶⁶ a été formulée dans le but de documenter, d'harmoniser et de gérer les catégories de données (et leurs

65 À des fins de cohérence avec les normes, nous utiliserons les barres obliques inversées lorsqu'indiquant une catégorie de données.

66 <https://www.iso.org/fr/standard/69550.html?browse=tc>.

spécifications) utilisées au sein d'une ressource terminologique. Une catégorie de données est définie comme une “*class of data items that are closely related from a formal or semantic point of view*” (ISO-12620, 2019). Les catégories de données sont donc des classes d'informations généralement fournies au sein des ressources et remplissent le rôle de propriété des entités que l'on veut représenter : par exemple / partie du discours/ pour l'entité-terme ou /identifiant de concept/ pour l'entité-concept. Comme l'explique la norme en question, ces catégories varient selon le type de ressource conçue. À titre d'exemple, la catégorie de données /nombre de significations/ est typique d'une ressource lexicographique et non terminographique. De plus, non seulement les catégories de données peuvent varier selon la ressource, mais aussi les mêmes valeurs que celles-ci admettent. Reprenant l'exemple fourni par la norme :

for terminology management, only a small set of values are needed for /part of speech/ (e.g. noun, verb, adjective, adverb), but in lexicographical resources, many additional values are required (e.g. preposition, pronoun, etc.) (ISO-12620, 2019).

En outre, chaque catégorie de données a sa propre spécification conçue comme une fiche descriptive qui :

provides the complete and formal representation of a data category (for example, its name, definition, examples, comments, etc.) (ISO-12620, 2019).

Or, la norme ISO-12620 intervient justement sur la représentation, la gestion et l'harmonisation de ces éléments. En particulier, ce document vise à promouvoir l'utilisation d'un répertoire de catégories de données librement accessible en ligne comme source commune de consultation contenant les spécifications des mêmes catégories de données afin d'homogénéiser leur représentation au sein de différentes ressources terminologiques. À cet égard, la norme fait référence au répertoire nommé DatCatInfo disponible en ligne au lien www.datcatinfo.net. En 2017, ce répertoire a remplacé le vieil ISOCAT *Data Category Repository* décrit dans la version précédente de la norme ISO 12620 de 2009 et développé et maintenu sous les auspices de l'ISO TC/37, avec l'Institut Max Planck de psycholinguistique de Nimègue, aux Pays-Bas, agissant en

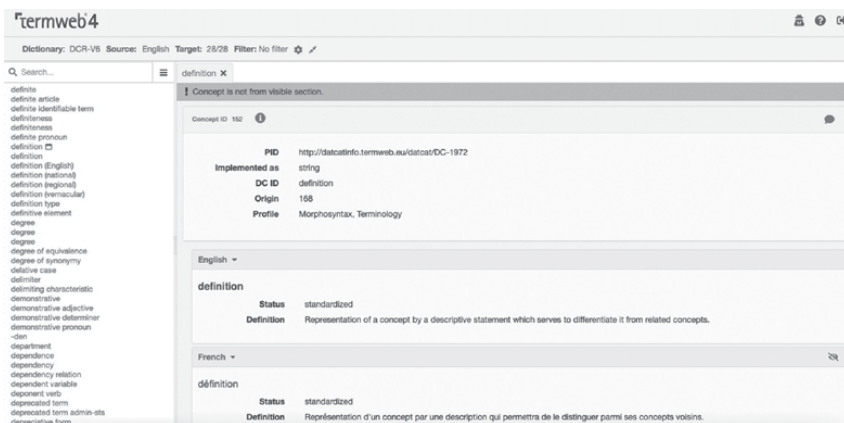


Image 3.3 – Exemple de catégorie de données sur DatCatInfo

tant qu'autorité d'enregistrement (Kemps-Snijders et al., 2008 ; Broeder et al., 2014 ; Windhouwer et Schuurman, 2014).

Depuis le site DatCatInfo il est possible d'accéder à un large éventail de spécifications de catégories de données. Le logiciel qui permet leur consultation s'appelle TermWeb et, dans l'image 3.3⁶⁷, nous fournissons une capture d'écran pour la spécification de la catégorie de données /définition/.

Comme l'illustre l'image, les informations contenues dans la fiche descriptive servent à désambiguïser et à représenter uniformément la catégorie de données : d'une part, la spécification nous fournit des informations relatives au nom canonique attribué à la catégorie et à sa définition afin de l'interpréter correctement ; d'autre part, des indications nous sont fournies pour référencer de manière unique la catégorie, au travers de son identifiant unique et persistant (PID), et pour l'implémenter de manière cohérente parmi les différentes ressources sous la forme de chaîne de caractères.

Les avantages découlant de l'adoption de cette norme par le terminologue se reflètent dans divers aspects liés à la curation des données

67 La capture d'écran a été prise le 18/10/2021 à partir de la page Web suivante : <https://datcatinfo.termweb.eu/search/terms/1802101>.

terminologiques numériques. Premièrement, la norme sensibilise à la cohérence de la dénomination des catégories de données (Wright, 2001), produisant, par conséquent, un impact en termes de meilleure interopérabilité entre les ressources terminologiques. Imaginons que nous voulions intégrer automatiquement les données de deux ressources terminologiques différentes qui présentent toutes deux la catégorie de données /définition/. Si le nom de la catégorie est différent (par exemple *définition* d'une part, et *déf* d'autre part), il serait difficile pour un logiciel d'interpréter les deux noms comme figurant la même catégorie et donc leur intégration ne serait pas automatique, mais demanderait un travail chronophage de formulation manuelle de liens entre les deux éléments. En ce sens, l'adoption du même nom canonique de la catégorie de données et de son interprétation favorise l'interopérabilité des données entre différentes ressources. Deuxièmement, la mise à disposition d'un répertoire de catégories de données contenant les spécifications permet de répondre aux exigences de trouvabilité et d'accessibilité des données. Chaque catégorie de données est en effet associée à un identifiant unique et persistant, est accompagnée d'une description riche, et est indexée au sein d'une ressource consultable. L'accessibilité est garantie à partir du moment où le PID fourni se présente sous la forme d'une URL qui fournit un accès Web direct à la spécification de données dans un répertoire en ligne conforme au protocole de communication standard HTTP. Enfin, tous ces éléments convergent naturellement pour faciliter la réutilisation des catégories de données. Le répertoire disponible sur DatCatInfo voit le jour précisément dans le but de fournir à tous les terminologues en charge de la conception et du développement d'une ressource numérique des informations pouvant être réutilisées de manière homogène. Il s'ensuit que la combinaison de tous ces éléments permet donc de répondre aux besoins préalablement définis à travers la formulation du paradigme de la « terminologie FAIR ».

Mais que se passe-t-il si la catégorie de données que nous voulons représenter dans notre ressource n'est pas disponible sur le répertoire standard ? À cet égard, la norme ISO-12620 prévoit toute une section consacrée à la mise en place d'un répertoire sur mesure qui vise à documenter uniquement les catégories de données non disponibles sur TermWeb ou dans tout autre répertoire. Ce répertoire devra suivre des conditions spécifiques afin de garantir également dans ce cas les

principes de trouvabilité, d'accessibilité, d'interopérabilité et de réutilisabilité des catégories de données⁶⁸.

3.4 Conclusion

La conception d'une ressource terminologique numérique implique donc une phase d'analyse des besoins répartis à différents niveaux et d'un processus de modélisation conceptuelle qui permet de définir les objets d'intérêt et leurs relations. À partir de la description du modèle entité-association, nous avons vu dans ce chapitre comment la norme ISO-16642 : 2017 permet de modéliser et d'organiser les entités concept, terme et langue, tandis que la norme ISO-12620 : 2019 permet de représenter et d'harmoniser les propriétés (catégories de données) choisies pour chaque entité.

À ce stade, nous passons au chapitre suivant de la phase de conception à la phase d'implémentation de la ressource terminologique. En particulier, nous nous concentrons sur la description de la troisième norme précédemment évoquée, illustrant le format de représentation *TermBase eXchange* (TBX) (ISO-30042, 2019), et identifiée comme modèle à suivre afin de favoriser l'adoption d'une approche FAIR de la terminologie. Finalement, nous proposons quelques pistes de réflexion sur les phases de mise à jour et de publication des données terminologiques contenues dans une ressource FAIR.

68 À ce propos, la section 5.5 du chapitre 5 sera consacrée à la description d'un répertoire conçu spécifiquement pour documenter un petit ensemble de catégories de données disponibles sur la ressource TriMED en tant qu'étude de cas d'application du paradigme de la terminologie FAIR.

4 Vers l'implémentation d'une ressource terminologique

What I cannot create, I do not understand.

Richard Feynman (1918–1988)

Ce chapitre est complémentaire au précédent, car il permet de faire évoluer la réflexion de la phase de conception à la phase d'implémentation d'une ressource terminologique. Cependant, une note de clarification s'impose à ce stade de l'exposition. Cette partie de l'ouvrage n'a pas vocation à prendre en considération tous les aspects liés à l'*implémentation* – d'où le choix du titre du chapitre – de la ressource terminologique. Notre intérêt est plutôt orienté vers la description, différemment approfondie, de trois aspects spécifiques de ce processus, à savoir :

1. la phase de représentation des données terminologiques et, conformément à ce qui précède, l'illustration d'un format standard spécifique qui favorise leur réutilisation ;
2. la phase d'actualisation (comprenant l'insertion, la modification et la suppression) des données terminologiques contenues dans la collection ;
3. la phase de publication de ces mêmes données au sein d'un répertoire existant ou conçu *ad hoc*.

Or, comme nous le verrons au fil des pages suivantes, si le premier aspect relève des activités propres du terminologue numérique, les deux derniers points évoqués ci-dessus concernent plutôt de compétences informatiques qui dépassent le rôle préalablement défini pour cette figure. Certains aspects du processus d'implémentation nécessitent en effet de la collaboration avec un expert en informatique qui, de manière synergique, collabore à la phase finale de mise en œuvre

de la base de données numérique. Cette synergie est, à notre avis, un nœud crucial caractérisant le travail du terminologue numérique. En ce sens, le caractère interdisciplinaire de cette figure implique une forte dimension collaborative pour la réalisation optimale du projet de recherche. Le volet interdisciplinaire grandement promu ne veut pas laisser entendre que le terminologue numérique est contraint de s'appropriier des compétences de l'informaticien (et/ou vice versa) pour la mise en œuvre de la base de données ou, par exemple, de l'expert du domaine de travail (le médecin, l'économiste, l'avocat, etc..) dont la terminologie est à l'étude. L'essence de l'interdisciplinarité évoquée réside plutôt dans la capacité du terminologue numérique à savoir communiquer et collaborer efficacement avec les experts des autres disciplines impliquées et à comprendre leurs besoins et spécificités sectorielles.

Pour ces raisons, ce chapitre abordera de manière approfondie l'aspect de représentation des données en introduisant la question à travers l'illustration d'un scénario qui justifie la nécessité d'adopter un format qui assure la réutilisation des données. Par la suite, nous nous concentrons sur le format de représentation *TermBase eXchange* (TBX) en tant que le seul modèle standard conçu pour l'échange de données terminologiques, et donc privilégiant les composantes d'interopérabilité et de réutilisabilité, telles que définies à travers la formulation du paradigme de la « terminologie FAIR » (section 3.2 du chapitre précédent). Enfin, nous décrivons de manière plus générique les aspects d'actualisation et de publication des données afin de tracer des pistes de réflexion qui font l'objet de discussion et d'évaluation en contexte collaboratif.

4.1 Représentation des données terminologiques

L'implémentation de ressources terminologiques et leur maintenance optimale sont des activités laborieuses et chronophages qui nécessitent souvent d'investissements économiques substantiels. La préservation de ces produits est donc un élément clé à la base de tout projet impliquant leur mise en œuvre. En ce sens, l'adoption d'un format de représentation des données terminologiques qui en garantit leur réutilisation

est une pratique essentielle qui guide le travail du terminologue numérique.

De nombreux scénarios peuvent se produire où il est nécessaire de disposer de données terminologiques structurées et normalisées afin de permettre leur exportation et/ou importation (Warburton, 2015). Par exemple, en considérant l'environnement commercial, la publication récente de Warburton (2021) illustre à cet égard de nombreux principes et lignes directrices pour la gestion optimale de la terminologie dans le cadre de collaborations avec des entreprises publiques et privées. En outre, du côté de la traduction spécialisée, l'étude de Schmitz (2012) illustre d'autres cas dans lesquels l'échange de terminologie s'avère être une activité fondamentale. Les activités commerciales de traduction entre clients et traducteurs *freelance* demandent parfois un transfert de données terminologiques dans les deux sens. Par exemple, dans le cas où une entreprise fournit une terminologie pour soutenir un projet de traduction et où les traducteurs renvoient les ressources terminologiques qui ont été mises à jour pendant le travail de traduction.

En outre, l'utilisation cohérente de la terminologie dans des projets plus importants avec plusieurs traducteurs *freelance* ne peut être garantie que si toutes les personnes impliquées ont accès aux mêmes ressources de données terminologiques. Un dernier scénario important mentionné par l'auteur pour l'échange de données terminologiques est la migration des collections de données terminologiques d'un système vers un autre. Cela peut être nécessaire si une entreprise, par exemple, change de fournisseur d'outils et doit convertir la terminologie de l'ancien système de gestion terminologique vers le nouvel outil. En ce sens :

a simple and easy terminology exchange in all these scenarios can be problematic, since the technical requirements for terminology management (languages, data categories, data value sets) can be quite different depending on the needs of the respective user groups and the organizational environments (Schmitz, 2012, p. 33–34).

Pour ces raisons, de nombreuses normes se sont succédé au fil du temps dans le même but de représenter et structurer de manière uniforme les données terminologiques contenues dans les ressources afin de permettre leur importation et/ou exportation. Parmi les formats d'échange qui se sont succédé on peut citer, entre autres : le format électronique

d'échange terminologique (ETIF) (Melby, 1995), le format lisible par machine d'échange terminologique (MARTIF) (Melby et al., 2001), et le format *TermBase Exchange* (TBX) (Melby, 2015)⁶⁹. Ce dernier – dans sa version de 2019 – fait l'objet de discussion de la section suivante et est actuellement configuré comme le seul format standard de représentation des données qui en assure une réutilisation optimale.

4.1.1 Format *TermBase eXchange*

La norme ISO-30042 : 2019 définit le cadre TBX, exprimé en langage de balisage XML, pour l'analyse, la représentation descriptive et la diffusion de données terminologiques structurées. Dans la vision d'une approche FAIR de la terminologie, cette norme est complémentaire aux deux normes précédemment mentionnées – relatives au méta-modèle TMF (ISO-16642 : 2017) et aux catégories de données (ISO-12620 : 2019) – car elle propose des lignes directrices pour la mise en œuvre du format à attribuer aux éléments composant la base de données terminologiques.

Le cadre TBX décrit une famille de formats qui partagent une structure de base commune et une gamme limitée de types d'informations, à savoir de catégories de données. La combinaison d'une structure de base et de modules de catégories de données constitue ce qui, dans la norme, est appelé un « dialecte » de TBX. Le site en ligne [TBXinfo.net](https://www.tbxinfo.net)⁷⁰ fournit de nombreux exemples pour ceux qui souhaitent organiser leurs données en format TBX. Le cadre TBX a pour objectif principal de garantir que les données terminologiques puissent être utilisées dans différents logiciels. Cette séparation entre les données et les logiciels offre des avantages pour les activités de formulation et de traduction, notamment la protection, la cohérence et l'interopérabilité des données. En particulier, comme expliqué dans l'introduction de la norme, le format TBX garantit : 1) l'intégration des données provenant de sources multiples ; 2) la comparaison du contenu de diverses CDT ;

69 Pour une étude approfondie à ce propos nous renvoyons le lecteur à la publication de Lommel et al. (2014) qui illustre, dans la section 2, l'historique des formats d'échange terminologique.

70 <https://www.tbxinfo.net>.

3) l'interrogation de plusieurs CDT à travers une interface commune ;
4) la mise en ligne des données et leur téléchargement ; et, enfin, 5) la dynamique de la terminologie qui sera disponible dans les applications en réseau par l'intermédiaire d'un service Web.

4.1.1.1 Dialectes TBX

La structure du format TBX est constituée de deux composants en interaction :

1. une structure de base qui reflète le modèle de données abstrait du métamodèle TMF (ISO-16642 : 2017) ;
2. un formalisme pour définir des modules contenant une liste de catégories de données (ISO-12620 : 2019).

La combinaison de ces deux composants définit un dialecte particulier, c'est-à-dire un langage de balisage XML conforme à TBX. Les catégories doivent être documentées à partir d'un répertoire de catégories de données en ligne comme nous l'avons décrit précédemment. Deux styles XML peuvent être utilisés pour représenter les catégories de données terminologiques : *Data Category as Attribute* – DCA (catégorie de données en tant qu'attribut) et *Data Category as Tag* – DCT (catégorie de données en tant que balise)⁷¹.

Comme spécifié par la norme en question, les dialectes peuvent différer en ce qui concerne les catégories de données autorisées et les niveaux de métamodèle où ces catégories peuvent être saisies. Principalement, ils sont divisés en deux catégories : publics et privés. Sur le site TBXinfo.net., trois dialectes publics sont décrits et recommandés pour l'échange de la terminologie. Ils sont composés à partir de modules publics et leur maintenance et leur support sont gérés par TBXinfo.net. Les dialectes publics sont :

71 Pour plus d'informations, veuillez consulter la section 6 de la norme et la page <https://www.tbxinfo.net/dca-v-dct/>. Pour le moment, il suffit d'explicitier que dans les deux cas, la valeur de la catégorie de données est le contenu de l'élément XML et que ces deux styles sont isomorphes, c'est-à-dire qu'ils peuvent être convertis dans les deux sens par un algorithme sans perte d'information.

1. **TBX-Core** : ce dialecte est le plus simple et comprend les catégories de données extraites du module *Core*. Le module *Core* est le module de base pour tous les fichiers TBX et contient la structure principale de TBX telle que décrite dans la norme ISO 30042 et les catégories de données /term/, /date/ et /note/⁷².
2. **TBX-Min** : ce dialecte TBX est conçu pour le stockage simple et direct de glossaires bilingues ou monolingues. Les catégories de données se basent sur le module *Min* et comprennent /part of speech/, /subject field/, /external cross reference/ et /administrative status/.
3. **TBX-Basic** : c'est le dialecte principal de l'échange de terminologie. TBX-Basic est conçu pour stocker efficacement un grand nombre de glossaires terminologiques dans un format XML simple. Il peut être utilisé pour gérer des glossaires monolingues, bilingues ou multilingues. Les catégories de données qui font partie du module *Basic* sont /context/, /definition/, /cross reference/, /grammatical gender/, /source/, /responsibility/, /term type/, /xGraphic/, /transaction type/, /term location/, /project subset/ et /geographical usage/. Ce dialecte est le plus riche et le plus structuré et comprend les modules *Core*, *Min*, *Basic*.

Les dialectes privés, en revanche, ne sont pas gérés par TBXinfo.net. Toutefois, une liste de dialectes privés connus est fournie en tant que service aux utilisateurs⁷³. Par conséquent, TBXinfo.net ne peut pas garantir la stabilité ou la fiabilité des dialectes privés. L'organisme offre cependant la possibilité de valider les dialectes privés conçus selon les indications fournies à partir de la page dédiée.

4.1.1.2 Conception d'un dialecte

Les raisons de la conception d'un nouveau dialecte TBX reposent sur la volonté de l'implémenteur d'utiliser des catégories de données qui ne sont pas présentes dans le module de base *Core*. Tous les dialectes conçus *ad hoc* doivent donc être configurés comme une extension du

72 En conformité avec la norme, les catégories de données sont en anglais.

73 <https://www.tbxinfo.net/tbx-private-dialects/>

module *Core* décrit précédemment, qui contient les catégories de données essentielles (/term/, /date/, /note/). Dans ce sens, les nouveaux dialectes TBX étendent la structure de base en ajoutant un ensemble de modules de catégories de données spécifiquement choisies aux fins des exigences de développement de la ressource numérique. TBX est donc configuré comme un format hautement flexible, car il permet aux groupes d'utilisateurs de sélectionner leurs catégories de données et de formuler leur propre dialecte adapté à leurs besoins.

Un nouveau dialecte TBX doit d'abord suivre des règles de dénomination. En particulier, il faut que le nom du dialecte commence par le préfixe « TBX- » et se termine par un indicateur du nouveau module qui est configuré comme une extension du module de base *Core*. À titre d'exemple, supposons que nous voulons concevoir un nouveau dialecte TBX composé du module de base *Core* et d'un nouveau module arbitraire *Multilang*. Dans ce cas, le nom du dialecte sera donc « TBX-Multilang ».

Une fois le nom du dialecte défini, l'implémenteur procède à la formulation du nouveau module TBX. Comme indiqué dans la norme (ISO-30042, 2019, p. 19), la définition d'un module est un processus qui se compose de 4 étapes :

1. dénomination du module ;
2. sélection d'une ou plusieurs catégories de données ;
3. définition des propriétés de la catégorie de données ;
4. établissement des contraintes de la catégorie de données.

Le nom du module, qui ne suit pas les conventions de dénomination des dialectes TBX, doit être unique parmi les modules publics. Par conséquent, en reprenant l'exemple proposé plus haut, un module peut être nommé *Multilang*, mais pas « TBX-Multilang ». De plus, il n'est pas nécessaire que les noms de modules reflètent le nom du dialecte TBX principal pour lequel ils ont été conçus. Cela signifie que le module *Multilang* aurait pu être nommé *Fiction*, même s'il s'agit du module de base du dialecte « TBX-Multilang ». Les catégories de données qui font partie du nouveau module sont sélectionnées à partir d'un répertoire de catégories de données en ligne (par exemple le répertoire *DatCatInfo* : voir sous-section 3.3.3 du chapitre précédent). Supposons que le nouveau module *Multilang* ajoute aux catégories essentielles du module *Core*, les éléments : /subject field/, /part of speech/ et /grammatical gender/. À

Dénomination	PID	Classification	Valeur	Niveau
subjectField	http://datcatinfo.termweb.eu/datcat/DC-489	<descrip>	chaîne de caractères	conceptEntry, langSec
partOfSpeech	http://datcatinfo.termweb.eu/datcat/DC-396	<termNote>	liste de sélection	termSec
grammaticalGender	http://datcatinfo.termweb.eu/datcat/DC-245	<termNote>	liste de sélection	termSec

Image 4.1 – Exemple de TBXMD du module *Multilang*

ce propos, il est nécessaire de sauvegarder le critère d'unicité des catégories de données : cela signifie que 1) les modules TBX conçus *ad hoc* ne doivent pas dupliquer les catégories déjà existantes, et 2) si plusieurs modules TBX sont utilisés dans le même dialecte, ils ne doivent partager aucune catégorie de données. Une fois sélectionnées, les catégories, leurs propriétés et leurs contraintes doivent être documentées et définies. À ce propos, pour chaque catégorie il faut expliciter :

- l'identifiant mnémonique ;
- le PID, c'est-à-dire un identifiant permanent unique pour chaque catégorie de données qui pointe vers son entrée sur un RCD ;
- la classification, c'est-à-dire la classe de données terminologiques qui peut être de type descriptif (<descrip>), administratif (<admin>), etc.
- la valeur de la catégorie de données, à savoir le contenu autorisé pour chaque catégorie de données comme une liste de valeurs au choix entre « nom », « verbe », « adjectif », etc.
- le niveau du métamodèle TMF où la catégorie de données est autorisée, à savoir le niveau du concept (conceptEntry), de la langue (langSec) ou du terme (termSec)⁷⁴.

La définition formelle d'un module doit être écrite en prose à travers le formalisme TBX *Module Description* (TBXMD) qui peut également être utilisé pour écrire une description exploitable par machine. Dans l'image 4.1, nous illustrons une capture d'écran d'un formalisme TBXMD sommaire du module *Multilang*⁷⁵. Une fois le dialecte

74 La section 8.3 de la norme en question (pages 12–14) décrit en détail les contraintes de positionnement des catégories de données. À titre d'exemple, les catégories de données appartenant à la classe <termNote> ne peuvent être positionnées qu'au niveau de la section de terme.

75 Le TBXMD doit également contenir d'autres informations de définition qui ne sont pas illustrées dans l'image. Pour ce niveau de détail, nous renvoyons

fictif « TBX-Multilang » défini, nous proposons ci-dessous un exemple d’instance :

```
<?xml version="1.0" encoding="UTF 8"?>
<tbx xmlns:tbx3="urn:iso:std:iso:30042:ed 2"
type="TBX Multilang" style="dct" xml:lang="en">
  <tbxHeader>
    <fileDesc>
      <sourceDesc>
        <p>A sample multilingual termbase
        consisting of one terminological entry</p>
      </sourceDesc>
    </fileDesc>
  </tbxHeader>

  <text>
    <body>
      <conceptEntry id="C7">
        <date>23/10/2021</date>
        <note>NA</note>
        <multilang:subjectField>medicine</subjectField>

        <langSec xml:lang="en">
          <termSec>
            <term>fever</term>
            <multilang:partOfSpeech>noun
            </multilang:partOfSpeech>
            <multilang:grammaticalGender>NA
            </multilang:grammaticalGender>
          </termSec>
        </langSec>
      </conceptEntry>
    </body>
  </text>
</tbx>
```

le lecteur à consulter la norme ISO-30042 : 2019 ou aux exemples de modules publics disponibles sur : <https://www.tbxinfo.net/tbx-modules/?id=1>.

```

    <langSec xml:lang="it">
<termSec>
    <term>febbre</term>
    <multilang:partOfSpeech>nome
</multilang:partOfSpeech>
    <multilang:grammaticalGender>femminile
</multilang:grammaticalGender>
</termSec>
</langSec>

    <langSec xml:lang="fr">
<termSec>
    <term>fièvre</term>
    <multilang:partOfSpeech>nom
</multilang:partOfSpeech>
    <multilang:grammaticalGender>feminin
</multilang:grammaticalGender>
</termSec>
</langSec>

</conceptEntry>
</body>
</text>
</tbx>

```

Dans l'élément racine <tbx>, il faut spécifier :

1. l'espace de nom (xmlns) ;
2. la valeur de l'attribut type, c'est-à-dire le nom du dialecte TBX ;
3. le style de l'instance (DCA ou DCT) ;
4. la langue de travail (xml:lang) du document, dans ce cas l'anglais.

Ce dernier attribut ne doit pas être confondu avec l'élément <langSec> : la définition de la langue de travail dans la racine du document indique la langue avec laquelle le fichier TBX est écrit (par exemple les balises), tandis que l'élément <langSec> spécifie la langue objet de description de la section (et donc le contenu des balises).

TBX-Multilang représente donc un exemple de base de données multilingue composée d'une seule entrée terminologique, à savoir un seul concept. L'instance de document suit le métamodèle hiérarchique de données TMF : dans l'ordre, <conceptEntry>, <langSec> et <termSec> sont définis. Au niveau du concept, représenté par son identifiant (id), les catégories de données /date/, /note/ (module *Core*) et /subject field/ (module *Multilang*) ont été placées. Ensuite, le même concept est exprimé en trois sections <langSec> pour l'anglais, l'italien et le français. Chaque <langSec> contient une <termSec> avec le terme qui, pour cette langue donnée, désigne le concept et d'autres catégories de données (/part of speech//grammatical gender/) appartenant au module *Multilang*.

En guise de conclusion, on peut affirmer que les trois niveaux de la structure hiérarchique TMF (<conceptEntry>, <langSec> et <termSec>) se retrouvent dans presque toutes les collections de données terminologiques. Cependant, comme nous l'avons vu dans la section 3.3.2 du chapitre précédent, le niveau supplémentaire <termCompSec> peut apparaître dans les collections dont la finalité est également la description des composantes de termes complexes. Ce niveau de composant de terme est implémenté via un module dédié. Pour plus d'informations, nous renvoyons le lecteur au lien <https://www.tbxinfo.net/tbx-modules/?id=3>.

4.2 Actualisation des données terminologiques

Une fois la phase de structuration des entrées terminologiques achevée selon le format TBX, nous nous concentrons dans cette section sur la phase suivante de compilation des données qui vont peupler chaque niveau hiérarchique (concept, langue, terme) de la fiche⁷⁶. Quel que soit

76 Nous soulignons que dorénavant nous utiliserons indifféremment les termes « entrée terminologique » et « fiche terminologique » se référant tous deux au conteneur qui collecte toutes les données qui peuvent être saisies au niveau du concept, de la langue et du terme.

le domaine de travail du projet terminologique, une réflexion s'impose sur le choix du logiciel pour effectuer les opérations d'actualisation (à savoir d'insertion, de modification, de suppression) des données au sein d'une fiche terminologique structurée. Comme nous l'avons vu, un fichier TBX est un document de type textuel. Cependant, l'utilisation d'éditeurs de texte pour la compilation de fiches terminologiques s'avère difficile à divers points de vue.

Le premier aspect auquel nous prêtons attention est celui relatif au contrôle et à la validation des données saisies. La structure du document TBX définie ci-dessus est généralement accompagnée d'un ensemble de fichiers d'accompagnement⁷⁷ visant à valider le document qui en garantissent une structure correcte (par exemple une catégorie de données ne peut pas être positionnée à un autre niveau TMF que celui pour lequel elle a été définie), et des valeurs uniformes (par exemple, les valeurs de la catégorie /part of speech/ ne peuvent pas apparaître sous forme de séquence de caractères libre, mais doivent être choisies à partir d'un ensemble prédéfini). Ce premier point exclut *a priori* l'utilisation d'un simple éditeur de texte qui risquerait d'invalider le processus de cohérence des données garanti par la norme ISO. À cet égard, la documentation de référence disponible en ligne⁷⁸ mentionne l'utilisation du logiciel Oxygen⁷⁹ pour la mise à jour et la validation des fichiers TBX en tant que logiciel optimisé à ces fins.

Cependant, il existe un deuxième aspect lié à des problèmes d'efficacité : une base de données terminologiques TBX suffisamment volumineuse (quelques centaines de MB) ne permet pas à l'éditeur de texte d'afficher rapidement les modifications apportées au fichier. Même les meilleurs logiciels optimisés pour différents types de fichiers XML imposent des limitations sur les fonctionnalités disponibles lors de l'ouverture de fichiers volumineux : par exemple, l'affichage et la mise en forme du texte sont limités, la validation XML est désactivée ainsi que

77 Dans l'annexe 3 de la norme en question (pages 39–42) référence est faite au langage de description Relax NG (*Regular Language for XML Next Generation*) qui permet de définir les différentes contraintes qui déterminent la classe des documents XML à des fins de validation.

78 <https://www.tbxinfo.net/validating-a-tbx-file/>.

79 <https://www.oxygenxml.com>.

la fonction de recherche et de remplacement et l'opération d'annulation d'une modification⁸⁰. Par conséquent, le choix du logiciel proposé par la documentation officielle de TBX devient en pratique difficilement utilisable (voire inutilisable) lorsque l'on travaille avec de grandes quantités de données, entravant ainsi les mêmes principes de cohérence et de validation des données.

Une dernière réflexion que nous avançons – toujours de manière générale – porte sur le fait que dans un projet de base de données terminologiques collaborative impliquant plusieurs utilisateurs, il faudrait aussi prévoir l'attribution de rôles entre les différents collaborateurs de la ressource. Par exemple, il pourrait être nécessaire de distinguer les utilisateurs « administrateurs » qui vérifient la qualité des fiches et valident leur contenu, d'autres utilisateurs « compilateurs » qui ont pour seule tâche celle de saisir des données et non de vérifier/valider le travail des collègues.

Pour toutes ces raisons, pour les projets de bases de données terminologiques qui impliquent l'insertion de grandes quantités de données, une meilleure alternative à l'utilisation d'éditeurs de fichiers XML est de concevoir et de mettre en œuvre une application *ad hoc* qui permet de travailler de manière plus efficace et répartie sur plusieurs utilisateurs. À ce stade le terminologue numérique devra solliciter une collaboration avec un collègue expert en conception de bases de données, afin de fournir toutes les informations nécessaires à la création d'une application qui permet l'actualisation de la base de données efficacement par différents utilisateurs, et avec différentes fonctionnalités selon sur le type d'utilisateur. C'est précisément ici qu'entre en jeu l'aspect interdisciplinaire du terminologue numérique, qui ne doit pas savoir développer ce logiciel, mais doit être capable de fournir au développeur les conditions préalables nécessaires pour conformer la base de données (de tout type, par exemple relationnelle ou orientée graphe)⁸¹ selon

80 À ce propos, le lecteur est renvoyé à la documentation Oxygen disponible au lien suivant : <https://www.oxygenxml.com/doc/versions/24.0/ug-editor/topics/huge-file-editor.html>.

81 Pour une étude approfondie à ce propos nous renvoyons le lecteur à l'article de Roy-Hubara et Sturm (2020) qui propose une revue systématique de la littérature sur les différents types de bases de données.

les normes ISO de référence. Dans cette perspective, un projet pilote mené en collaboration entre le Département d'études linguistiques et littéraires et le Département d'ingénierie de l'information de l'Université de Padoue a conduit au développement d'une Application Web appelée FAIRterm, pour faciliter la compilation de fiches terminologiques multilingues à travers une interface graphique conviviale qui permet d'exporter les données directement au format TBX standard (Vezzani, 2021). L'application Web FAIRterm fait partie de l'initiative susmentionnée de la « terminologie FAIR » et est accessible en ligne à partir du lien suivant : <http://purl.org/fairterm>. À l'heure actuelle, l'application sert de ressource de référence pour la compilation de fiches terminologiques dans le cadre du projet européen « Terminologie sans frontières » de l'Unité de Coordination de la Terminologie (Term-Coord)⁸² du Parlement européen et constitue l'un des cas d'utilisation de la task force *FAIR Metrics and Data Quality* de l'*European Open Science Cloud Association*⁸³.

Pour conclure, la prise en compte de tous ces aspects conduit également à l'investigation des questions relatives à comment et où les données de la ressource sont publiées, ce qui est le sujet de discussion de la section suivante.

4.3 Publication des données terminologiques

Une fois les données terminologiques actualisées, revenons à la discussion liée au processus de FAIRisation des données, en s'intéressant notamment à sa dernière étape, à savoir celle relative à la publication du jeu de données⁸⁴. En effet, pour être réutilisées, les données doivent également être stockées dans un espace qui permet de les retrouver et de les accéder à partir d'autres ressources et programmes. Le lieu

82 <https://yourterm.org/toolbox/>.

83 https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf.

84 <https://www.go-fair.org/fair-principles/fairification-process/>.

virtuel où les données peuvent être diffusées est généralement appelé « répertoire » et consiste en un ensemble de services qui permettent la conservation et l'accès aux données de la recherche⁸⁵.

Un répertoire de données est un élément fondamental de la science ouverte qui va au-delà de l'idée de partager des fichiers individuels (par exemple le fichier TBX contenant l'ensemble de la collection de données terminologiques) qui peuvent être téléchargés à partir de pages Web ou d'archives Cloud. Un répertoire de données remplit un ensemble de fonctions spécifiques nécessaires à la conservation de données FAIR⁸⁶, par exemple :

- conserver les données pendant de longues périodes, avec des politiques de sauvegarde et de migration des données pour un stockage correct ;
- attribuer des identifiants uniques aux ensembles de données déposés ;
- vérifier la qualité des données et des métadonnées ;
- gérer l'accès aux données par les utilisateurs et attribuer des licences d'utilisation.

L'un des exemples les plus importants de répertoire de données générique – non spécialisé pour un domaine scientifique particulier – est Zenodo⁸⁷, géré par le CERN⁸⁸ pour le compte d'OpenAIRE⁸⁹. Zenodo permet aux chercheurs et aux institutions de partager, de préserver et d'afficher gratuitement les résultats de la recherche multidisciplinaire. Alternativement, en revenant au dernier paragraphe de la section précédente, il est possible de développer son propre répertoire de données spécifique garantissant les fonctionnalités mentionnées précédemment. En ce sens, l'European Open Science Cloud⁹⁰ promeut, à travers ses plateformes, une fédération d'infrastructures visant à rendre l'accès

85 <https://www.fairsfair.eu/articles-publications/boosting-fair-data-adoption-eu-outcomes-fairsfair-week-2021>.

86 <https://www.nature.com/sdata/policies/repositories>.

87 <https://zenodo.org>.

88 <https://home.cern>.

89 <https://www.openaire.eu/>.

90 <https://eosc-portal.eu>.

aux données scientifiques plus efficace et plus facile. À cet égard, le « marketplace EOSC »⁹¹ devient un point d'accès pour les « prestataires » (*providers*) certifiés par EOSC lui-même. Un exemple parmi tant d'autres est celui de l'*European Language Social Science Thesaurus* (ELSST) qui met à disposition les données d'un thésaurus multilingue en 14 langues⁹². Ce prestataire fournit le répertoire de données et propose des fonctions de consultation et de réutilisation des données non seulement pour l'utilisateur humain, mais aussi pour l'utilisateur machine.

4.4 Conclusion

Ce chapitre avait pour objectif d'aborder, à différents niveaux de granularité, certains aspects fondamentaux pour la mise en œuvre d'une ressource terminologique numérique : c'est-à-dire la représentation, l'actualisation et la publication de données terminologiques. Afin de concrétiser ce qui a été évoqué jusqu'ici, nous proposons dans le chapitre suivant la description d'une étude de cas spécifique pour la conception et l'implémentation d'une ressource terminologique multilingue et polyvalente pour le domaine médical formulée sur la base du paradigme de la « terminologie FAIR ».

91 <https://marketplace.eosc-portal.eu>.

92 <https://elsst.cessda.eu/about/>.

5 Étude de cas : la ressource TriMED

Je ne sais si cela se peut ; mais je sais bien que cela est.

Molière, L'Amour Médecin (1655)

Les principes formulés à travers le paradigme de la « terminologie FAIR » peuvent s'adapter à n'importe quel domaine de travail de tout(e) terminologue/terminographe numérique. Ce chapitre sera consacré à la description d'une étude de cas spécifique concernant la conception et l'implémentation d'une ressource terminologique FAIR pour le domaine médical⁹³. La ressource nommée TriMED est une base de données terminologiques multilingue et polyvalente conçue pour l'utilisation de différentes catégories d'utilisateurs impliqués, selon des perspectives différentes, dans le contexte médico-sanitaire (Vezzani et Di Nunzio, 2020b,a).

En particulier, nous décrivons les raisons qui ont conduit à choisir la terminologie médicale comme domaine de travail dans une perspective intra- et inter-linguistique. Par la suite, une analyse des exigences de potentiels utilisateurs de la ressource sera proposée. Enfin, un grand espace sera consacré à la description des phases de conception et d'implémentation de la ressource, ainsi que le développement de son interface Web et la présentation des données qui y figurent et qui sont actuellement disponibles pour consultation.

93 L'étude de cas présentée a été développée dans le cadre d'un projet de recherche doctoral de 3 ans (2017–2020) mené au Département d'études linguistiques et littéraires de l'Université de Padoue. Ce chapitre rassemble, synthétise et retravaille une série d'articles publiés depuis le début du projet (Di Nunzio et Vezzani, 2021 ; Vezzani et Di Nunzio, 2020b, 2019b ; Vezzani et al., 2018). En particulier, nous réutilisons et développons des parties du matériel publié en anglais dans Vezzani et Di Nunzio (2020a) : <https://www.jbe-platform.com/content/journals/10.1075/term.00053.vez>

5.1 Langue médicale

La langue médicale est une « langue de spécialité » au plein sens du terme. Cette expression désigne les langues utilisées dans des situations de communication (orales ou écrites) qui impliquent la transmission d'une information relevant d'un champ d'expérience particulier (Gotti, 1991 ; Cortelazzo, 1990 ; Bianco, 2002). D'un point de vue plus pragmatique, la notion de langue spécialisée désigne une langue naturelle considérée en tant que vecteur de connaissances spécialisées. En tant que telle, la langue médicale devrait présenter des caractéristiques spécifiques vérifiables dans ses productions. Serianni (2003) identifie un premier trait distinctif dans la mono-référentialité, à savoir la caractéristique d'univocité entre le signe et le référent qui implique la non-utilisation de l'homonymie et de la synonymie, contrairement à l'attitude de la langue commune où la désignation d'un concept par le même mot est presque considérée comme une forme de déficit linguistique. Le deuxième trait devrait être la neutralité émotionnelle, selon laquelle le ton du texte est neutre et objectif, privé de toute hédonicité, accordant donc plus d'importance à l'aspect dénotatif qu'à l'aspect connotatif. En outre, Gotti (1991) introduit d'autres aspects comme 1) la précision, selon laquelle chaque terme devrait exprimer son concept de façon immédiate ; 2) la concision, selon laquelle les concepts sont exprimés dans la forme la plus brève possible ; et enfin, 3) la transparence, par laquelle le lexique des langues spécialisées devrait permettre un décodage rapide du sens d'un terme grâce à l'analyse de sa forme de surface. Or, d'un point de vue strictement terminologique, nous pouvons affirmer, en toute objectivité, que les traits distinctifs susmentionnés sont souvent des composants souhaitables – en vertu d'une communication efficace – plutôt que des manifestations réelles du discours médical. Dans les sous-sections suivantes, nous fournissons un aperçu de certains problèmes découlant de la complexité de la terminologie médicale qui se reflètent, en particulier, sur deux fronts : 1) d'un point de vue intra-linguistique, nous décrirons les difficultés qui se posent dans la communication entre le médecin et le patient ; 2) dans une perspective inter-linguistique, nous allons plutôt nous concentrer sur les défis qui se posent au traducteur scientifique.

5.1.1 Perspective intra-linguistique

La communication dans le domaine médical est souvent caractérisée par l'utilisation d'une terminologie hautement spécialisée et difficile à comprendre pour les profanes. Habituellement, les professionnels de la santé tendent à utiliser leur propre jargon, connoté négativement de *medicalese*, en tant que composante du discours spécialisé de la médecine (Hadden et al., 2018). Par conséquent, les non-initiés à ce domaine, par exemple les patients, sont souvent exposés à des termes médicaux qui peuvent être sémantiquement complexes et difficiles à comprendre : parmi les différents auteurs qui traitent du sujet sur la scène internationale, voir par exemple Serianni (2005) ; Castro et al. (2007) ; Balliu (2010) ; Vecchiato et Gerolimich (2013) ; Magris (2016). Dans ce contexte, de nombreuses études se concentrent sur la complexité de la langue médicale en termes de « compréhensibilité » et de « lisibilité » des informations de santé contenues, par exemple, dans les notices⁹⁴ de médicaments (Patel et al., 2002) ou dans les sites Web (McCray, 2005 ; Jucks et Bromme, 2007 ; Tran et al., 2009). La différence entre la terminologie utilisée par les professionnels de la santé et la langue utilisée par les patients est donc citée comme une source de « mauvaise communication » (Elhadad et Sutaria, 2007). En effet, le domaine médical rassemble des personnes de différents statuts sociaux et de différents niveaux d'expertise qui éprouvent un crucial besoin d'interagir et de se comprendre de manière simple et efficace.

Du point de vue terminologique, les problèmes de communication dans le dialogue (verbal ou textuel) entre le patient et le médecin sont strictement liés au phénomène de la variation terminologique, c'est-à-dire des différentes représentations/manifestations d'un même signe linguistique. À ce propos, de nombreuses études se concentrent sur différents aspects et types de variations de termes qui surviennent fréquemment dans le discours médical, tels que, entre autres, les variantes

94 Il est intéressant de noter qu'en italien, les notices des médicaments sont communément appelées *bugiardini* (mot dérivé de l'adjectif *bugiardo*, menteur en français) avec une connotation ironique qui visait à souligner, en particulier pendant les années de boom de la pharmacologie, la tendance à négliger les effets indésirables du médicament pour en mettre en relief les seules qualités : <https://bit.ly/2YYtcQj>.

réversibles (Bowker et Hawkins, 2006) ; les synonymes, les variantes flexionnelles et dérivatives (Llanos et al., 2016) ; et les variantes lexicales (Delavigne, 2017). L'une des manifestations les plus importantes de variation terminologique qui se produit dans le domaine médical est la variation de registre, ou variation diastratique (Coseriu, 1969). Selon la proposition de Freixa (2006) sur les typologies de la variation dénomminative, ce phénomène est attribuable à des causes fonctionnelles lorsque le niveau de spécialisation linguistique des interlocuteurs doit être adapté :

in medicine, denominative variation according to patients' level of comprehension is a common occurrence. The opposite intention has also been observed, i.e., when the specialist does not want the patient to understand what he says. In this case we could speak of "perverted adequacy" (Freixa, 2006).

De plus, l'étude de Picton et Dury (2017) montre que le phénomène de la variation diastratique se manifeste également entre des communautés d'experts qui, bien que travaillant dans le même domaine, ont leur propre micro-langue de spécialité.

Dans ce contexte, de nombreuses ressources ont été spécifiquement implémentées pour gérer la variation terminologique dans le domaine médical et, en particulier, pour combler les différences de registre qui se manifestent dans le dialogue médecin-patient. Ces ressources sont conçues à l'intention des non-experts, pour faciliter la communication et la transmission d'informations dans le domaine de la santé. L'objectif est de favoriser la compréhension des termes médicaux plutôt que d'en normaliser la signification. Par exemple, le *Consumer Health Vocabulary Initiative* (Zeng et Tse, 2006) a conduit au développement d'une ressource utile⁹⁵, pour la langue anglaise, qui permet de traduire des termes techniques dans une langue populaire. En parallèle, l'initiative de Cardillo et al. (2009) a conduit à l'implémentation de la ressource *Italian Consumer-oriented Medical Vocabulary*⁹⁶ qui est un dictionnaire italien montrant la différente façon dont les patients et les professionnels de santé en général expriment les thèmes de la santé. Dans ce contexte, le Ministère espagnol de l'Économie et de la Compétitivité

95 <http://bit.ly/2L3tSw4>.

96 <http://bit.ly/2L31Fpp>.

a financé un autre projet qui a également mené au développement de la base de données terminologiques *VariMed*⁹⁷ (Variation dénomminative en médecine. Ressource multifonctionnelle plurilingue pour la recherche et la vulgarisation) : cette ressource se concentre sur la variation terminologique dans le domaine médical enrichie d'informations pragmatiques et vise à faciliter la communication entre les professionnels de la santé et les patients (Tercedor Sánchez et Prieto Velasco, 2013). Un autre projet disponible pour la consultation des patients est le moteur de recherche *CureHunter*⁹⁸ : cet index numérique intègre tous les médicaments connus, agents biologiquement actifs, maladies et données empiriques déclarations de tous les résultats cliniques efficaces publiés dans la Bibliothèque nationale de médecine des États-Unis⁹⁹. La base de données est conçue pour mettre l'accent sur la définition, la variation et les relations entre traitements-maladies, et elle fournit des documents faciles à lire pour chaque maladie.

D'un point de vue multilingue, plusieurs initiatives ont été développées au niveau européen. La Commission européenne (DG III) a commandé le Glossaire multilingue des termes médicaux populaires et techniques¹⁰⁰. Le projet a été mis en œuvre par le Département de linguistique appliquée du Heymans Institute of Pharmacologie et Mercator School en 1995–2000. La ressource regroupe neuf glossaires de 1830 termes médicaux scientifiques et populaires sur les notices de médicaments dans neuf langues officielles de l'Union européenne. En 2018, l'Unité de Coordination Terminologique (TermCoord) du Parlement européen, en partenariat avec de nombreuses universités européennes, s'est engagée dans le projet *YourTerm MED*¹⁰¹ consistant à réaliser une base de données terminologiques fondée sur la structure cognitive de la Terminologie des Cadres (*Frame-based Terminology*) (Faber, 2015). Cet outil multilingue est conçu pour faciliter la communication entre les professionnels de santé en mission et leurs patients. En particulier, il s'efforce de répondre aux besoins terminologiques réels de Médecins

97 <http://varimed.ugr.es>.

98 <http://www.curehunter.com>.

99 <https://www.nlm.nih.gov>.

100 <http://bit.ly/2UEMxxK>.

101 <http://bit.ly/2VqClO0>.

Sans Frontières (MSF) lors de leurs missions internationales, afin que les médecins puissent consulter sur place une ressource multilingue qui permette de prendre en charge les patients.

La perspective intra-linguistique de la langue médicale reflète donc le problème de la communication entre médecin et patient en termes de variation diastatique de la terminologie. Dans la section suivante, nous nous concentrerons plutôt sur les défis auxquels se confronte la traduction scientifique dans une perspective inter-linguistique.

5.1.2 *Perspective inter-linguistique*

La complexité de la langue médicale affecte non seulement la communication intra-linguistique, mais également la transmission inter-linguistique d'informations médicales, c'est-à-dire la traduction de documents médicaux d'une langue source vers une langue cible (Fischbach, 1962). Les traducteurs scientifiques, en tant qu'experts linguistiques et non professionnels de la santé, sont engagés dans la tâche complexe de transmettre fidèlement les informations médicales dans différentes langues en respectant une terminologie appropriée, le niveau de spécialisation du registre linguistique et la structure syntaxique de la langue cible (Montalt, 2011 ; Montalt et Gonzalez-Davies, 2014). Dans son article, Maurice Rouleau (1995) affirme :

comme une langue de spécialité naît du besoin que ressentent les spécialistes de communiquer entre eux de façon concise et sans ambiguïté, elle est forcément réservée aux initiés, c'est-à-dire aux spécialistes. Pour maîtriser cette langue, le traducteur ne doit pas seulement connaître les termes ; il lui faut savoir les enchaîner dans un discours cohérent qui reflète les usages du domaine.

La formation du traducteur est donc un aspect fondamental : de nombreuses études proposent différentes méthodologies, du point de vue didactique, pour l'exécution efficace d'une traduction dans le domaine médical (Wakabayashi, 1996 ; Lee-Jahnke, 2001 ; Vandaele, 2001 ; Balliu, 2005). En général, le processus de traduction médicale est divisé en deux étapes : la phase de décodage et la phase de transcodage du texte (Jammal, 1999). Le décodage implique une démarche documentaire méthodique, une prise de conscience de certains aspects particuliers qui pourraient faire obstacle à la compréhension du texte et, enfin, une

recherche terminologique systématique en langue source et cible. Le transcodage consiste plutôt dans la formulation et l'adaptation d'un texte cible en respectant les contraintes imposées par deux aspects : les destinataires et le type discursif du texte.

Du point de vue strictement terminologique, Rouleau (2003) identifie certains facteurs à la base des difficultés propres à la langue médicale, comme l'usage, le changement de terminologie, la synonymie, la polysémie et l'éponymie. L'usage désigne, par exemple, la façon dont les médecins s'expriment habituellement et donc l'utilisation particulière de ce type de langue de spécialité. Par exemple, la construction de l'adjectif en hypallage est d'usage fréquent : l'adjectif s'accorde syntaxiquement à un mot présent, mais qui est associé à un terme absent. Un exemple pourrait être le syntagme « circulation veineuse » : il est bien évident que ce ne sont pas les veines qui circulent, mais le sang qui circule *dans* les veines. L'utilisation de l'hypallage est fréquente dans le discours médical : « diabète sucré », « cure ambulatoire » ou bien « fièvre pré-tibiale ». De même, certains cas lexicaux s'imposent par leur fréquence d'usage : un médecin parlera d'« infarctus du myocarde », utilisant le groupe propositionnel au lieu de l'adjectif relationnel « myocardique », et inversement, d'« adhérence cicatricielle » et pas d'« adhérence des cicatrices », même si, du point de vue syntagmatique, les deux versions sont possibles dans les deux cas¹⁰².

Le deuxième facteur identifié par l'auteur est le changement de terminologie. L'auteur fait référence aux œuvres *Nomina Anatomica* et *Terminologica Anatomica* qui naissent d'un besoin d'uniformisation de la terminologie. Cette exigence se voulait une réponse à la difficulté que rencontraient les anatomistes de différents pays (et donc de différentes langues) à communiquer entre eux. Toutefois, l'imposition d'un terme ne reçoit pas toujours l'accueil souhaité et le souci d'uniformisation s'est doublé d'un souci de correction de termes qui persistaient dans l'usage. En outre, la nomenclature internationale a ensuite fait l'objet, au niveau national, d'une adaptation dont le traducteur doit forcément tenir compte (Vandaele et Gingras Harvey, 2013).

Le troisième facteur de difficulté de la terminologie médicale concerne le phénomène de la synonymie. Idéalement, le rapport entre

102 Les exemples cités sont tirés de l'article précité de Rouleau (2003).

un terme et sa notion devrait, en langue de spécialité, être monosémique (un terme par notion). Dans la pratique, il n'en est pas toujours ainsi. Il arrive qu'une même notion soit désignée par plusieurs termes (à ce propos, voir aussi Lurquin (1986)) : « pathologie »/« maladie », « céphalée »/« mal de tête », « hémorragie »/« saignement », etc. Au problème que pose la synonymie, à savoir bien cerner la réalité désignée par tel terme, vient s'en ajouter un autre, celui de l'éponymie. Dans le domaine médical, le recours à des éponymes, patronymes ou toponymes est courant, et cela peut engendrer des difficultés du point de vue inter-linguistique : Henri Van Hoof décrit à plusieurs reprises les différences d'utilisation des éponymes qui peuvent varier selon la langue et la culture (Van Hoof, 1986, 1993, 2001).

Le phénomène de la variation terminologique se reflète donc également dans la perspective inter-linguistique. Comme l'affirme Magris (2012), l'étude de la synonymie et d'autres formes de variation a longtemps été négligée par les terminologues, en raison de l'importance accordée à une correspondance biunivoque idéale entre termes et concepts. Ces dernières années, cependant, ont produit un changement d'attitude : divers auteurs ont commencé à remettre en question ce principe et à concentrer leur attention sur le comportement réel des synonymes et des variantes dans une perspective traductologique. En outre, pour ce qui est de la langue médicale, le savoir est diffusé pour différents publics à différents niveaux de spécialisation, des articles de synthèse aux textes journalistiques. Une telle diversité constitue pour les traducteurs un défi rarement mentionné et largement sous-estimé. Par exemple, l'article de Raffo (2016) présente les résultats d'une étude sur un corpus bilingue centrée sur la manière dont les concepts spécialisés sont présentés à un public profane. Il vise à approfondir la caractérisation du discours spécialisé médical – plus spécifiquement dans les textes de vulgarisation – du point de vue de la traduction.

Compte tenu des différentes caractéristiques du lexique médical et, par conséquent, des difficultés qui peuvent survenir lors du processus de traduction spécialisée, de nombreuses ressources terminologiques ont été conçues et mises en œuvre pour soutenir le traducteur professionnel dans sa pratique. Bien qu'elles ne soient pas toutes exclusives pour le domaine médical, le panorama international propose de nombreux glossaires, vocabulaires, lexiques et bases de données

terminologiques multilingues. Le site de l'Unité de Coordination de la Terminologie (TermCoord) du Parlement européen propose une liste exhaustive consultable en ligne¹⁰³. Par ailleurs, le même TermCoord s'est récemment engagé, dans le cadre du projet « Terminologie sans frontières », dans le recueil par thèmes de ressources terminologiques monolingues ou multilingues : des ressources spécifiques au domaine médical peuvent donc être consultées sur le site en ligne¹⁰⁴. Dans ce vaste panorama des outils, nous nous concentrons exclusivement sur la description des ressources terminologiques multilingues destinées à la collection de fiches terminologiques qui regroupent les informations linguistiques accompagnant le terme dans la langue source et la langue cible. Dans ce contexte, il faut mentionner la ressource précédemment décrite *VariMed*¹⁰⁵ qui s'adresse non seulement aux patients, mais également aux traducteurs scientifiques. Comme l'expliquent Alarcón-Navío et al. (2016), le projet est né afin de gérer et d'organiser le phénomène concernant la variation terminologique dans une perspective traductologique. Les traducteurs sont souvent contraints de faire un choix parmi une multitude de variantes terminologiques ayant une signification similaire. La ressource est donc axée sur les motivations cognitives et communicatives de la variation dénomminative, et la manière dont elles se reflètent dans les variantes terminologiques d'un concept. La ressource est disponible pour l'anglais et l'espagnol. Pour chaque entrée terminologique associée à un concept, les variantes dénomminatives du terme sont illustrées. Par exemple, le traducteur qui consulte le couple de termes désignant le concept *fiebre-fever*¹⁰⁶ aura accès à une fiche terminologique qui fournit des informations relatives à : la catégorie grammaticale, le registre, l'usage géographique, les variantes de mots, les contextes d'utilisation du terme et les images liées au concept même.

Une autre ressource multilingue développée par le groupe de recherche LexiCon et spécifique pour le domaine de l'oncologie est la base de données terminologiques OncoTerm¹⁰⁷. La ressource, conçue

103 <http://bit.ly/2W5KggH>.

104 <https://yourterm.org/med-resources-centre/>.

105 <http://varimed.ugr.es>.

106 <http://varimed.ugr.es/index.php?op=viewconcepto&idconcepto=32>.

107 <https://www.ugr.es/~oncoterm/>.

pour le besoin de gérer et de cataloguer la terminologie oncologique à des fins différentes (Faber, 2002), offre un modèle de fiche terminologique de type bilingue (espagnol et anglais). En cherchant le terme *carcinoma* en anglais, par exemple, le traducteur peut visualiser simultanément une fiche bilingue avec les mêmes informations pour le terme de départ et le traduisant : la partie du discours, le numéro grammatical, la définition, le contexte, la date de création et la personne chargée de la compilation de la fiche terminologique.

En général, nous pouvons affirmer que la tendance retrouvée dans ce type de ressource est d'aller au-delà du simple glossaire ou dictionnaire, offrant des informations pertinentes au processus de traduction (comme la variation ou le contexte d'utilisation) qui ne se limitent pas à une liste des termes et de ses traduisants. Cependant, compte tenu des nombreux phénomènes terminologiques qui caractérisent cette langue de spécialité, nous croyons fermement que le traducteur technico-scientifique, n'étant pas forcément spécialisé dans le domaine médical, doit disposer d'une ressource qui fournit un plus grand nombre d'informations terminologiques afin de pouvoir consulter une image globale et exhaustive de la terminologie analysée.

5.2 Analyse des exigences

À la lumière du panorama présenté jusqu'à présent sur la complexité de la terminologie médicale, la définition de ses caractéristiques et les ressources disponibles et consultables par le grand public, nous décrirons ci-dessous le projet à la base de notre recherche. L'état de l'art présenté dans le domaine de la terminologie médicale et les exigences de normalisation de la structure des ressources terminologiques numérique nous ont menés à réfléchir sur la nécessité de concevoir une nouvelle ressource terminologique normalisée pour le domaine médical. Cet objectif est motivé par deux exigences en particulier :

1. fournir une ressource numérique qui se veut suffisamment exhaustive, c'est-à-dire contenant un grand nombre de

- catégories et de données terminologiques, afin de répondre aux besoins d'information de plusieurs utilisateurs ;
2. répondre aux exigences formulées dans la modélisation conceptuelle de la « terminologie FAIR », en concevant une ressource terminologique multilingue qui fournisse des données numériques trouvables, accessibles, interopérables et réutilisables.

Le fruit de nos recherches s'est matérialisé dans la conception et l'implémentation de la ressource multilingue TriMED¹⁰⁸. Le nom de la ressource s'inspire de son caractère tripartite. Trois sont les langues de travail pour lesquelles les données terminologiques sont fournies, c'est-à-dire le français, l'anglais et l'italien, et trois sont les typologies d'utilisateurs auxquels la ressource s'adresse :

1. les patients ;
2. les traducteurs ;
3. les médecins.

Les raisons qui nous ont permis d'identifier ces utilisateurs potentiels découlent de l'analyse de leurs besoins d'information. Quant aux patients, ou plus généralement au grand public non initié au domaine médical, il existe un réel problème de compréhension de la terminologie médicale dû au niveau de spécialisation élevé de la langue. L'exigence du patient est donc de disposer d'une ressource qui mette en évidence le phénomène de variation diastratique. En ce sens, la ressource TriMED vise à fournir, pour chaque terme technique, son équivalent en langue populaire et sa définition.

Dans une perspective inter-linguistique, les traducteurs sont confrontés à de nombreux défis liés au décodage et au transcodage des informations médicales à traduire. Comme nous l'avons vu, bien que le marché offre un grand nombre d'outils pour des domaines de travail très variés, rares sont les ressources linguistiques spécifiquement conçues pour le domaine médical. De plus, la majorité de ces outils ne fournissent pas un aperçu complet des données linguistiques

108 Le lien pour la consultation de la ressource est le suivant : <https://purl.org/trimed>

liées aux termes techniques. Or, nous avons vu que le processus de traduction médicale doit prendre en compte différents axes d'analyse. Pour cette raison, la ressource terminologique TriMED présentera un nouveau modèle de fiche terminologique contenant jusqu'à 42 catégories de données afin de fournir un cadre global sur le comportement syntaxique, sémantique et phraséologique du terme source et de son traduisant en langue cible. En outre, la structuration de TriMED en tant que ressource numérique FAIR permettra aux traducteurs d'exporter les données fournies et de les importer dans les systèmes de traduction assistée par ordinateur.

Enfin, soulignant la composante d'interopérabilité entre les ressources, nous nous tournons vers les professionnels de santé en offrant un point d'accès unique à la consultation des autres terminologies, nomenclatures ou codes de classification internationaux utilisés habituellement par des experts. TriMED permettra donc un accès direct aux ressources médicales, telles que les codes ICPC2 et ICD10, et les vocabulaires SNOMED CT et MeSh Terms.

Le caractère innovant de la base de données se reflète donc dans 1) la richesse des données fournies, 2) la prise en compte systématique de différentes catégories d'utilisateurs et 3) la normalisation de sa structure. Dans les sections suivantes, nous décrivons donc la modélisation conceptuelle de TriMED comme la première ressource terminologique à adopter les principes « FAIR » susmentionnés à travers la formulation du paradigme. La structure de la ressource est inspirée du paradigme de la « terminologie FAIR » proposé et décrit dans les chapitres précédents. Compte tenu de la publication récente des trois normes ISO sur lesquelles repose la modélisation conceptuelle, la ressource TriMED est configurée – à notre connaissance – comme la première à les implémenter dans leur version actuelle.

5.3 Modèle structurel de TriMED

La conception de la structure de la ressource TriMED se fonde sur la norme ISO 16642 concernant le métamodèle *Terminological Markup*

Framework (TMF) (ISO-16642, 2017). Comme nous l'avons vu dans la section 3.3.2 du chapitre 3, le TMF est une norme internationale qui fournit un cadre pour la représentation de collections de données terminologiques, telles que les bases de données terminologiques multilingues, dans le langage de balisage *eXtensible Markup Framework* (XML). La norme TMF s'articule autour de deux niveaux d'abstraction. Le premier niveau concerne une description du métamodèle qui sous-tend l'analyse, la conception et les échanges de données à un niveau très général. Le métamodèle est donc indépendant de toute implémentation ou de tout logiciel spécifique. Le deuxième niveau concerne plutôt les catégories de données qui peuvent être associées aux niveaux du métamodèle et qui sont spécifiques à chaque collection des données terminologiques (voir la section 5.4).

Le modèle structurel de la ressource TriMED est donc basé sur le cadre TMF qui adopte l'approche selon laquelle : un concept est décrit dans un nombre générique n de langues et est désigné par un nombre générique n de termes pour chaque langue. En particulier, le modèle structurel que nous proposons est configuré comme suit :

- TriMED est une collection de données terminologiques (CDT) qui contient un nombre quelconque d'entrées terminologiques (ET).
- Chaque ET renvoie à un concept unique qui peut donc être représenté dans n langues dans les sections de langue (SL).
- Pour chaque langue, il existe n sections de terme (ST) contenant les termes qui, dans cette langue particulière, décrivent le concept.
- Chaque ST peut contenir un nombre quelconque de sections de composants de terme (SCT) fournissant des informations sur des parties d'un terme telles que des morphèmes, des phonèmes, des syllabes ou des mots isolés à partir d'un terme composé.

Les relations entre les différentes instances (entrée terminologique, section de langue, section de terme et section de composants de terme) sont réglées par les cardinalités, c'est-à-dire les valeurs qui indiquent le nombre minimum et maximum (x, y) des éléments des instances qui sont en relation entre eux. Dans l'image 5.1, nous proposons le schéma entité-association (Chen, 1976) du TMF pour la représentation de TriMED, où les entités (rectangles) sont les instances et les associations (losanges) sont les relations qui unissent ces derniers : la collection de

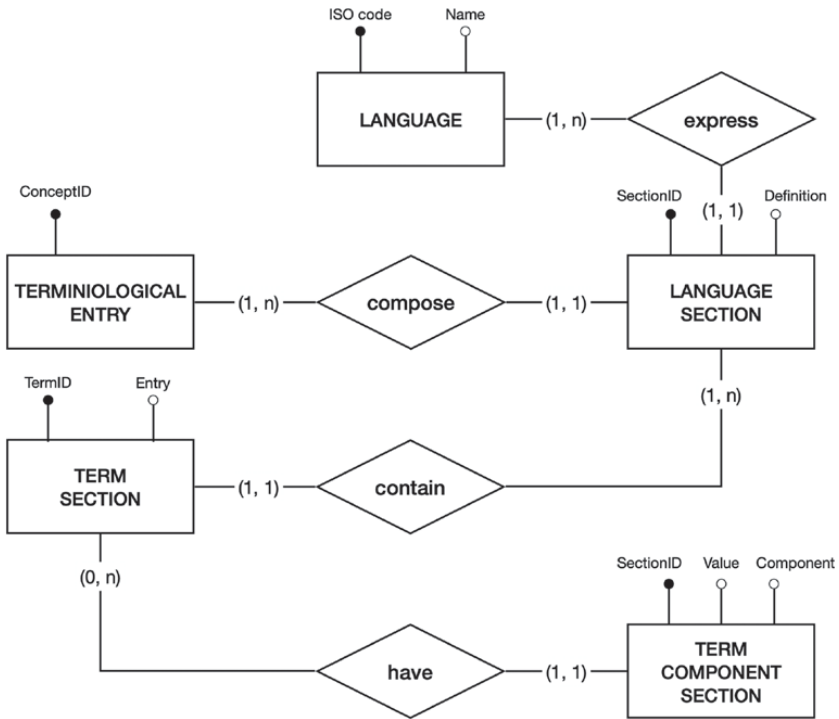


Image 5.1 – Schéma entité-association de TriMED

données terminologiques contient un nombre quelconque d'entrées terminologiques (0, n). Une ET doit contenir au moins une SL (1, n). Une SL, qui fait référence à une et une seule langue, doit contenir au moins une ST (1, n) et une ST peut contenir un nombre quelconque de SCT (0, n). L'organisation hiérarchique est assurée par les cardinalités (1,1) exprimées entre les nœuds du métamodèle TMF.

Cette structure distingue les niveaux hiérarchiques (Romary, 2001a) auxquels différentes catégories de données (propriétés) peuvent être associées :

1. les données de type conceptuel communes à toutes les langues ;
2. les données spécifiques à une langue ;
3. les données spécifiques à un terme.

Dans la ressource TriMED un concept unique, exclusif pour une entrée terminologique et défini par son identifiant, peut être exprimé par n langues. Les sections de langue ont leur propre identifiant et incluent la définition exprimant le concept associé, ainsi que le code de langue défini par la norme ISO 639¹⁰⁹. Par la suite, pour chaque langue, tous les termes désignant ce concept-là sont associés avec toutes leurs données de type terminologique. Enfin, dans le cas d'un terme complexe, les caractéristiques des parties du terme sont regroupées dans la SCT.

5.4 Fiche terminologique

La ressource TriMED est donc une collection de données terminologiques qui recueille un ensemble d'entrées terminologiques. L'une des nouveautés de cette ressource réside dans le modèle de fiche terminologique fournie afin de collecter un grand nombre de données terminologiques pour chaque entrée.

Une fiche terminologique est un outil pour enregistrer, dans un ensemble structuré de champs, les données terminologiques d'un concept spécialisé (Pavel et al., 2001). Autrement dit, la fiche terminologique se présente comme « un ensemble structuré de données terminologiques relatives à une notion »¹¹⁰. Une ressource terminologique est généralement organisée autour d'une collection de fiches terminologiques rassemblant des informations pertinentes sur les termes techniques. Le choix des données relatives à un terme à afficher à l'utilisateur dépend de la finalité de la ressource. Comme l'affirme (Gouadec, 1990) :

une fiche de terminologie ou un dossier de terme ne s'improvisent pas : toute décision concernant la liste et le contenu de chacune des rubriques de la fiche ou du dossier est lourde de conséquences.

109 <https://www.iso.org/iso-639-language-codes.html>.

110 <https://www.iso.org/standard/5591.html>.

En effet, l'omission d'une information nécessaire à l'analyse oblige à reprendre tout le travail et à modifier la fiche, occasionnant une perte de temps considérable.

Dans la ressource TriMED, la structure de la fiche vise la complétude, afin de respecter les exigences des diverses catégories d'utilisateurs (patients, traducteurs et médecins) : elle permet un repérage optimal des informations et assure une bonne gestion et une organisation optimale des données. Le choix des informations à montrer à l'utilisateur a été guidé par le désir de fournir une ressource qui souhaite :

1. aider les patients à comprendre correctement les informations médicales, compte tenu de l'aspect de la variation diastratique de la terminologie ;
2. soutenir le traducteur dans le processus de traduction spécialisée en fournissant un cadre sur le comportement syntaxique, sémantique et phraséologique du terme source et de son traduisant en langue cible ;
3. fournir un point d'accès unique pour la consultation des professionnels de la santé aux autres terminologies, nomenclatures ou codes de classification internationaux généralement utilisés par des experts.

En particulier, la fiche terminologique de TriMED fournit jusqu'à 42 champs terminologiques (c'est-à-dire « catégories de données » en utilisant la terminologie de la norme ISO-12620 (2019)) qui peuvent être associés à différents niveaux du modèle structurel TMF. Pour cette raison, la fiche de TriMED contient des catégories de données associées 1) au concept, 2) à la langue de travail et 3) au terme analysé. Dans la sous-section 5.6.2, nous discuterons les contraintes de position pour les catégories de données et les raisons qui nous ont menés à ces choix. Pour le moment, nous décrivons dans la sous-section suivante les catégories de données fournies dans la fiche, en les classant en catégories de type terminologique, conceptuel et administratif pour des raisons de logique discursive. Nous avons décidé de présenter, d'abord, les catégories de type terminologique, c'est-à-dire les propriétés linguistiques qui peuvent être associées au terme qui désigne un concept donné pour une langue donnée. Vient ensuite la description des catégories de type conceptuel qui devraient être communes à toutes les langues. Enfin, nous présentons les catégories de type administratif, qui permettent d'assurer la traçabilité de la fiche terminologique.

5.4.1 Catégories de type terminologique

La fiche est articulée autour de différents axes d'analyse du terme. Le tableau 5.1 montre la liste des catégories de données de type terminologique qui sont fournies dans le modèle de fiche. Tout d'abord, la fiche contient la description des caractéristiques formelles du terme en rassemblant les informations concernant la partie du discours, le genre et le nombre grammatical. Chaque terme est donc étiqueté comme « nom », « verbe », « adjectif », ou « adverbe ». Le genre grammatical est fourni comme caractéristique intrinsèque du nom qui est donc classé comme « masculin », « féminin » ou « neutre ». Enfin, le nombre grammatical indiquant la quantité d'unités du terme est prévu : « singulier », « pluriel », « collectif » ou « partitif ». La fiche s'étend également à la famille lexicale du terme : elle contient tous les noms, verbes, adjectifs et adverbes qui dérivent du terme analysé et qui peuvent être réunis sous la même sphère sémantique. Par exemple, la fiche du terme « virologie » comprendra un champ relatif aux formes dérivées indiquant comme données terminologiques les termes « virologique », « virologue » et « virologue »¹¹¹. La fiche fournit aussi la prononciation en alphabet phonétique international (API) et d'autres informations de type étymologique, comme la dérivation et la composition du terme (dans le cas de l'étymologie de « virologie » : « comp. des élém. formants viro-, tiré de virus* et -logie*. Cf. angl. *Virology* ».)

Tableau 5.1 – Catégories de type terminologique

Axe d'analyse	Catégorie de données
Morphologie	partie du discours, genre grammatical, nombre
Phonétique	transcription API
Étymologie	dérivation, composition
Variation	variante orthographique, abréviation, expansion, acronyme
Sémantique	définition, analyse sémique, synonyme, hyponyme, hyperonyme
Phraséologie	unité phraséologique, collocation
Pragmatique	contexte d'utilisation
Registre	nom populaire, nom scientifique, codes de classification internationaux
Domaine	domaine, sous-domaine

111 <https://www.cnrtl.fr/definition/virologie>.

Une attention particulière est consacrée au phénomène de la variation des termes. Dans la fiche, apparaissent également les éventuelles variantes orthographiques, les abréviations et les acronymes fréquemment employés dans la langue médicale (Monin, 1993 ; Bowker et Hawkins, 2006). Les variantes terminologiques renseignent également les éventuelles variations orthographiques du terme qui peut s'écrire de plus d'une façon (Dugas, 1994 ; Jaffré, 2010) : « xénoœstrogène » ou « xéncœstrogène », « acupuncture » ou « acuponcture », etc. Les abréviations concernent plutôt les formes abrégées résultant de l'omission de lettres d'un terme (« acdts » pour « accidents »), alors que l'acronyme est formé des initiales de plusieurs termes et est prononcé syllabiquement comme un mot : par exemple « SLA » pour « Sclérose Latérale Amyotrophique ». Inversement, si l'entrée du terme est un acronyme, la fiche renseigne sur son expansion.

La section consacrée à la sémantique du terme propose, tout d'abord, une définition des termes à travers l'analyse du sens qui leur est traditionnellement attribué par une communauté de personnes partageant le même savoir et ayant un objectif commun (L'Homme et San Martín, 2016). Les définitions constituent un système de connaissance structuré afin de contenir la signification d'un terme ; elles doivent « donner d'une notion une image mentale exacte » (Dubuc, 2002) et permettre de la « différencier des autres notions à l'intérieur d'un système notionnel » (Boutin-Quesnel, 1985). En outre, nous nous concentrons sur le point de vue sémantique en fournissant l'analyse sémique du terme (Rastier, 2009). L'analyse sémique est une méthode d'étude utilisée dans la sémantique compositionnelle visant à décomposer la signification des termes en unités minimales de sens qui ne peuvent pas être segmentées : les sèmes. L'union de multiples traits sémantiques (sèmes) constitue la signification d'un lexème (Hébert, 2007). Par exemple :

- « pathologie » : /étude/ /médecine/ /développement/ /cause /symptôme/ /maladie/ ;
- « maladie » : /altération/ /santé/ /être vivant/ /signe/ /symptôme/ /causes internes/ /causes externes/ /évolution/ /organisme/ ;

- « bistouri » : /instrument/ /chirurgie/ /forme de couteau/ /lame fixe/ /lame repliable/ /incision/.

De plus, la fiche fournit aussi les variantes synonymiques du terme, c'est-à-dire tout terme synonyme (ou quasi-synonyme) désignant le même concept. Or, la question de la synonymie parfaite est un sujet très discuté dans la littérature, en particulier pour les langues de spécialité, car les soi-disant synonymes couvrent rarement tous les aspects sémantiques du même concept (pour ne citer que quelques auteurs sur ce sujet : Lurquin (1986) ; Lethuillier (1989) ; Durieux (1996) ; Quérin (2013)). Dans la fiche terminologique de TriMED, la catégorie de données /synonyme/ est fournie en conformité avec la norme ISO-12620 (2019) et nous avons ajouté un espace de réflexion pour les considérations liées à l'équivalence sémantique des termes à travers la catégorie de données /note/. D'un point de vue traductologique, cet espace peut également être utilisé pour mener des réflexions sur les équivalences ou sur les différences de dénominations entre les langues. Par exemple, en français le synonyme de « Sclérose Latérale Amyotrophique » est le terme complexe sous forme d'éponyme « Maladie de Charcot ». En italien, il y a équivalence du terme scientifique *Sclerosi Laterale Amiotrofica*, tandis que pour le synonyme, nous parlerons plutôt de *Malattia di Lou Gehrig*, du nom du joueur de baseball, dont la maladie en 1939 a attiré l'attention du public. Enfin, la fiche prévoit une catégorisation des termes selon leurs relations sémantiques. En particulier, deux champs sont consacrés à la collecte d'hyperonymes et d'hyponymes du terme analysé. L'hyperonyme est le terme qui désigne le générique ; l'hyponyme désigne le spécifique (L'Homme, 2004) : dans le couple « imagerie médicale » – « scintigraphie », par exemple, le premier est l'hyperonyme ; le second, l'hyponyme. L'« imagerie médicale » regroupe les moyens d'acquisition et de restitution d'images du corps humain et la « scintigraphie » est l'une parmi les méthodes possibles.

Pour l'évaluation du comportement sémantique du terme, nous prenons également en analyse sa phraséologie en considérant les unités phraséologique du terme, en particulier les locutions et les collocations (Galinski, 1990 ; Rousseau, 1993). Reprenant la définition fournie par Polguère (2015), nous considérons la locution comme une unité lexicale – une lexie – à part entière :

une locution est une expression phraséologique sémantiquement non compositionnelle. En tant que telle, elle doit être considérée comme non construite : elle ne se modélise pas comme assemblage syntaxique de ses éléments lexicaux dans l'interface sémantique-syntaxe.

Pour sa part, la collocation est plutôt considérée comme une structure binaire, formée d'une base et d'un collocatif, constituant une association lexicale privilégiée dont le sens est compositionnel. Dans la fiche terminologique du terme « cœur », par exemple, les cas de locution « chirurgie à cœur ouvert » ou « souffle au cœur »¹¹² et les cas de collocation « cœur anémique, bon, défaillant, excellent, fatigué, fragile, instable, malade »¹¹³ sont illustrés.

Dans la fiche, toute une section consacrée au corpus fournit les contextes spécialisés dont les termes ont été extraits. Le contexte « renseigne sur le contenu notionnel du terme sans avoir la rigueur formelle d'une définition » (Boutin-Quesnel, 1985) et permet de dégager une image précise du concept. Le contexte est donc un élément fondamental dans notre modèle de fiche terminologique, car il permet aussi de distinguer les cas d'homonymie et de polysémie du terme spécialisé. Par exemple, l'adjectif « cervical » désigne 1) la partie postérieure du cou, et 2) la partie anatomique relative au col de l'utérus¹¹⁴. Dans ce sens, les contextes permettent de dégager la signification du terme :

1. « La dystonie *cervicale* se caractérise par des contractions involontaires prolongées (chroniques) ou par des contractions intermittentes périodiques des muscles du cou, induisant le cou à pivoter de différentes manières »¹¹⁵ ;
2. « L'insuffisance *cervicale* est l'ouverture indolore du col de l'utérus, entraînant la naissance du bébé au cours du 2e trimestre de la grossesse »¹¹⁶.

112 <https://www.cnrtl.fr/definition/coeur>.

113 <https://bit.ly/3BP62L>.

114 Du latin *cervical*, de *cervix* (« nuque, cou, tête, goulot ») : <https://www.cnrtl.fr/definition/cervical>.

115 <https://msdmnls.co/3pzyfkH>.

116 <https://msdmnls.co/36wGW8t>.

Dans la base de données TriMED, ce phénomène est géré en insérant deux ou plusieurs entrées pour le même terme, car il désigne deux ou plusieurs concepts différents.

Nous procédons ensuite à l'identification du domaine et du registre d'emploi du terme. La notion de domaine joue en effet un rôle central dans la terminologie en tant que constituant du concept :

l'ensemble concept-définition est incomplet et donne l'impression d'être un peu bancal. [. . .] Le concept, sa définition (et son terme) appartiennent obligatoirement à un domaine. [. . .] Le domaine permet d'indiquer le système conceptuel auquel appartient le concept (De Bessé, 2000).

Le terme et sa définition prennent donc sens au moment où ils sont reliés à un domaine spécifique. Dans la fiche, nous identifions donc le domaine (« médecine ») et les sous-domaines d'usage du terme, par exemple « chirurgie », « pathologie », « pharmacologie », etc.

La fiche offre également la spécification du niveau de registre de langue dans lequel le terme est employé : « populaire », « argotique », « vulgaire », « familier », « courant » ou « standard » et « soutenu ».

La fiche terminologique accorde une attention particulière au phénomène de variation de registre, ou variation diastratique de la langue (Coseriu, 1969). Sur la base de la proposition de typologie de Freixa (2006), ce phénomène est attribuable aux causes de type fonctionnel lorsque le niveau de spécialisation linguistique des interlocuteurs doit être adapté (voir par exemple Picton et Dury (2017)). Dans ce sens, la fiche prévoit deux catégories de données /nom populaire/ et /nom scientifique/, afin d'illustrer les différences de registre et de promouvoir la compréhension de l'information médicale de la part des non-experts. Des exemples de ce type sont : « rougeole » (nom scientifique) et « première maladie » (nom populaire), ou « coqueluche » (nom scientifique) et « toux des cent jours » (nom populaire). Enfin, dans le but d'interopérabilité entre les ressources dans le domaine médical, chaque terme de la ressource TriMED est mappé avec un lien direct vers d'autres nomenclatures et codes de classification internationaux. La fiche fournit donc les catégories de données 1) /Code ICPC2/, 2) /Code ICD10/, 3) /Terme SNOMED CT/ et 4) /Terme Mesh/ contenant l'URL du terme correspondant dans les suivantes ressources médicales : 1)

Tableau 5.2 – Catégories de type conceptuel

Axe d'analyse	Catégorie de données
Concept	concept superordonné, concept subordonné, relation entre concepts, sphère conceptuelle

*International Classification of Primary Care (ICPC-2)*¹¹⁷, 2) *International Classification of Diseases (ICD10)*¹¹⁸, 3) *Systematized Nomenclature of Human and Veterinary Medicine—Clinical Terms (SNOMED CT)*¹¹⁹ et, 4) *Medical Subject Headings (MeSh)*¹²⁰.

5.4.2 Catégories de type conceptuel

La fiche terminologique de TriMED a une orientation conceptuelle afin de répondre au principe de conformité à la norme ISO-16642 décrite ci-dessus : chaque entrée terminologique fait référence à un concept unique auquel une définition est associée et qui regroupe tous les termes qui le désignent. Dans le tableau 5.2, les catégories de type conceptuel fournies dans notre modèle de fiche sont illustrées. Tout d'abord, il faut préciser que les données fournies au niveau conceptuel sont exclusivement de type alphanumérique et servent à établir des relations entre concepts – et, par conséquent, entre termes – au sein de notre système conceptuel. Par exemple, le terme « thrombose » a pour identifiant unique le code alphanumérique « TPLT19015 » et est directement lié au concept ayant le code d'identification « C5 ». Les identifiants du terme et du concept sont exclusifs pour chaque fiche terminologique et permettent d'identifier et de désambigüiser la relation directe entre

117 <https://www.who.int/classifications/icd/adaptations/icpc2/>.

118 <https://www.who.int/classifications/icd/en/>.

119 <http://www.snomed.org>.

120 <https://meshb.nlm.nih.gov/search>.

les deux, en particulier dans la phase d'implémentation en langage de programmation.

Nous avons vu qu'au niveau linguistique, la fiche prend en considération les relations hiérarchiques entre les termes en fournissant les informations sur l'/hyperonyme/ et l'/hyponyme/ du terme analysé. De même, au niveau conceptuel, nous entendons maintenir cette représentation hiérarchique en fournissant les catégories du /concept supérieur/ (concept de niveau supérieur) et du /concept subordonné/ (concept de niveau inférieur). Comme l'affirme Depecker (2017), les relations de superordination et de subordination dérivent de la propriété de certains concepts de subsumer d'autres concepts, c'est-à-dire d'en englober d'autres sous eux (principe de subsumption). Ces catégories de données contiendront donc les identifiants alphanumériques de l'entrée terminologique au niveau conceptuel et seront directement liées aux identifiants des termes désignant ces concepts et exprimés dans la relation hyperonymie–hyponymie. Les relations entre les concepts ne se limitent pas seulement à la hiérarchisation entre la superordination et la subordination. Il existe en fait d'autres types de relations : coordination, association, opposition, etc. (Depecker, 2017). La fiche terminologique de TriMED permet d'explicitier ces liens sémantiques à travers la catégorie de données /relation entre concepts/.

La dernière catégorie de type conceptuel fournie par la fiche terminologique de TriMED concerne la /sphère conceptuelle/ du terme. Avec cette expression, nous nous référons à une macroclassification qui regroupe les concepts et, par conséquent, les termes sémantiquement liés par des critères spécifiquement identifiés pour le domaine médical. La définition des sphères conceptuelles a été réalisée avec l'aide d'un expert. En particulier, nous avons identifié sept sphères conceptuelles classées comme suit :

- A – Diagnostic : types, méthodes et outils (en particulier, tests et examens) pour les investigations, les évaluations et les formulations diagnostiques et pronostiques des maladies.
- B – Epidémiologie et prévention : systèmes de prévention, méthodes d'apparition, propagation et fréquence des maladies en fonction des conditions de l'organisme, de l'environnement et de la population.

- C – Pathologie : maladies, conditions pathologiques, troubles et leurs signes (manifestations) et symptômes.
- D – Spécialités médicales.
- E – Pharmacologie : médicaments, produits chimiques, préparations artificielles.
- G – Anatomie : organisme humain, cellules, tissus, organes et systèmes.
- H – Thérapie : outils, techniques et opérations chirurgicales, préventives (ou prophylactiques), thérapies de soutien, psychologiques ou psychothérapeutiques pour le traitement des maladies, des conditions pathologiques et des troubles.

Chaque terme contenu dans la ressource TriMED est donc associé manuellement à une /sphère conceptuelle/ selon le champ sémantique auquel il appartient. Par exemple, les termes « scintigraphie », « angiotomodensitométrie » et « échographie » seront regroupés sous la sphère conceptuelle ayant comme identifiant la lettre A en tant qu'outils d'investigation diagnostique et pronostique de la maladie. Les termes « autoévaluation », « dépistage » et « contagion » seront associés à la sphère conceptuelle B en tant que désignant des concepts liés à la prévention et à la propagation des maladies. La sphère conceptuelle C comprendra plutôt les noms de maladies telles que le « syndrome de Down », la « maladie d'Alzheimer » et la « diphtérie ». Les différentes spécialités de la médecine telles que la « virologie », la « bactériologie » ou la « biologie » seront plutôt regroupées dans la sphère conceptuelle plus générique D. Les termes relatifs aux substances pharmaceutiques ou préparations chimiques, telles que les « alcaloïdes », le « chloroforme » ou la « solution saline » sont regroupés sous la sphère conceptuelle E. Enfin, les termes concernant le corps humain (« rate », « colonne vertébrale » ou « foie ») seront étiquetés avec la sphère conceptuelle G, tandis que les termes désignant les thérapies et techniques pour le traitement chirurgical de la maladie, tels que « résection », « ablation », « excision », feront partie de la sphère conceptuelle identifiée par la lettre H. Le choix de cette catégorie de données permet donc de regrouper les termes sémantiquement liés. Cette catégorie peut être donc classée comme une entité plus générique du même concept. De cette

manière, des termes tels que 1) « diagnostic » (concept A1), 2) « diagnostiquer » (concept A2) et 3) « diagnostique » (concept A3) peuvent être considérés comme trois concepts différents (chacun correspondant à sa propre définition exprimée en langage naturel) indiquant 1) une pratique, 2) une action et 3) une qualification, regroupés sous la même sphère conceptuelle A.

5.4.3 Catégories de type administratif

Afin d'assurer la traçabilité des données fournies et leur interprétation correcte, la fiche terminologique de TriMED fournit également des catégories de données de type administratif illustrées dans le tableau 5.3. Pour que les données terminologiques soient trouvables, accessibles, interopérables et réutilisables, il est strictement nécessaire qu'elles soient associées à des identifiants uniques et persistants. Les identifiants sont associés à chaque instance du modèle structurel TMF : cela signifie que chaque entrée terminologique fournira un code exclusif pour le niveau du concept, de la langue et du terme.

Tableau 5.3 – Catégories de type administratif

Axe d'analyse	Catégorie de données
Admin	Identifiants, source, référence croisée, référence croisée externe, responsabilité, type de transaction, date, note

L'image 5.2 montre que le concept ayant le code A13 (où la lettre « A » se réfère à la sphère conceptuelle du « Diagnostic » précédemment décrite) est représenté en trois langues qui ont comme identifiant : « FR » pour le français, « IT » pour l'italien et « EN » pour l'anglais. Ensuite, un terme est associé à chaque langue, c'est-à-dire la désignation de ce concept dans cette langue, chacune ayant son

propre identifiant exclusif : *scintigraphie* (TPLT19035), *scintigrafia* (TPLT18056) et *scintigraphy* (TPLT18055). Le fait d'avoir trois termes traduisants avec des identifiants exclusifs qui pointent vers le même identifiant de concept nous permet de gérer efficacement les données et de récupérer automatiquement les traduisants du concept lui-même.

Un autre aspect important pris en compte dans le modèle de fiche est la traçabilité des sources consultées notamment pour la /définition/ et le /contexte/ du terme. En ce sens, la fiche fournit les catégories de données /source/ et /référence croisée externe/ afin d'explicitier 1) la source des informations dans le cas des documents papier et 2) l'URL de la ressource externe consultée dans le cas de documentation numérique. Par ailleurs, en cas de références internes à la ressource, la catégorie de données /référence croisée/ permet d'orienter l'utilisateur vers des liens et d'associations entre les fiches terminologiques. Les autres catégories administratives concernent la /responsabilité/, c'est-à-dire l'individu qui a formulé ou modifié l'entrée terminologique, le /type de transaction/ réalisé (généralement, origine ou modification), la /date/ dans laquelle la transaction a eu lieu et toute /note/ complémentaire.

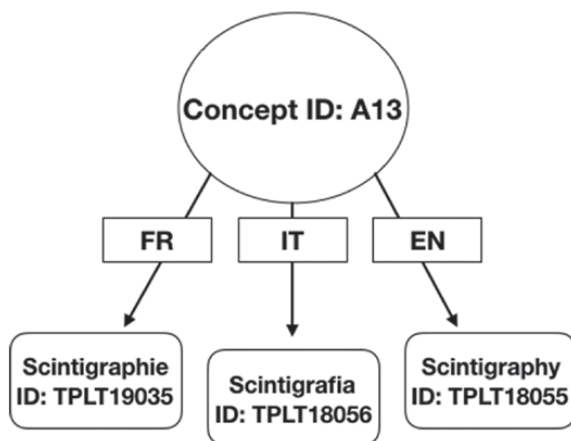


Image 5.2 – Identifiants dans TriMED

5.5 Répertoire des catégories de TriMED

Dans la section 3.3.3 du chapitre 3, nous avons décrit l'importance de documenter les catégories de données présentes dans une ressource terminologique afin de comprendre leur signification, de pouvoir les réutiliser efficacement et de garantir leur interopérabilité. En particulier, le document de référence pour leur normalisation est la norme ISO-12620 (2019) concernant les *Spécifications des catégories de données*. Une spécification de catégorie de données fournit la représentation complète et formelle d'une catégorie. Pour leur description et leur implémentation, la norme fait référence à un répertoire de catégories de données nommé DatCatInfo¹²¹, qui recueille une liste de toutes les catégories de données et leurs spécifications disponibles pour la terminologie et produites au sein de l'Organisation Internationale de Normalisation (ISO).

La nouveauté de cette norme réside dans la possibilité que tous les implémenteurs conçoivent leur propre répertoire spécifique pour leur ressource terminologique afin de documenter les catégories de données qui ne sont pas présentes sur DatCatInfo, ou sur tout autre répertoire partagé. De cette façon, la définition d'un cadre clair pour spécifier, gérer et utiliser les catégories de données améliore l'interopérabilité entre les ressources terminologiques. À cette fin, en suivant les directives prévues dans la norme précitée, nous avons conçu le répertoire de catégories de données de TriMED qui rassemble toutes les spécifications, c'est-à-dire la description de toutes les catégories de données terminologiques que nous fournissons dans notre modèle de fiche terminologique. L'application Web a été implémentée à l'aide du *package Shiny R* (Chang et al., 2021) et est disponible en ligne¹²². De plus, pour permettre la visualisation au format papier, nous fournissons dans l'annexe A de cet ouvrage la description de toutes les spécifications de données fournies sous forme de tableau. À notre connaissance, il s'agit du

121 <http://www.datcatinfo.net/>.

122 Le lien pour la consultation du répertoire est le suivant : <http://purl.org/trimed/dcr>.

premier répertoire implémenté selon cette norme, et il recueille 42 spécifications de catégorie de données. Comme le montre l'image 5.3¹²³, les informations fournies pour chaque catégorie de données sont :

Data Category Repository - TriMED

The screenshot shows the Data Category Repository interface. On the left, there are two dropdown menus: 'Language' (set to 'en') and 'Data Category' (set to 'Common name'). On the right, the search results are displayed. The first result is for a category with the following details:

- Description:** XML
- PID:** <http://www.datcatinfo.net/datcat/DC-328>
- Identifier:** commonName
- Module:** Trimed
- Level (TMF):** termSec
- Classification:** <termNote>
- Implemented as:** string
- Value(s):** NA

Below these details, there is a description: 'Description: A synonym for an international scientific term that is used in general discourse in a given language.' and other fields like 'Explanation: NA', 'Note: NA', and 'Example: Mountain laurel, as opposed to the international scientific name, Kalmia latifolia'. At the bottom, there is a 'Download CSV' button.

Image 5.3 – Répertoire des catégories de TriMED

1. Un identifiant unique et persistant (PID), c'est-à-dire une URL qui fournit l'accès Web direct à la spécification de la catégorie de données dans le répertoire en ligne.
2. Un identifiant mnémotechnique unique et stable de la catégorie de données qui ne doit pas inclure d'espaces entre les mots, car il est utilisé dans les environnements de codage comme élément ou comme valeur d'attribut. Dans l'exemple fourni, cet identifiant est *common Name*.
3. Le module de catégories de données TBX auquel la catégorie se réfère (voir la section 5.6).
4. Le niveau du métamodèle TMF (concept, langue et terme) auquel la catégorie de données est associée dans notre ressource. L'exemple illustre la catégorie /common name/ associée à la section de terme.

¹²³ La capture d'écran a été prise le 21/10/2021 à partir de la page suivante : <http://purl.org/trimed/dcr>.

5. La typologie de contenu de la catégorie de données, c'est-à-dire les types d'informations que la catégorie de données permet par sa mise en œuvre, comme une liste déroulante ou une chaîne de caractères.
6. L'ensemble de valeurs énumérées que la catégorie de données peut avoir si elle est implémentée en tant que liste déroulante.
7. La définition de la catégorie de données.
8. D'autres explications et notes sur la catégorie de données.
9. Quelques exemples d'utilisation de la catégorie de données, tels que *singulier* ou *pluriel*.
10. La traduction du nom canonique de la catégorie de données dans les autres langues de travail de la ressource. Dans l'exemple fourni, *nom populaire* pour le français et *nome popolare* pour l'italien.

Il faut noter que, afin d'éviter la duplication des métadonnées et d'assurer une certaine traçabilité, les informations fournies dans le répertoire de TriMED concernant les PID des catégories de données se réfèrent à l'URL correspondant sur DatCatInfo pour les catégories qui sont déjà documentées dans ce répertoire. Par exemple, le PID de la catégorie de données /common name/ est le même que celui fourni dans DatCatInfo, à savoir <http://www.datcatinfo.net/datcat/DC-328> Il existe également un nombre restreint de catégories de données qui sont exclusives pour notre ressource et ne sont donc pas illustrées sur DatCatInfo. Ces catégories de données terminologiques comprennent :

- 1) l'analyse sémique,
- 2) l'hyperonyme,
- 3) l'hyponyme,
- 4) le sous-domaine,
- 5) le code ICPC2,
- 6) le code ICD10,
- 7) le terme SNOMED CT,
- 8) le terme MeSh,
- 9) la sphère conceptuelle.

Pour ces catégories, nous fournissons un PID qui correspond à l'URL de la page associée dans notre répertoire TriMED. Par exemple, pour

l'analyse sémique le PID fourni est : <http://purl.org/trimed/dcr/dc/dc-1>¹²⁴.

Le choix de formuler notre répertoire de catégories de données nous permet de désambiguïser la signification de certaines catégories de données. Les utilisateurs qui consultent la ressource peuvent donc trouver toutes les informations nécessaires pour comprendre les données fournies. Le répertoire de TriMED incarne les lignes directrices pour compiler correctement les fiches terminologiques en termes de « cohérence » des données. En particulier, nous nous référons à la dénomination des catégories de données représentées au sein de la ressource, qui doit suivre une certaine rigueur (Wright, 2001) : l'ensemble des valeurs de chaque catégorie de données est défini par le même répertoire. La catégorie /common name/ possède donc un identifiant unique, c'est-à-dire la chaîne de caractères lisible par machine *commonName*, pour laquelle le système n'accepte aucune autre variante. Enfin, comme suggéré par la norme et pour répondre au besoin de réutilisation des données, l'utilisateur peut exporter les informations fournies dans le répertoire dans le format *Comma-Separated Values* (CSV) et *eXtensible Markup Language* (XML).

5.6 Mise en œuvre TBX

Après avoir défini 1) le modèle structurel de TriMED, 2) les catégories de données que nous fournissons dans le modèle de fiche et, 3) leurs spécifications collectées dans un répertoire conçu à cet effet, nous décrivons dans cette section le format d'implémentation choisi pour notre collection de données terminologiques. Sa mise en œuvre est fondée sur la norme ISO-30042 concernant le format *TermBase eXchange* (TBX) décrit dans la section 4.1 du chapitre 4. Ce document définit le cadre TBX, exprimé en langage de balisage XML, pour l'analyse, la représentation descriptive et la diffusion de données terminologiques

124 Un exemple visuel de ce type sera également fourni dans la section 5.8.

structurées. En particulier, l'objectif principal du cadre TBX est de garantir que les données terminologiques peuvent être utilisées dans différents logiciels. Par exemple, dans le processus de traduction avec les outils de Traduction Assistée par Ordinateur (TAO), le format TBX est le plus couramment utilisé pour les systèmes de gestion terminologique (Bowker et Fisher, 2010).

En résumé, le format TBX est basé sur deux composants en interaction : 1) une structure de base qui reflète le métamodèle TMF, et 2) un formalisme visant à définir des modules avec une liste de catégories de données. La combinaison de ces deux composants définit un « dialecte » particulier, c'est-à-dire un langage de balisage XML conforme à TBX. Les dialectes peuvent différer en termes de catégories de données autorisées et de niveaux de métamodèle où ces catégories peuvent être saisies. Sur le site Web [TBXinfo.net](http://www.tbxinfo.net)¹²⁵, trois dialectes publics sont recommandés pour l'échange de terminologie : 1) TBX-Core, 2) TBX-Min, 3) TBX-Basic que nous avons précédemment décrits. Ces dialectes fournissent un ensemble restreint de catégories de données. Pour cette raison, la norme permet donc la formulation des « dialectes privés » afin de représenter les catégories de données terminologiques qui ne sont pas incluses dans les modules publics.

5.6.1 Formulation du dialecte TBX-TriMED

Compte tenu des catégories de données de TriMED, nous avons donc décidé de formuler un nouveau dialecte qui s'appelle « TBX-TriMED » selon les conventions de dénomination établies sur la page suivante : <https://www.tbxinfo.net/tbx-dialectes/>. Pour les catégories de données déjà documentées, la norme précise que les modules publics (*Core*, *Min*, *Basic*) doivent être référencés, tandis que pour les autres catégories, il est nécessaire de compiler un nouveau module TBX spécifique.

Nous avons donc formulé le module *Trimed* qui contient les catégories de données exclusives pour notre ressource. Pour se conformer à la norme, la définition formelle du module *Trimed* a été écrite en prose

125 <https://www.tbxinfo.net/tbx-dialectes/>.

et en langage XML en utilisant le formalisme *TBX Module Description* (TBXMD). Dans l'image 5.4¹²⁶, nous fournissons une liste partielle des catégories de données contenues dans notre module avec toutes les informations pertinentes selon la norme internationale qui sont : l'identifiant, le PID, la classification, la valeur et le niveau TMF. Dans l'annexe B de cet ouvrage, nous fournissons le document qui contient la définition complète du module *Trimed*.

Comme spécifié en ligne sur la page consacrée aux modules privés¹²⁷, la pratique recommandée pour les définitions d'un nouveau module est d'utiliser le répertoire public GitHub¹²⁸. Un *package* de définition du module doit contenir les informations suivantes :

- définition du module en prose ;
- schémas de module ;
- formalisme TBXMD.

À ce propos, le *package* de définition du dialecte TBX-TriMED est disponible en consultant le lien suivant : <https://github.com/trimed-dialect-2020/TriMED>¹²⁹. Le dialecte TBX-TriMED comprend donc les modules suivants :

- *Core* qui contient les catégories de données : /terme/, /date/ et /note/.
- *Min* qui contient (entre autres) les catégories de données : /partie du discours/ et /domaine/.
- *Basic* qui contient (entre autres) les catégories de données : /contexte/, /définition/, /référence croisée/, /genre grammatical/, /source/, /responsabilité/, /type de transaction/ et /référence croisée externe/.
- *Trimed* qui contient les catégories de données : /concept superrordonné/, /concept subordonné/, /sous-domaine/, /identifiant de concept/, /identifiant de langue/, /identifiant/, /nombre grammatical/.

126 La capture d'écran a été prise le 21/10/2021 à partir du répertoire GitHub en ligne qui contient la description du dialecte : <https://bit.ly/3E4pTX4>.

127 <https://www.tbxinfo.net/tbx-private-modules/>.

128 <https://github.com>.

129 Actuellement, le *package* a été envoyé pour validation au contact indiqué sur le site officiel de TBXinfo.net : <https://www.tbxinfo.net/tbx-private-dialects/>.

Data Category Module Definition

Name: Trimed

Namespace: http://shiny.dei.unipd.it/TriMED/data_category_repository/

Description: The Trimed Module is a data category module collecting data categories selected from 1) DatCatInfo (<http://www.datcatinfo.net/#/>) and 2) TriMED Data Category Repository.

Data Categories:

Name	PID	Classification	Value(s)	Level (TMF)
superordinateConcept	http://www.datcatinfo.net/datcat/DC-495	<descrip>		conceptEntry
subordinateConcept	http://www.datcatinfo.net/datcat/DC-490	<descrip>		conceptEntry
subDomain	http://purl.org/trimed/dcr/dc/dc-7	<descrip>		conceptEntry
conceptIdentifier	http://www.datcatinfo.net/datcat/DC-139	<descrip>		conceptEntry
languageIdent	http://www.datcatinfo.net/datcat/DC-279	<descrip>		langSec
identifier	http://www.datcatinfo.net/datcat/DC-3894	<descrip>		termSec
grammaticalNumber	http://www.datcatinfo.net/datcat/DC-251	<termNote>	singular, plural, dual, mass, otherNumber	termSec
derivative	http://www.datcatinfo.net/datcat/DC-4611	<termNote>		termSec
pronunciation	http://www.datcatinfo.net/datcat/DC-407	<termNote>		termSec

Image 5.4 – Formalisme partiel du module *Trimed*

/formes dérivées/, /prononciation/, /étymologie/, /variante orthographique/, /abréviation/, /expansion/, /acronyme/, /analyse sémique /, /synonyme/, /hyponyme/, /hyperonyme/, /unité phraséologique/, /collocation/, /registre/, /nom populaire/, /nom scientifique/, /code ICPC2/, /code ICD10/, /terme SNOMED CT/, /terme MeSh/, /relation entre concepts/ et /sphère conceptuelle/.

Une fois que les composants du dialecte TBX-TriMED et que les catégories de données pour chaque module ont été définies, nous fournissons ci-dessous un extrait d’instance de notre dialecte :

```
<?xml version="1.0" encoding="UTF-8"?>

<tbx type="TBX-TriMED" style="dct" xml:lang="en"
xmlns="urn:iso:std:iso:30042:ed-2"
xmlns:min="http"
xmlns:basic="http://www.tbxinfo.net/ns/basic"
xmlns:trimed="https://github.com/trimed-dialect-2020/TriMED">
  <tbxHeader>
  <fileDesc>
```

```

<sourceDesc>
  <p>A sample of the TriMED termbase
  consisting of one concept entry</p>
</sourceDesc>
</fileDesc>
</tbxHeader>
<text>
  <body>
    <conceptEntry>
      <trimed:conceptIdentifier>A2</trimed:conceptIdentifier>
      <trimed:conceptualSphere>A</trimed:conceptualSphere>
      <basic:crossReference></basic:crossReference>
      <trimed:conceptRelation></trimed:conceptRelation>
      <basic:transaction>origination</basic:transaction>
      <basic:responsibility>Federica Vezzani</basic:responsibility>
      <date>04/04/2020</date>
      <note></note>
      <trimed:superordinateConcept></trimed:superordinate-
      Concept>
      <trimed:subordinateConcept></trimed:subordinateConcept>
      <min:subjectField>medicine</min:subjectField>
      <trimed:subDomain></trimed:subDomain>

      <langSec xml:lang="en">
        <basic:definition>The act or process of identifying or
        determining the nature and cause of a disease or injury
        through evaluation of patient history,
        examination, and review of laboratory data.</basic:definition>
        <basic:externalCrossReference> https://www.thefreediction-
        nary.com/diagnosis
        </basic:externalCrossReference>
        <basic:source></basic:source>
        <note></note>
        <trimed:semicAnalysis>/process/
        /identification/ /nature/ /cause/
        /disease/</trimed:semicAnalysis>

```

```

<termSec>
<term>diagnosis</term>
<trimed:identifier>TPLT18167</trimed:identifier>
<basic:transaction>origination</basic:transaction>
<basic:responsibility>student TPLT</basic:responsibility>
<date>01/01/2018</date>
<note></note>
<min:partOfSpeech>noun</min:partOfSpeech>
<basic:grammaticalGender></basic:grammaticalGender>
<trimed:grammaticalNumber></trimed:grammaticalNumber>
<trimed:pronunciation></trimed:pronunciation>
<trimed:etymology></trimed:etymology>
<trimed:variant>plural: diagnoses</trimed:variant>
<trimed:acronym></trimed:acronym>
<trimed:fullForm></trimed:fullForm>
<trimed:abbreviation></trimed:abbreviation>
<trimed:derivative>to diagnose, to diagnosticate,
diagnostic, diagnostically, diagnostician</trimed:derivative>

<trimed:collocation>accurate diagnosis, age of diagnosis, clinical
diagnosis, correct diagnosis,definitive diagnosis,
early diagnosis, differential diagnosis,
diagnosis of, difficulty of diagnosis</trimed:collocation>
<trimed:phraseologicalUnit></trimed:phraseologicalUnit>

<trimed:synonym>analysis, examination, investigation,
conclusion, interpretation</trimed:synonym>
<trimed:hypernym></trimed:hypernym>
<trimed:hyponym></trimed:hyponym>
<basic:crossReference></basic:crossReference>
<basic:context>Consequently, people with suspected
hypertension have been required to undergo repeated measurements
of their clinic BP on repeated clinic visits to confirm or refute
the diagnosis of hypertension</basic:context>
<basic:externalCrossReference>https://bit.ly/3n8ud0B
</basic:externalCrossReference>
<basic:source></basic:source>

```

```

<trimed:register>specialized</trimed:register>
<trimed:commonName>diagnosis</trimed:commonName>
<trimed:scientificName>diagnosis</trimed:scientificName>
<trimed:mesh>https://bit.ly/2Z2P8da</trimed:mesh>
<trimed:snomed>https://bit.ly/3BV8ZsW</trimed:snomed>
<trimed:icpc2Code></trimed:icpc2Code>
<trimed:icd10Code></trimed:icd10Code>

```

```
</termSec>
```

```
</langSec>
```

```

<langSec xml:lang="fr">
<basic:definition>Connaissance que l'on peut avoir d'une
maladie
en observant les signes de celle-ci ; art du diagnostic.</
basic:definition>
<basic:externalCrossReference>https://bit.ly/3pegwQk
</basic:externalCrossReference>
<basic:source></basic:source>
<note></note>
<trimed:semicAnalysis>etude/maladie/symptome/medecine/
</trimed:semicAnalysis>

```

```
<termSec>
```

```

<term>diagnose</term>
<trimed:identifier>TSF18068</trimed:identifier>
<basic:transaction>Origine</basic:transaction>
<basic:responsibility>etudiant TRADFR</basic:responsibi-
lity>
<date>01/01/2018</date>
<note></note>

```

```

<min:partOfSpeech>nom</min:partOfSpeech>
<basic:grammaticalGender>feminin</basic:grammati-
calGender>
<trimed:grammaticalNumber>singulier</trimed:grammati-
calNumber>

```


<trimed:pronunciation>\djagno:z</trimed:pronunciation>
 <trimed:etymology>Empr. au gr. proprement action de dis-
 cerner
 spec. terme de med. diagnose, diagnostic.</trimed:etymology>
 <trimed:variant></trimed:variant>
 <trimed:acronym></trimed:acronym>
 <trimed:fullForm></trimed:fullForm>
 <trimed:abbreviation></trimed:abbreviation>
 <trimed:derivative>diagnostic, diagnostique, diagnostiquer</
 trimed:derivative>

<trimed:collocation>faire une diagnose, effectuer une dia-
 gnose</trimed:collocation>
 <trimed:phraseologicalUnit></trimed:phraseologicalUnit>

<trimed:synonym>diagnostic</trimed:synonym>
 <trimed:hypernym></trimed:hypernym>
 <trimed:hyponym></trimed:hyponym>
 <basic:crossReference></basic:crossReference>

<basic:context>Les courbes de temperature (. .) ne sont pas
 suffisamment typiques pour qu'elles puissent servir a la dia-
 gnose</basic:context>
 <basic:externalCrossReference><https://bit.ly/3pegwQk>
 </basic:externalCrossReference>
 <basic:source></basic:source>

<trimed:register>specialise</trimed:register>
 <trimed:commonName>diagnose</trimed:commonName>
 <trimed:scientificName>diagnose</trimed:scientificName>
 <trimed:mesh><https://bit.ly/2Z2P8da></trimed:mesh>
 <trimed:snomed><https://bit.ly/3BV8ZsW></trimed:snomed>
 <trimed:icpc2Code></trimed:icpc2Code>
 <trimed:icd10Code></trimed:icd10Code>

</termSec>

</langSec>

```

<langSec xml:lang="it">
<basic:definition>Procedura consistente nell'interpretazione di
segni e sintomi raccolti quali manifestazioni di un processo
patologico in atto o pregresso</basic:definition>
<basic:externalCrossReference>https://bit.ly/3DSh6Y8
</basic:externalCrossReference>
<basic:source></basic:source>
<note></note>
<trimed:semicAnalysis>procedura/
/interpretazione/ /sintomi/ /malattia/</trimed:semicAnalysis>

```

```

<termSec>
<term>diagnosi</term>
<trimed:identifier>TPLT18168</trimed:identifier>
<basic:transaction>origination</basic:transaction>
<basic:responsibility>studente TPLT</basic:responsibility>
<date>01/01/2018</date>
<note></note>

```

```

<min:partOfSpeech>sostantivo</min:partOfSpeech>
<basic:grammaticalGender>femminile</basic:grammaticalGender>
<trimed:grammaticalNumber>invariabile
</trimed:grammaticalNumber>
<trimed:pronunciation>dia-gno-si</trimed:pronunciation>
<trimed:etymology></trimed:etymology>
<trimed:variant></trimed:variant>
<trimed:acronym></trimed:acronym>
<trimed:fullForm></trimed:fullForm>
<trimed:abbreviation></trimed:abbreviation>
<trimed:derivative>diagnosticare, diagnostico, diagnosta</trimed:derivative>

```

```

<trimed:collocation>fare la diagnosi,
formulare una diagnosi</trimed:collocation>
<trimed:phraseologicalUnit></trimed:phraseologicalUnit>

```

```

    <trimed:synonym>anamnesi, prognosi, analisi, accertamento</trimed:synonym>
    <trimed:hypernym></trimed:hypernym>
    <trimed:hyponym></trimed:hyponym>
    <basic:crossReference></basic:crossReference>

    <basic:context></basic:context>
    <basic:externalCrossReference>https://bit.ly/2Z5ZGrZ</basic:externalCrossReference>
    <basic:source></basic:source>

    <trimed:register>specializzato</trimed:register>
    <trimed:commonName>diagnosi</trimed:commonName>
    <trimed:scientificName>diagnosi</trimed:scientificName>
    <trimed:mesh>https://bit.ly/2Z2P8da</trimed:mesh>
    <trimed:snomed>https://bit.ly/3BV8ZsW</trimed:snomed>
    <trimed:icpc2Code></trimed:icpc2Code>
    <trimed:icd10Code></trimed:icd10Code>

  </termSec>
</langSec>
</conceptEntry>
</body>
</text>
</tbx>

```

Cette instance montre la représentation TBX de la fiche terminologique multilingue du concept « A2 » désignée par les termes *diagnosis* (EN), *diagnose* (FR) et *diagnosi* (IT). Dans l'élément racine < *tbx* >, les informations suivantes sont spécifiées :

1. les espaces de noms (xmlns) du TBX-Core¹³⁰, du TBX-Min, du TBX-Basic et du TBX-TriMED ;
2. la valeur de l'attribut type, c'est-à-dire le nom du dialecte TBX-TriMED ;

130 C'est-à-dire *urn: iso: std: iso:30042: ed-2* qui doit être utilisé comme espace de noms par défaut pour les instances de document TBX de tous les dialectes.

3. le style de l'instance, c'est-à-dire le style *Data Category as Tag* (DCT) que nous avons choisi¹³¹ ;
4. la langue de travail (xml:lang) du document.

L'instance illustre comment les données terminologiques de TriMED sont organisées. Elle suit le métamodèle hiérarchique TMF : dans cet ordre, i) < *conceptEntry* >, ii) < *langSec* >, et iii) < *termSec* > sont définis.

Au niveau du concept, représenté par son identifiant (id = A2), les catégories de données fournies sont : /sphère conceptuelle/, /référence croisée/, /relation entre concepts/, /type de transaction/, /responsabilité/, /date/, /note/, /concept superordonné/, /concept subordonné/, /domaine/ et /sous-domaine/. Le même concept est ensuite exprimé en trois sections de langue, chacune contenant les termes désignant le concept en anglais, en français et en italien.

Au niveau de la langue, nous fournissons les informations sur la /définition/, /référence croisée externe/, /source/, /note/ et /analyse sémique/ du concept. Chaque < *langSec* > contient un < *termSec* > avec le terme désignant le concept et d'autres catégories de données, telles que /partie du discours/, /genre grammatical/, /formes dérivées/ etc. Pour chaque catégorie de données de l'instance TBX, le module correspondant est spécifié dans le nœud XML.

5.6.2 Position des catégories de données

La norme ISO-30042 : 2019 définit, enfin, les contraintes de positionnement des catégories de données pour chaque niveau du métamodèle TMF. En général, les catégories de données les plus génériques ont un certain degré de liberté en termes de positionnement. Par exemple, la /définition/ peut être placée à la fois au niveau du concept et au niveau de la langue, tandis que les informations descriptives du terme, telles que le /nombre grammatical/ ne doivent être placées qu'au niveau de la section de terme.

131 La norme permet de choisir entre le style DCT et DCA (*Data Category as Attribute*). Les différences entre les deux sont illustrées au lien suivant : <https://www.tbxinfo.net/dca-v-dct/>.

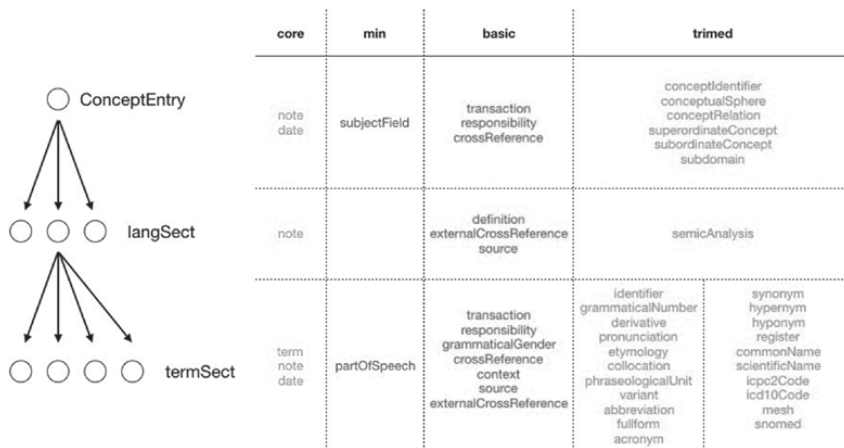


Image 5.5 – Position des catégories dans TriMED

Dans l’image 5.5, nous montrons les catégories de données de TriMED associées à chaque niveau hiérarchique de la structure TMF (concept, langue et terme) et les modules TBX auxquels les catégories appartiennent (*Core*, *Min*, *Basic* et *Trimed*). Les catégories ont été positionnées à l’aide des indications fixées par la norme où les niveaux de contraintes sont explicites. Pour ces catégories sans restrictions, nous avons volontairement décidé leur niveau de positionnement. Les catégories qui apparaissent au niveau du concept concernent des informations générales de type administratif sur l’entrée terminologique, telles que : /date/, /note/, /type de transaction/, /responsabilité/, /référence croisée/ et /identifiant du concept/. D’autres catégories qui devraient être partagées par toutes les langues sont fournies toujours au niveau conceptuel comme : /domaine/, /sous-domaine/, /sphère conceptuelle/, /relation entre concepts/, /concept superordonné/ et /concept subordonné/. Au niveau de la langue, les catégories fournies sont : /note/, /définition/, /analyse sémique/, /référence croisée externe/ et /source/. En particulier, nous avons choisi de positionner la catégorie de la /définition/ au niveau de la langue (et pas au niveau du concept comme, par exemple, dans la ressource TermScience de Khayari et al. (2006)) afin de ne pas attribuer une identité linguistique à un concept dont la

signification devrait être équivalente pour toutes les langues analysées dans l'entrée terminologique. Si la définition avait été placée au niveau du concept, une langue arbitraire aurait dû être choisie pour sa représentation. Au lieu de cela, placer la définition au niveau de la langue permet de l'exprimer dans chacune des langues de travail de la ressource (anglais, italien et français). Ainsi, le concept reste une entité *super partes* indépendante de toute attribution linguistique privilégiée¹³². Enfin, une telle décision permet de saisir, au niveau inférieur à la section langue, tous les termes (un par section de terme) qui correspondent exactement à cette définition avec toutes les catégories de données décrites précédemment. De cette façon, un concept peut être exprimé dans une langue par le truchement de termes équivalents véhiculant la même signification.

Cependant, le problème de la gestion des termes partiellement synonymes n'est toujours pas résolu. Par exemple, si nous considérons les termes « fièvre » et « pyrexie », nous pouvons sans aucun doute dire que ceux-ci véhiculent la même signification, c'est-à-dire l'augmentation de la température corporelle. Toutefois, les termes n'auront pas les mêmes contextes d'occurrence, car ils appartiennent à des registres linguistiques différents : « pyrexie » comme nom scientifique, « fièvre » comme nom populaire. Or, si les conditions de synonymie prévoient 1) l'équivalence de signification et 2) leur substituabilité dans n'importe quel contexte (Jezek, 2005), les termes « pyrexie » et « fièvre » ne peuvent pas être traités en tant que synonymes purs dans la même entrée terminologique. Ces types de considérations nous permettent de mener des réflexions sur certains aspects formels de la normalisation de la structure d'une ressource terminologique numérique. À l'heure actuelle, nous avons décidé de gérer ce problème en renseignant autant

132 Une autre solution qui aurait pu être adoptée est celle de 1) formuler des définitions ontologiques formelles indépendantes de la langue à positionner au niveau du concept, et 2) développer ou mettre à jour des définitions en langage naturel à positionner au niveau de la langue. À l'heure actuelle, nous n'approfondissons pas cette question car au-delà de la portée du projet de recherche, cependant nous renvoyons le lecteur intéressé par ce sujet aux publications suivantes : Roche (2015) et, en particulier, pour le domaine médical (sous-domaine endométriose) les articles liés au projet EndoTerm de Carvalho et al. (2018) ; Roche et al. (2019).

de catégories de données que possible afin de classer les différents phénomènes linguistiques (notamment la variation terminologique). Un espace particulier est consacré aux notes du terminologue qui peuvent être saisies au niveau conceptuel, linguistique et terminologique.

5.7 Application Web TriMED

Après avoir décrit la conception de la ressource et l'organisation des données selon les principes de la « terminologie FAIR », nous consacrerons cette section à la présentation iconographique de l'application Web de la ressource TriMED consultable en ligne par le grand public¹³³. En particulier, nous analyserons les aspects de visualisation et d'interaction dans une perspective inter- et intra-linguistique des trois catégories d'utilisateurs identifiées. Comme nous l'avons avancé dans les sections précédentes, la ressource a été conçue afin de :

1. aider les patients à comprendre correctement les informations médicales, compte tenu de l'aspect de la variation diastratique de la terminologie ;
2. soutenir le traducteur dans le processus de traduction spécialisée en fournissant un cadre sur le comportement syntaxique, sémantique et phraséologique du terme source et de son traduisant en langue cible ;

133 Toutes les captures d'écran illustrées dans les sections suivantes ont été prises le 21/10/2021 et proviennent du site Web du projet : <https://purl.org/trimed>



Image 5.6 – Page d'accueil TriMED

3. fournir un point d'accès unique pour la consultation des professionnels de la santé aux autres terminologies, nomenclatures ou codes de classification internationaux généralement utilisés par des experts.

L'application Web de TriMED, implémentée à l'aide du code source fourni par Vezzani et al. (2018) avec le package Shiny R (Chang et al., 2021), a été développée au sein du groupe de recherche *Information Management System*¹³⁴ du Département de Génie de l'Information de l'Université de Padoue.

La ressource TriMED est accessible en ligne via l'application Web disponible au lien suivant : <https://purl.org/trimed>. L'image 5.6 capture la page principale de l'application Web en français. À partir de cette visualisation, l'utilisateur peut sélectionner différents onglets qui permettent d'accéder à différentes informations liées à la conception et à la mise en œuvre de la ressource. À partir de l'onglet de la page

¹³⁴ <http://ims.dei.unipd.it>.

Projet

TriMED est une base de données terminologique multilingue, conçue pour répondre à divers problèmes liés à la complexité de la terminologie médicale en trois langues: italien, anglais et français. En particulier, la ressource est conçue pour:

1. Les patients, qui peuvent chercher les équivalents de termes techniques formulés dans un langage plus familier et plus compréhensible;
2. Les traducteurs, qui peuvent consulter une fiche terminologique structurée qui fournit un large éventail d'informations dans la langue source et la langue cible;
3. Les médecins qui, en termes d'interopérabilité, peuvent consulter d'autres ressources médicales utiles telles que MeSH terms, SNOMED CT etc.

Le travail de compilation des fiches est en cours depuis 2018. Pour le moment, nous pouvons fournir le nombre suivant de fiches compilées par langue:

- 436 pour l'anglais
- 410 pour le français
- 289 pour l'italien

La ressource a été mise en œuvre conformément aux dernières normes ISO pour la gestion de la terminologie, ce qui répond aux principes sous-jacents à la formulation des données FAIR dans le contexte de l'European Open Science Cloud (EOSC).

La description des catégories de données est fournie dans le **Répertoire des catégories de données** TriMED. Notre fiche terminologique contient 42 catégories de données différentes.

Personnes

Le projet TriMED a été conçu et développé par **Federica Vezzani** actuellement doctorante au Département d'études linguistiques et littéraires de l'Université de Padoue sous la direction de la Professeure Geneviève Henrot et du Professeur Giorgio Maria Di Nunzio.

La ressource TriMED a été mise en œuvre en tant qu'application Web Shiny par le Professeur Giorgio Maria Di Nunzio, qui fait partie du groupe Information Management System du Département de génie informatique de l'Université de Padoue.

Image 5.7 – À propos de TriMED

d'accueil, les utilisateurs peuvent sélectionner l'une des trois catégories de consultation, puis accéder aux trois interfaces qui fournissent différents sous-ensembles d'informations linguistiques en fonction de l'identification de l'utilisateur. L'accès est autorisé en cliquant directement sur l'image respective ou sur l'étiquette attribuée à chaque catégorie d'utilisateur : « Patient », « Traducteur » et « Médecin ». En outre, les utilisateurs peuvent accéder directement à l'interface qui leur est dédiée en cliquant directement sur l'onglet « Utilisateurs ».

Dans la section suivante (image 5.7), des informations descriptives sont fournies concernant la nature du projet, le nombre de fiches terminologiques actuellement disponibles sur la ressource¹³⁵, et les personnes qui ont contribué à son développement. Les informations concernant les collaborations et les publications relatives au projet sont fournies sur la même page, mais ne figurent pas dans l'image 5.7 pour des raisons d'espace. En particulier, le projet TriMED est promu par

135 Le nombre fourni est constamment mis à jour au fur et à mesure de la compilation des fiches qui est toujours en cours.

l'Unité de Coordination de Terminologie (TermCoord)¹³⁶ de l'Union européenne avec laquelle une collaboration est en cours concernant le projet nommé « Terminologie Sans Frontières »¹³⁷ qui reflète et soutient l'objectif de la Direction Générale de la Traduction (DG TRAD) du Parlement européen de communiquer avec les citoyens dans une langue compréhensible. Le développement de la ressource TriMED est également soutenu par le projet ExaMode (*Extreme-scale Analytics via Multimodal Ontology Discovery & Enhancement project*)¹³⁸, financé par l'Union européenne dans le cadre du programme Horizon 2020¹³⁹.

Depuis la page principale de la ressource TriMED, il est possible d'accéder directement au « Répertoire des Catégories de Données », implémenté selon la norme ISO-12620 de 2019, dont l'application Web sera discutée dans la section 5.8 de ce chapitre. De plus, l'application Web fournit une recherche rapide des contacts dans la section « Contacts » et le « Journal des modifications » afin de conserver la traçabilité des dernières mises à jour de la ressource. Enfin, le dernier onglet d'affichage de la barre horizontale permet de choisir la langue de navigation de la ressource : actuellement, TriMED est une ressource multilingue envisageant l'exploration en ligne en français, en anglais et en italien, mais des développements futurs sont envisagés.

5.7.1 Utilisateurs

Les utilisateurs qui s'identifient dans une des trois catégories peuvent accéder à l'interface qui leur est dédiée via l'onglet « Utilisateurs », ou directement via la page d'accueil en cliquant sur l'image ou sur l'étiquette correspondante. Les URL suivantes correspondent aux liens directs vers l'interface du :

- Patient : <https://purl.org/trimed/patient>
- Traducteur : <https://purl.org/trimed/translator>
- Médecin : <https://purl.org/trimed/physician>

136 <https://termcoord.eu>.

137 <https://yourterm.org>.

138 <https://www.examode.eu>.

139 <https://ec.europa.eu/programmes/horizon2020/en>.

Dans les sous-sections suivantes, nous décrivons l’affichage et l’interaction de l’utilisateur avec le système en ligne.

5.7.1.1 Le patient

Pour gérer le problème de la variation diastratique dans les dialogues entre les experts et les non-experts et pour faciliter la compréhension des informations médicales, les patients peuvent rechercher un terme technique puis consulter le terme populaire correspondant et sa définition, si disponible.

Un patient à la recherche d’informations sur un terme populaire commence la requête en sélectionnant une des langues disponibles (anglais, français ou italien), puis en saisissant le terme dans le champ de recherche. Le système – grâce à la fonction *selectInput*¹⁴⁰ du package Shiny RStudio – filtre automatiquement les mots, caractère par caractère, et affiche les alternatives possibles dans la boîte. Nous avons décidé d’afficher à la fois le terme populaire et le terme technique lors de la recherche pour fournir aux utilisateurs un retour immédiat. Lorsque



Image 5.8 – Interface « Patient »

140 <https://shiny.rstudio.com/reference/shiny/0.12.2/selectInput.html>.

l'utilisateur sélectionne le terme, la définition et les informations sur la variation terminologique sont affichées sur la page.

Dans l'image 5.8, nous avons simulé la recherche d'un patient qui sélectionne la langue française et demande au système l'équivalent populaire du terme scientifique « rougeole ». La ressource renvoie en sortie l'équivalent couramment utilisé par le grand public « première maladie » et sa définition « Maladie infectieuse, contagieuse et épidémique, due à un virus, caractérisée par une éruption sur la peau de taches rouges peu saillantes, apparaissant successivement à la tête, sur le corps, puis sur les membres », extraite à partir du portail lexical du Centre National de Ressources Textuelles et Lexicales (CNRTL)¹⁴¹.

5.7.1.2 Le traducteur

Dans la sous-section 5.1.2 précédente, nous avons présenté plusieurs aspects (plus ou moins problématiques) liés à la traduction de textes médicaux. Dans ce contexte, avec la base de données TriMED, nous nous efforçons de fournir un outil utile à l'usage des traducteurs technico-scientifiques afin de soutenir le processus de traduction spécialisée. Le modèle de fiche terminologique consultable par le traducteur comprend 42 catégories de données différentes pour le terme source et le terme cible, afin d'offrir un cadre complet sur le comportement syntaxique, sémantique et phraséologique des termes.

Dans l'image 5.9, nous simulons la recherche d'un traducteur qui sélectionne la paire de langues de travail anglais (langue source) et italien (langue cible), et saisit dans la boîte de recherche le terme objet d'analyse en langue source : *scintigraphy*. Le système renvoie automatiquement en sortie une fiche bilingue contenant des informations concernant le terme de départ du côté gauche (*scintigraphy*) et le traduisant (*scintigrafia*) du côté droit. Compte tenu du grand nombre d'informations terminologiques consultables, la fiche a été organisée en six panneaux fournissant différentes catégories de données résumées ci-dessous :

141 <https://cnrtl.fr/definition/rougeole>.

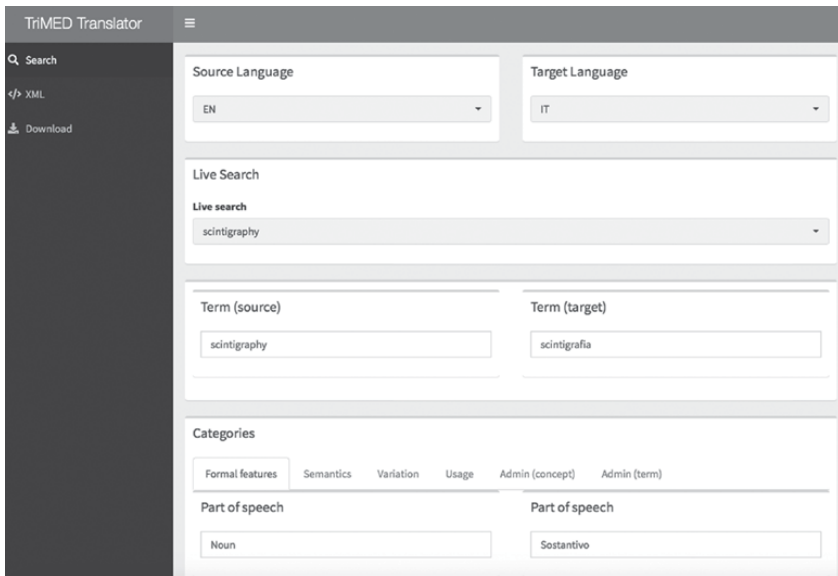


Image 5.9 – Interface « Traducteur »

1. « Caractéristiques formelles » : /partie du discours/, /genre grammatical/, /nombre grammatical/, /prononciation/, /étymologie/ et /formes dérivées/ ;
2. « Sémantique » : /définition/, /référence croisée externe /, /source/, /note/ et /analyse sémique/ ;
3. « Variation » : /variante orthographique/, /acronyme/, /expansion/, /abréviation/, /synonyme/, /hyperonyme/, /hyponyme/, /référence croisée/, /nom commun/ et /nom scientifique/ ;
4. « Utilisation » : /contexte/, /référence croisée/, /source/, /collocation/, /unité phraséologique/ et /registre/ ;
5. « Informations administratives sur le concept » : / identifiant du concept/, /date/, /note/, /responsabilité/, /type de transaction/, /sphère conceptuelle/, /domaine/, /sous-domaine/, /relation entre concepts/, /référence croisée/, /concept superordonné/ et /concept subordonné/ ;

6. « Informations administratives sur le terme » : /identifiant du terme/, /type de transaction/, /responsabilité/, /date/, /note/.

L'image 5.10 montre un exemple d'affichage du panneau « Usage » de la fiche terminologique bilingue *Scintigraphy – Scintigrafia*. Le traducteur peut ainsi accéder facilement au contexte dans lequel le terme est utilisé, revenir à sa source et, comme le montre l'image, consulter les cas de collocations récurrentes avec les deux termes analysés.

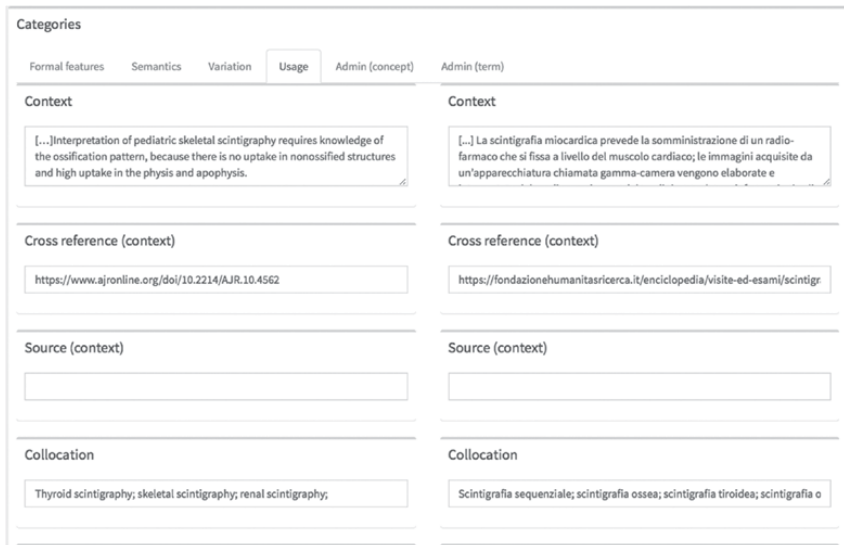


Image 5.10 – Panneau « Usage » fiche bilingue *Scintigraphy – Scintigrafia*

La ressource TriMED se distingue d'autres ressources sur le marché non seulement par la quantité d'informations fournies, mais également par la possibilité offerte aux traducteurs professionnels de réutiliser les données terminologiques. Pour répondre aux principes formulés à travers la modélisation conceptuelle de la « terminologie FAIR », l'application permet au traducteur d'afficher et de télécharger le format TBX standard de la fiche, compatible avec tout logiciel de traduction assistée par ordinateur (TAO) qui prend en charge la norme ISO-30042 du 2019.

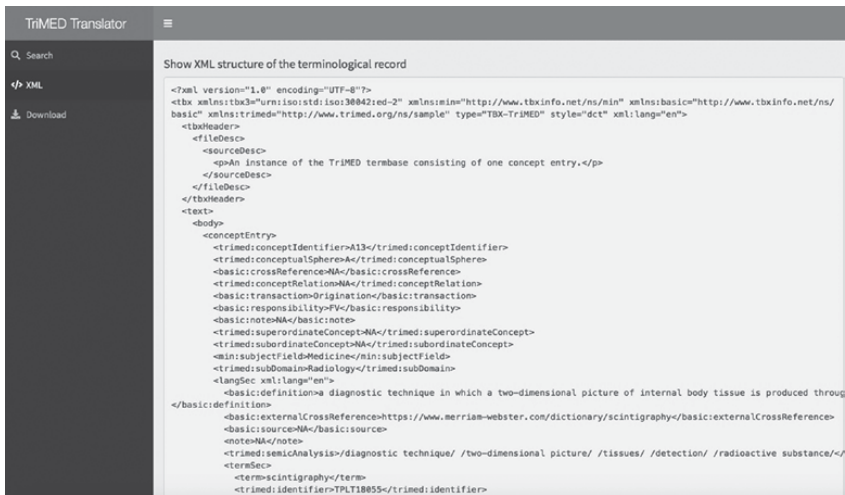


Image 5.11 – Affichage de la fiche en format TBX

À partir de la barre latérale gauche de l'interface, le traducteur peut sélectionner l'une des trois options de visualisation :

« Rechercher », « XML » et « Télécharger ». En cliquant sur l'étiquette « XML », l'utilisateur pourra accéder à la fiche bilingue consultée qui apparaîtra automatiquement dans le format standard (image 5.11) et pourra être téléchargée localement pour être réutilisée.

5.7.1.3 Le médecin

L'interopérabilité des ressources est une exigence fondamentale lors de l'adoption d'une approche FAIR de la terminologie. Dans ce contexte, l'un des objectifs les plus ambitieux de TriMED est de fournir un système capable de collecter des informations médicales à partir de plusieurs sources, via un point d'accès unique. À ce propos, lors de la consultation de la base de données TriMED, les médecins peuvent rechercher un terme technique dans l'une des trois langues de travail. Ils peuvent ainsi accéder aux terminologies, aux ontologies et aux systèmes de classification médicale standard,

tels que les ressources *Medical Subject Headings* (MeSH)¹⁴² et *Systematized Nomenclature of Human and Veterinary Medicine – Clinical Terms* (SNOMED CT)¹⁴³.

Dans l'image 5.12, nous montrons la visualisation de l'interface consultable par le médecin. Tout d'abord, l'utilisateur peut sélectionner la langue de travail puis, comme dans le cas de l'interface consacrée au patient, il peut saisir le terme dans le champ de recherché : le système filtrera automatiquement les mots et les caractères indiquant les options possibles pour l'utilisateur. Le système établira automatiquement le lien avec d'autres ressources externes et fournira un accès direct au même terme et/ou à la combinaison de termes qui sont corrélés à la recherche principale. L'exemple fourni dans l'image 5.12 simule l'expert recherchant le terme anglais *Diphtheria*. La sortie automatique de la ressource fournira à l'utilisateur les liens pour accéder directement à d'autres informations médicales sémantiquement liées, telles que *Diphtheria Antitoxin*, *Diphtheria Toxin*, et *Diphtheria Toxin Receptor*, fournies, dans ce cas spécifique, par la ressource MeSH *terms*. À l'heure actuelle, les liens directs des termes contenus dans la ressource TriMED sont garantis avec les ressources MeSH et SNOMED. Nous sommes en train de mettre en œuvre un système qui permet également la cartographie automatique des codes internationaux de classification des maladies de la CIM10¹⁴⁴ et ICPC2¹⁴⁵. Enfin, le système montre à l'utilisateur au bas de la page les termes qui désignent le concept dans les trois langues de travail.

142 <https://meshb.nlm.nih.gov/search>.

143 <http://www.snomed.org>.

144 <https://www.who.int/classifications/icd/icdonlineversions/en/>.

145 <https://www.who.int/classifications/icd/adaptations/icpc2/en/>.

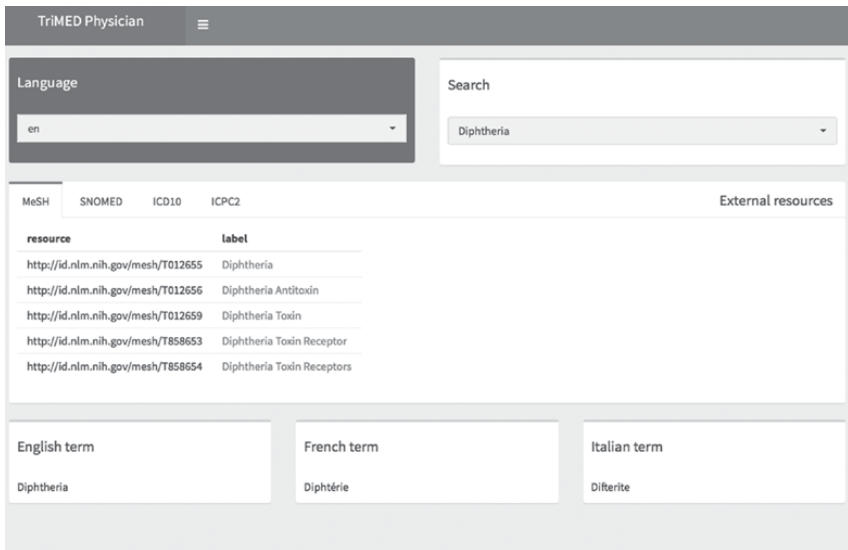


Image 5.12 – Interface « Médecin »

5.8 Application Web RCD

Dans la section 5.5 de ce chapitre, nous avons argumenté et illustré la conception et l’implémentation du Répertoire de Catégories de Données (RCD) de TriMED qui, à notre connaissance, est le premier répertoire implémenté selon la norme ISO-12620 de 2019. Cet outil a été conçu pour la documentation des métadonnées terminologiques de TriMED : il contient une liste descriptive des 42 catégories de données présentes dans la ressource multilingue¹⁴⁶. En particulier, le répertoire a été élaboré afin de garantir la trouvabilité, l’accessibilité, la réutilisabilité et l’interopérabilité des (méta)données terminologiques conformément à la formulation des principes FAIR appliqués à la terminologie. L’application Web a été implémentée à l’aide du *package Shiny R* (Chang et al., 2021) et est disponible en ligne directement à partir de

146 La liste est disponible en ligne et sous forme de tableau dans l’Annexe A mentionné ci-dessus.

la page d'accueil de la ressource TriMED en cliquant sur l'onglet de référence « Répertoire des Catégories de Données », ou au lien direct suivant : <http://purl.org/trimed/dcr>

Dans l'image 5.13, nous montrons un exemple de visualisation pour une nouvelle catégorie de données que nous avons introduite dans TriMED, à savoir l'analyse sémique/. Comme pour l'interface principale de TriMED, l'utilisateur peut tout d'abord sélectionner l'une des trois langues de travail. Ensuite, il peut saisir la dénomination de la catégorie dans la boîte de recherche : le système filtrera automatiquement les mots et les caractères indiquant les options possibles pour l'utilisateur. Une fois sélectionnée la catégorie de données, le système affichera en sortie automatique ce qui est défini par la norme ISO-12620 comme une « spécification » de la catégorie de données, c'est-à-dire une collection de toutes les métadonnées utiles pour son utilisation lors de la compilation d'une fiche terminologique.

Data Category Repository - TriMED

The screenshot shows the 'Data Category Repository - TriMED' interface. On the left, there are two dropdown menus: 'Langue' (Language) set to 'fr' and 'Catégorie de donnée' (Data Category) set to 'Analyse sémique'. On the right, there is a 'Description' tab and an 'XML' tab. Below the tabs, the following information is displayed:

PID: <http://purl.org/trimed/dcr/dc/dc-1>

Identifiant: semicAnalysis
 Module: Trimed
 Niveau (TMF): langSec
 Classification: <descrip>
 Typologie de contenu: chaîne de caractères
 Valeur(s): NA

Description: L'analyse sémique d'une production sémiotique, un texte par exemple, vise à en dégager les sèmes, c'est-à-dire les éléments de sens, à définir leurs regroupements (isotopies et molécules) et à stipuler les relations entre ces regroupements (relations de présupposition, de comparaison, etc., entre isotopies).

Explication: NA
 Note: NA
 Exemple: Pathologie: /étude/ /médecine/ /développement/ /cause /symptôme/ /maladie/. Maladie: /altération/ /santé/ /être vivant/ /signe/ /symptôme/ /causes internes/ /causes externes/ /évolution/ /organisme/.

Download CSV

Image 5.13 – Interface RCD : catégorie de données /analyse sémique/

Le répertoire de TriMED fournit également à l'utilisateur un système d'affichage multilingue complet. En ce sens, des informations telles que « description », « explication », « notes » et « exemples » sont

adaptées en fonction de la langue de consultation de l'utilisateur (voir image 5.13). Enfin, ce type d'affichage multilingue se reflète également dans la génération automatique du fichier XML : comme le recommande la norme, l'utilisateur peut télécharger et réutiliser une spécification de la catégorie de données dans un format lisible par la machine et adapté à sa langue de travail.

5.9 Collection terminologique

Les utilisateurs qui consultent la ressource TriMED auront donc accès à une collection de termes qui relèvent de différentes sphères conceptuelles du domaine médical. Les travaux sur la population de la base de données terminologique ont commencé en 2018 et se poursuivent grâce au développement de projets terminologiques et de traduction spécialisée : en conséquence, les chiffres fournis dans cette section augmentent constamment. À l'heure actuelle, la base de données recueille un total de 1.135 fiches terminologiques multilingues affichées pour les trois utilisateurs¹⁴⁷. Le nombre de fiches normalisées disponible par langue est :

- 436 pour l'anglais ;
- 410 pour le français ;
- 289 pour l'italien.

Les fiches terminologiques ont été élaborées grâce à la précieuse collaboration de différents groupes de travail composés d'étudiants des cours de *Tecnologie per la traduzione*¹⁴⁸ (Systèmes de traduction assistée par ordinateur) et *Traduzione specializzata francese 2*¹⁴⁹ (Traduction spécialisée française 2), pendant les années académiques 2017/2018, 2018/2019, 2019/2020, dans le cadre du Master en Langues Modernes pour

147 Dernière mise à jour 21/10/2021.

148 <https://didattica.unipd.it/off/2019/LM/SU/IF0314/000ZZ/SUP6075958/N0>.

149 <https://it.didattica.unipd.it/off/2018/LM/SU/IF0314/000ZZ/SUP6076038/N0>.

la Communication et la Coopération internationales de l'Université de Padoue¹⁵⁰.

Le travail de compilation des fiches a été effectué lors d'expériences visant à étudier l'utilisation de la ressource TriMED d'un point de vue méthodologique et didactique comme support au processus de traduction spécialisée. En outre, de nombreuses expériences ont été menées pour l'accomplissement d'une tâche liée au domaine de la Recherche d'Information consistant à reformuler les requêtes médicales en adoptant une approche terminologique pour effectuer des revues systématiques. Pour la description détaillée des expériences menées jusqu'à présent et les spécifications relatives aux corpus textuel utilisés, au processus d'extraction terminologique et à la compilation de fiches multilingues, nous renvoyons le lecteur aux publications suivantes : Vezzani et Di Nunzio (2019b,a) ; Di Nunzio et Vezzani (2018). Dans cette section finale, nous nous limitons à l'aspect de l'affichage des données terminologiques pour les utilisateurs.

Le tableau 5.4 suivant rassemble des exemples de termes présents au sein du TriMED visibles par les utilisateurs et pour lesquels des fiches terminologiques multilingues ont été compilées. Le tableau présente les termes dans les trois langues, ordonnés selon le critère de classification des sphères conceptuelles dont la méthodologie de mise en œuvre a été décrite dans la sous-section 5.4.2 de ce chapitre.

150 <https://elearning.unipd.it/scienzeumane/course/view.php?id=4226>.

Tableau 5.4 – Exemples de termes dans TriMED classés par sphères conceptuelles

Sphère conceptuelle	ID Concept	Français	Anglais	Italien
A - Diagnostic	A2	diagnostic	diagnosis	diagnosi
A - Diagnostic	A18	test de détection rapide d'antigène	rapid antigen detection test	test rapido di rilevamento dell'antigene
A - Diagnostic	A23	imagerie	imaging	diagnostica per immagini
A - Diagnostic	A97	signe	signs	segno
A - Diagnostic	A98	symptôme	symptom	sintomo
B - Épidém. et prévention	B1	éradiquer	eradicate	debellare
B - Épidém. et prévention	B2	cocooning	cocoon strategy	strategia del bozzolo
B - Épidém. et prévention	B4	prévention	prevention	prevenzione
B - Épidém. et prévention	B5	patient zéro	patient zero	paziente zero
B - Épidém. et prévention	B6	autoévaluation	selfmonitoring	automonitoraggio
C - Pathologie	C2	cancer du col utérin	cervical cancer	tumore del collo dell'utero
C - Pathologie	C3	varice	varix	varice
C - Pathologie	C5	thrombose	thrombosis	trombosi
C - Pathologie	C6	typhoïde	typhoid	tifoide
C - Pathologie	C9	pneumonie communautaire	community-acquired pneumonia	polmonite acquisita in comunità
D - Spécialités médicales	D1	cytologie	cytology	citologia
D - Spécialités médicales	D3	virologie	virology	virologia
D - Spécialités médicales	D4	épidémiologie	epidemiology	epidemiologia
D - Spécialités médicales	D5	biologiste	biologist	biologo
D - Spécialités médicales	D7	bactériologie	bacteriology	batteriologia

Suite à la page suivante

Tableau 5.4 suite de la page précédente

Sphère conceptuelle	ID Concept	Français	Anglais	Italien
E - Pharmacologie	E2	florbetapir	florbetapir	florbetapir
E - Pharmacologie	E3	solution saline	saline	soluzione salina
E - Pharmacologie	E15	vaccin	vaccine	vaccino
E - Pharmacologie	E16	couverture vaccinale	vaccination coverage	copertura vaccinale
E - Pharmacologie	E40	médicament	drug	farmaco
G - Anatomie	G1	plaquette sanguine	platelet	piastrina
G - Anatomie	G2	rate	spleen	milza
G - Anatomie	G4	foie	liver	fegato
G - Anatomie	G5	veine porte	portal vein	vena porta
G - Anatomie	G11	colonne vertébrale	spine	colonna vertebrale
H - Thérapie	H1	réparation endoluminale	endoluminal repair	riparazione endoluminale
H - Thérapie	H2	résécabilité	resectability	resecabilità
H - Thérapie	H3	drainer	drain	drenare
H - Thérapie	H4	drain	drain	drenaggio
H - Thérapie	H5	résection pancréatique	pancreatic resection	resezione pancreatica

Le regroupement par sphères conceptuelles a été effectué *a posteriori* : une fois le corpus de travail sélectionné et la terminologie extraite, nous avons donc procédé à une classification par abstraction selon la nature sémantique des termes. Comme nous l'avons déjà évoqué, la classification par sphères conceptuelles et les termes qui peuplent chaque groupe ont été revus et validés par un médecin spécialiste. À cet égard, il est important de souligner que la classification est encore dans un état approximatif et doit être affinée. Parmi les difficultés rencontrées dans la formulation de cette cartographie conceptuelle, il est apparu nécessaire d'établir comme travaux futurs des « sphères – passerelles » intersécables qui permettent la présence simultanée du même terme dans plusieurs sphères conceptuelles : par exemple, le terme « dépistage » peut être classé à la fois comme appartenant au diagnostic (sphère A)

et à la prévention (sphère B). Au moment présent, nous nous sommes concentrés sur la définition des sphères conceptuelles délimitées, mais étendues (la catégorie concernant la pathologie, par exemple, comprend à la fois les agents pathogènes, tels que les virus et les bactéries, et les pathologies proprement dites) pour le regroupement de tous les termes sémantiquement proches et pour l'attribution d'identificateurs uniques au niveau conceptuel qui permettent la gestion automatique des désignations multilingues pour le même concept.

Quant à la nature morphologique des termes, la collection de TriMED est principalement composée de noms sous la forme de termes simples (un seul mot-forme) et de termes complexes (plusieurs mots-formes). L'application Web consultable en ligne contient également une modeste collection d'adjectifs analysés individuellement et identifiés dans la sphère conceptuelle provisoire « F – Qualificateurs », par exemple : « non invasif/ve », « congénital/e », « chronique », « virulent/e ». Compte tenu de leur nature, les adjectifs peuvent être classés en différentes sphères conceptuelles selon le nom auquel ils sont associés : on peut parler à la fois de « maladie non invasive » (sphère conceptuelle C – Pathologie) et d'« intervention non invasive » en référence à la pratique chirurgicale (sphère conceptuelle H – Thérapie). Par conséquent, nous opérons à rebours pour revenir au contexte d'utilisation du terme dans le corpus de travail afin de l'associer au nom qui permettra de positionner le terme composé ainsi formé dans la sphère conceptuelle de référence.

Enfin, en ce qui concerne l'état de révision, nous procédons à un nettoyage global de la forme et du contenu des fiches terminologiques visibles par les utilisateurs. Les normes ISO sur lesquelles repose notre proposition de modélisation conceptuelle ont été publiées après le début des travaux de collecte des termes et de compilation des fiches terminologiques. En ce sens, nous nous occupons actuellement d'harmoniser la forme de certaines catégories de données, en particulier celles pour lesquelles il existe une dénomination ou un format standard. Pour conclure, nous prévoyons, en fonction des ressources disponibles, une phase de validation effectuée par des experts du domaine en ce qui concerne les informations médicales fournies dans les fiches.

5.10 Conclusion

Ce long chapitre contient un parcours de recherche doctorale qui ouvre de nombreuses et fructueuses perspectives d'avenir. L'enquête visant à formuler le paradigme de la « terminologie FAIR » nous a conduits à développer une ressource numérique multilingue et polyvalente qui promet d'être vivante dans les années à venir grâce, notamment, à la synergie des travaux menés par tous les collaborateurs impliqués dans le projet.

À cette occasion, nous ne nous attardons pas sur les perspectives spécifiques de la ressource TriMED car sa description visait exclusivement à fournir au lecteur une étude de cas concrète sur l'application des principes de curation des données terminologique à l'ère numérique. D'un point de vue méthodologique, les orientations proposées peuvent en effet s'appliquer à toute autre ressource terminologique numérique conçue pour tout autre domaine de travail. Revenant donc à des aspects qui dépassent le cas particulier du TriMED, le prochain chapitre visera, en guise de conclusion, une réflexion critique sur les aspects de FAIRisation des données terminologiques numériques et jette les bases d'éventuelles implémentations futures à partir d'une structure de données conforme au format standard TBX.

6 *Is it FAIR enough?*

*Often when you think you're at the end of something,
you're at the beginning of something else.*

Fred Rogers (1928–2003)

L'étude de cas de la ressource numérique TriMED permet de développer des réflexions sur les évolutions possibles et prometteuses en termes de FAIRisation des données terminologiques. Pour résumer ce qui précède, le paradigme de la « terminologie FAIR » englobe un ensemble de lignes directrices pour la conception et l'implémentation d'une ressource qui :

1. suit un modèle structurel interopérable (TMF) ;
2. permet l'accès aux données terminologiques via des protocoles de communication standard ;
3. fournit des (méta)données rigoureusement documentées et trouvables, à travers un répertoire (de catégories) de données ;
4. garantit la réutilisation des données grâce à l'application de formats pour l'échange terminologique (TBX).

À ce stade, la première réflexion à mener porte sur l'état d'avancement actuel du processus de FAIRisation de la ressource TriMED illustrée ci-dessus. Comme nous l'avons vu à l'occasion de la description des principes FAIR et des données liées (sections 1.2.1, 1.2.2 et 3.1), l'attribution d'identifiants uniques et persistants (PID) pour chaque entité apparaissant au sein d'une collection est configurée comme une étape essentielle afin de désambiguïser leur interprétation. Or, alors que pour les catégories de données présentes dans TriMED ces identifiants ont été attribués de manière unique et globale (en utilisant les PID disponibles sur DatCatInfo ou en créant des nouveaux pour les catégories non encore référencées), pour les autres données contenues dans la ressource ce

processus est toujours en cours de développement. En particulier, comme nous l'avons évoqué dans le chapitre précédent, les entités-concept et les entités-terme de la fiche terminologique de TriMED sont associées à des identifiants alphanumériques uniques au sein de la ressource spécifique, mais non globaux par rapport à l'ensemble du Web sémantique. Dans la perspective d'une numérisation optimale de la ressource, ces codes seront ensuite transformés en identifiants globaux sous forme d'URI afin que les entités concernées puissent être nommées de manière unique sur le Web conformément aux principes des données liées (section 1.2.2). De cette façon, les données objet d'intérêt ne seront pas uniques seulement par rapport au monde limité de la ressource terminologique en question, mais aussi respectivement à toute autre entité disponible sur le Web sémantique. Par ailleurs, toujours dans la vision de l'interrelation et de l'interopérabilité entre les données, il est possible dans ce cas de se référer aux identifiants de ces entités déjà existants sur d'autres ressources (ontologies ou terminologies) actuellement publiées sur le Web.

Le deuxième aspect à examiner concerne le format de représentation des données. Comme nous l'avons vu, la norme ISO-30042 : 2019 promeut l'adoption du format TBX basé sur le métalangage XML afin de faciliter l'échange de données terminologiques. Cependant, il est important de souligner que l'idée d'échange promue ne fait pas nécessairement référence à des données présentes sur le Web. En vue de rendre disponible une collection de données terminologiques au sein du Web, il est donc nécessaire de réaliser une étape supplémentaire de transformation du format des données afin de se conformer aux standards du Web sémantique. En ce sens, nous avons mentionné au chapitre 1 que le modèle de données commun promu pour la modélisation de l'information sur le Web est le *Resource Description Framework* (RDF), en tant que standard d'échange et d'interrelation de données sur le Web. La transformation du format de représentation des données terminologiques de XML/TBX vers RDF est donc configurée comme un procédé avantageux à divers points de vue. Cette opération permet notamment de conformer la collection terminologique aux principes des données liées, tirant ainsi tous les bénéfices que cet environnement peut générer en termes de découverte, de conservation, d'interrelation et de réutilisation des données (Cimiano et al., 2020).

À ce stade, on peut se demander pourquoi ne pas structurer les données terminologiques directement au format RDF. Pour répondre

à cette question, nous citons un extrait de l'article de (Cimiano et al., 2015, p. 505) dans lequel est proposé un système de conversion de TBX vers RDF, que nous explorons dans les sections suivantes :

it is important to mention that we are not proposing to replace TBX by a new format. In fact, we regard our work as providing an alternative serialization of terminologies in RDF format. We assume that terminologies will be natively stored and managed using the TBX data model, but that in addition they will be exposed in RDF to support the linking of terminologies across datasets, thus supporting the creation of the above mentioned ecosystem.

En ce sens, nous soutenons l'idée que les données terminologiques en tant que telles sont nativement produites au format TBX comme standard de référence pour la gestion terminologique. La transformation en RDF est configurée plutôt comme une extension facultative et qui devient nécessaire si, à la base des besoins d'implémentation des ressources, il y a la volonté de soutenir l'écosystème des données liées et de bénéficier des avantages qui en découlent.

Dans cette perspective, ce chapitre vise à illustrer les possibilités qui se présentent au terminologue numérique une fois arrivé au format de représentation TBX. Dans les sections suivantes, nous décrivons donc le standard RDF et les études actuellement menées en termes de transformation des données terminologiques.

6.1 *Resource Description Framework* – RDF

Le modèle de données de base derrière le Web sémantique et les données liées est le *Resource Description Framework* (RDF). RDF est un standard développé par le *World Wide Web Consortium* (W3C) qui permet de décrire des ressources, principalement via des attributs et leurs relations avec d'autres ressources¹⁵¹. Ce modèle de données permet de structurer les informations à l'aide de déclarations constituées d'un sujet, d'un prédicat et d'un objet. L'idée de base est que les choses

151 <https://www.w3.org/RDF/>.

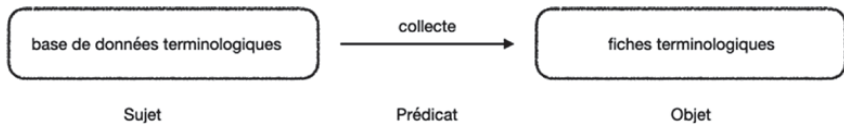


Image 6.1 – Exemple de graphe RDF

décrites ont des propriétés, qui ont des valeurs, et que les ressources peuvent être décrites par le biais de déclarations. RDF prescrit comment faire des déclarations sur les ressources, en particulier les ressources Web, sous la forme d’expressions sujet–prédicat–objet. Les déclarations sont également appelées triplets, car elles ont toujours la structure sujet–prédicat–objet.

Par exemple, dans la déclaration « une base de données terminologiques collecte des fiches terminologiques » :

- le sujet (« base de données terminologiques ») représente la ressource décrite ;
- le prédicat (« collecte ») indique le type de relation existant entre le premier et le dernier composant ;
- l’objet (« fiches terminologiques ») représente une donnée ou une autre ressource.

Chaque triple RDF constitue une déclaration, et, à son tour, ces déclarations peuvent être représentées comme un graphe, c’est-à-dire avec le sujet et l’objet étant les nœuds, et le prédicat reliant les deux (voir l’image 6.1). Enfin, il existe différents formats pour sérialiser les données RDF afin qu’elles puissent être publiées sur le Web. À titre d’exemple, nous listons les éléments suivants et renvoyons le lecteur à (Cimiano et al., 2020, p. 14–19) pour leur description détaillée : a) RDF/XML¹⁵², b) le langage N-Triples¹⁵³, c) Turtle¹⁵⁴, d) RDFa¹⁵⁵ et, enfin, le plus récent e) JSON-LD¹⁵⁶.

152 <https://www.w3.org/TR/rdf-syntax-grammar/>.

153 <https://www.w3.org/TR/n-triples/>.

154 <https://www.w3.org/TR/2014/REC-turtle-20140225/>.

155 <https://www.w3.org/TR/rdfa-primer/>.

156 <https://json-ld.org>.

6.2 De TBX à RDF

Comme nous l'avons mentionné précédemment, l'impact des données liées et les bénéfices découlant de cet écosystème ont conduit de nombreux chercheurs à développer des systèmes de transformation des données terminologiques du format XML/TBX au format RDF. Loin d'entrer dans les technicités de ces procédés qui dépasseraient le cadre de ce travail, nous regroupons ci-dessous quelques études menées depuis 2014 constituant l'état de l'art actuel sur le sujet¹⁵⁷. En outre, il faut noter que, pour des raisons strictement chronologiques, les études mentionnées ci-dessous se réfèrent soit au format TBX dans sa version obsolète de 2008 soit à sa version actuelle mise à jour de 2019.

Le premier effort visant à transformer le format des données terminologiques est représenté par l'étude de Reineke (2014). Dans cet article, l'auteur présente une routine de conversion presque automatisée de TBX (2008) à la sérialisation RDF/XML et vice versa. L'objectif était de mapper complètement le modèle de données TBX en RDF/XML et de reconvertir le document de fichier de sortie RDF/XML en TBX sans perte de données. Dans l'expérience décrite, la proposition de conversion s'est avérée efficace tant qu'une feuille de style est préparée pour mapper certains éléments, tels que les PID et le type de catégories de données. L'auteur lui-même prévoit comme développement futur de cette stratégie de transformation une approche entièrement automatique qui permet une transformation de toutes les instances TBX sans intervention de l'utilisateur. Une autre tentative importante de conversion TBX-RDF est représentée par l'étude Cimiano et al. (2015). Les auteurs proposent une approche de transformation basée sur le modèle *Ontolex-lemon* (McCrae et al., 2011), un modèle RDF développé pour soutenir la publication de ressources lexicales sous forme de données liées. La méthodologie de travail a conduit à la mise en place d'un service de conversion en ligne TBX2RDF¹⁵⁸ et à la publication d'une série de bonnes pratiques pour transformer les terminologies de

157 Pour une présentation plus approfondie des méthodologies de conversion, le lecteur est renvoyé au chapitre 9 de l'ouvrage de Cimiano et al. (2020).

158 <http://tbx2rdf.lider-project.eu/converter/>.

TBX au format de données liées¹⁵⁹. Comme dans l'étude précédente, le convertisseur et les lignes directrices se réfèrent à la version TBX de 2008. En outre, les auteurs fournissent une démonstration de faisabilité découlant de ce projet pilote en appliquant la méthodologie à deux bases de données terminologiques, à savoir l'*Inter-Active Terminology for Europe* (IATE), convertie de TBX en RDF¹⁶⁰, et le glossaire *European Migration Network* (EMN), converti à partir du format HTML vers RDF¹⁶¹. Cependant, une question qui mérite d'être explorée à propos de cette étude est soulevée dans la publication récente de Reineke et Romary (2019) dans laquelle les auteurs proposent un système de conversion impliquant le format TBX et le *Simple Knowledge Organisation System* (SKOS), une recommandation du W3C s'appuyant sur le modèle de données RDF¹⁶². Les auteurs soulignent qu'un problème avec la conversion de TBX en Ontolex-Lemon est que :

OntoLex is essentially a semasiological (word-to-sense) model which, by definition, cannot be naturally mapped onto the concept-to-term model of TBX. As a consequence, the mapping mainly consists of inserting TBX components into various unrelated places in the OntoLex format.

Pour cette raison, Reineke et Romary (2019) proposent plutôt une étude de conversion impliquant TBX (de 2019) et SKOS car tous les deux fournissent, dans une perspective onomasiologique, des modèles de données orientés vers le concept. Cependant, les auteurs soulignent que le problème majeur de mappage réside dans la différence entre SKOS et TBX en ce qui concerne la profondeur de la description de l'information et la flexibilité de l'instanciation. Par exemple, la catégorie de données / définition / est un élément qui dans SKOS est toujours positionné au niveau du concept, tandis que dans TBX les définitions peuvent être instanciées au niveau du concept ou au niveau de la langue. Par conséquent, lors de la mise en correspondance de TBX

159 <https://www.w3.org/2015/09/bpmlod-reports/multilingual-terminologies/>.

160 La version « données liées » de IATE est disponible sur : <http://tbx2rdf.lider-project.eu/data/iate>.

161 La version « données liées » du glossaire EMN est également disponible en ligne : <http://data.lider-project.eu/emn>

162 <https://www.w3.org/TR/skos-primer/>.

à SKOS, les définitions ne sont pas problématiques, puisque l’instanciation des informations relatives au concept dans SKOS est restreinte à un seul niveau. Dans le scénario de mappage inverse, cependant, la décision du niveau où placer la définition ne peut pas être résolue sans une désambiguïsation (humaine) préalable. En illustrant d’autres cas de conversion problématiques, les auteurs concluent que TBX est un format d’échange clair ayant une grande flexibilité d’instanciation, alors que SKOS fournit un formalisme de description relativement rigide et ambigu. Par conséquent, l’intégrité des données n’est pas affectée dans la conversion de SKOS à TBX mais un appauvrissement des données est généré lors du mappage de TBX à SKOS. Pour conclure ce panorama, nous évoquons enfin la récente étude de Di Buono et al. (2020) dans laquelle les auteurs présentent Terme-à-LLOD (TAL), un nouveau paradigme de virtualisation – composé de trois éléments – pour faciliter le processus de transformation des ressources terminologiques en RDF. En particulier, l’élément convertisseur gérant la transformation automatique est basé sur le service TBX2RDF susmentionné (Cimiano et al., 2015) qui permet de mapper les entrées TBX, y compris les dialectes publics TBX, c’est-à-dire TBX-Core, TBX-Min et TBX-Basic, au format RDF, en réutilisant un ensemble de classes et de propriétés à partir de vocabulaires de données liés existants (tels que OntoLex-Lemon). Le paradigme proposé dans cette étude s’inscrit dans le cadre plus large du projet européen Prêt-à-LLOD dont l’objectif est de fournir des services de TALN basés sur des données linguistiques liées afin qu’ils soient durables et puissent être facilement utilisés et déployés par des tiers¹⁶³.

6.3 Conclusion

Ce chapitre de conclusion avait pour but d’illustrer au lecteur certaines des possibilités auxquelles le terminologue numérique est confronté

163 Pour plus d’informations sur le projet, veuillez consulter le site Web suivant : <https://pret-a-llod.github.io>.

une fois qu'il arrive au format standard d'implémentation TBX. Pour répondre à la question posée au début de ce chapitre, le processus de FAIRisation des données terminologiques nécessite une mise à jour constante afin de tirer le meilleur parti des formats de représentation des données les plus efficaces en termes de trouvabilité, d'accessibilité, d'interopérabilité et de réutilisabilité. La question de la mesurabilité de la *FAIRness* des données reste un problème ouvert et l'objet de discussions actuelles par plusieurs groupes de travail de l'association *European Open Science Cloud* dont les efforts sont orientés vers la définition de critères d'évaluation efficaces¹⁶⁴.

Dans cette perspective, le chemin tracé jusqu'à présent du côté de la terminologie numérique n'est en aucun cas à considérer comme un point d'arrivée, mais plutôt comme un tremplin pour faire avancer des réflexions dynamiques futures sur l'organisation optimale des collections de données terminologiques.

164 https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf.

Annexe A

Tableau 6.1 – Spécifications des catégories de données en TriMED

Catégorie de données	Spécification	Description
Identifiant du concept	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-139
	Identifiant	conceptIdentifier
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Un code utilisé pour identifier une fiche terminologique au niveau du concept afin de relier des éléments physiques pour former une entrée de concept virtuelle.
Explication		
Note	L'hypothèse de l'ISO 12620 : 1999 était qu'une spécification du concept pouvait consister en plusieurs entrées de données réelles, d'où la répartition entre /identifiant du concept /et / identifiant/.	
	Exemple	
	Niveau (TMF)	conceptEntry
Concept superordonné	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-495
	Identifiant	superordinateConcept
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description	
	Description	Un concept qui est soit un concept générique soit un concept global.	
	Explication		
	Note		
	Exemple		
	Niveau (TMF)	conceptEntry	
Concept subordonné	Module	Trimed	
	PID	http://www.datcatinfo.net/datcat/DC-490	
	Identifiant	subordinateConcept	
	Classification	<descrip>	
	Typologie de contenu	chaîne de caractères	
	Valeur(s)		
	Description	Un concept dans un système hiérarchique qui peut être regroupé avec au moins un autre concept du même niveau pour former un concept de rang supérieur.	
	Explication		
	Note		
	Exemple		
	Niveau (TMF)	conceptEntry	
Relation entre concepts	Module	Trimed	
	PID	http://datcatinfo.termweb.se/datcat/DC-142	
	Identifiant	conceptRelation	
	Classification	<descrip>	
	Typologie de contenu	chaîne de caractères	
	Valeur(s)		
	Description	Un lien sémantique entre les concepts	
	Explication	Les relations conceptuelles constituent la base des systèmes conceptuels. Les types de relation conceptuelle peuvent inclure : relation générique, relation partitive, relation séquentielle, relation temporelle, relation spatiale, relation associative [liste ouverte].	

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Note	Dans ce champ, il convient d'exprimer la relation entre les concepts en indiquant le type de relation et en fournissant l'ID des concepts liés.
	Exemple Niveau (TMF)	conceptEntry
Sphère conceptuelle	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description	Trimed http://purl.org/trimed/dcr/dc/dc-9 conceptualSphere <descrip> liste déroulante A, B, C, D, E, G, H Par sphère conceptuelle, nous entendons une macroclassification qui regroupe des concepts et des termes sémantiquement liés.
	Explication	Pour la base de données TriMED, nous avons identifié 8 sphères conceptuelles différentes liées à différents champs sémantiques pour le domaine médical. Ce sont : A - Diagnostic ; B - Epidémiologie et prévention ; C - Pathologie; D- Spécialités médicales; E - Pharmacologie ; G - Anatomie ; H - Thérapie.
	Note Exemple Niveau (TMF)	conceptEntry
Identifiant de langue	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description	Trimed http://www.datcatinfo.net/datcat/DC-279 languageIdent <descrip> chaîne de caractères Un identifiant unique dans une fiche terminologique qui indique la langue décrite dans une section particulière.

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Explication Note Exemple Niveau (TMF)	EN, FR, IT langSec
Identifiant	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://www.datcatinfo.net/datcat/DC-3894 identifier <descrip> chaîne de caractères Indication d'un identifiant unique (id). Dans la fiche TriMED, cette catégorie de donnée fait référence *exclusivement* à la catégorie / terme/. termSec
Terme	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Core http://www.datcatinfo.net/datcat/DC-508 Term <term> chaîne de caractères Désignation d'un concept dans un domaine spécifique. Les termes peuvent être des unités lexicales simples ou complexes. terme simple : rougeole ; terme complexe : maladie d'Alzheimer termSec
Partie du discours	Module PID Identifiant Classification	Min http://www.datcatinfo.net/datcat/DC-396 partOfSpeech <termNote>

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Typologie de contenu Valeur(s)	liste déroulante nom, verbe, adjectif, adverbe, pronom, nom propre, autre
	Description	Une catégorie attribuée à un mot en fonction de ses propriétés grammaticales et sémantiques.
	Explication Note Exemple Niveau (TMF)	termSec
Genre grammatical	Module PID	Basic http://www.datcatinfo.net/datcat/ DC-245
	Identifiant Classification	grammaticalGender <termNote>
	Typologie de contenu Valeur(s)	liste déroulante masculin, féminin, neutre, autre genre
	Description	Une catégorie grammaticale qui indique les relations grammaticales entre les mots dans les phrases.
	Explication Note Exemple Niveau (TMF)	termSec
Nombre grammatical	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/ DC-251
	Identifiant Classification	grammaticalNumber <termNote>
	Typologie de contenu Valeur(s)	liste déroulante singulier, pluriel, collectif ou partitif
	Description	Dans de nombreuses langues, la distinction grammaticale qui indique le nombre d'objets auxquels fait référence le terme ou le mot.

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Explication	Dans les situations où le singulier et le pluriel d'un terme ne désignent pas nécessairement le même concept, le singulier et le pluriel doivent être signalés dans des entrées distinctes afin de conserver l'orientation conceptuelle de la base de données.
	Note	
	Exemple	
	Niveau (TMF)	termSec
Formes dérivées	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-4611
	Identifiant	derivative
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Un lexème qui est lié à un autre lexème par une règle de dérivation.
	Explication	
	Note	
	Exemple	virologique : forme dérivée de virologie
	Niveau (TMF)	termSec
Prononciation	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-407
	Identifiant	prononciation
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Représentation de la manière dont un terme ou un mot est prononcé.
	Explication	
	Note	
	Exemple	
	Niveau (TMF)	termSec

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
Etymologie	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-221
	Identifiant	etymology
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Informations sur l'origine d'un mot et l'évolution de sa signification.
	Explication	
	Note	
	Exemple	Vaccin : (1801) De vaccine, lui-même issu du latin vaccinus (« de vache »). Le virus ayant été historiquement recueilli dans les pustules du pis des vaches puis inoculé à l'être humain pour le préserver de la variole.
	Niveau (TMF)	termSec
Domaine	Module	Min
	PID	http://www.datcatinfo.net/datcat/DC-489
	Identifiant	subjectField
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Un champ de connaissances spécialisées.
	Explication	
	Note	
	Exemple	médecine, lois, économie etc.
	Niveau (TMF)	conceptEntry
Sous-domaine	Module	Trimed
	PID	http://purl.org/trimed/dcr/dc/dc-7
	Identifiant	subDomain
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Description	Un sous-domaine de connaissances spécialisées.
	Explication	
	Note	
	Exemple	virologie, pathologie, chirurgie
	Niveau (TMF)	conceptEntry
Définition	Module	Basic
	PID	http://www.datcatinfo.net/datcat/DC-168
	Identifiant	definition
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Une représentation d'un concept par un énoncé descriptif.
	Explication	
	Note	
	Exemple	
	Niveau (TMF)	langSec
Analyse sémique	Module	Trimed
	PID	http://purl.org/trimed/dcr/dc/dc-1
	Identifiant	semicAnalysis
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	L'analyse sémique d'une production sémiotique, un texte par exemple, vise à en dégager les sèmes, c'est-à-dire les éléments de sens, à définir leurs regroupements (isotopies et molécules) et à stipuler les relations entre ces regroupements (relations de présupposition, de comparaison, etc., entre isotopies).
	Explication	
	Note	

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Exemple	Pathologie : /étude/ /médecine/ / développement/ /cause / symptôme/ /maladie/. Maladie : /altération/ /santé/ /être vivant/ /signe/ / symptôme/ /causes internes/ / causes externes/ /évolution/ / organisme/.
	Niveau (TMF)	langSec
Contexte	Module PID	Basic http://www.datcatinfo.net/datcat/DC-149
	Identifiant	context
	Classification	<descrip>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Un texte qui illustre un concept ou l'utilisation d'une désignation dans son environnement naturel.
	Explication	
	Note	
	Exemple	
	Niveau (TMF)	termSec
Variante orthographique	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-330
	Identifiant	variant
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Un terme, une appellation ou un lexème dans sa présentation orthographique alternative.
	Explication	
	Note	
	Exemple	acupuncture vs acuponcture
	Niveau (TMF)	termSec

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
Abréviation	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-1418
	Identifiant	abbreviation
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Une forme abrégée résultant de l'omission de mots ou de lettres.
Expansion	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-321
	Identifiant	fullForm
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Terme, appellation ou lexème dans sa présentation complète (lorsqu'il existe une forme abrégée).
Acronyme	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-334
	Identifiant	acronym
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Forme abrégée résultant de la combinaison des lettres ou syllabes initiales (de chacun ou de certains des éléments) de la forme complète et prononcée syllabiquement comme un mot.

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Explication Note Exemple Niveau (TMF)	CA : canal artériel termSec
Synonyme	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://www.datcatinfo.net/datcat/DC-317 synonym <termNote> chaîne de caractères Tout terme ou lexème représentant le même ou un concept très semblable à celui de l'entrée principale. La synonymie est généralement relative, c'est-à-dire que les synonymes couvrent rarement tous les aspects du même concept dans tous les cas. Le degré de synonymie résultant est traité en utilisant les conventions définies pour le degré d'équivalence. affection vs maladie termSec
Hyponyme	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note	Trimed http://purl.org/trimed/dcr/dc/dc-5 Hyponym <termNote> chaîne de caractères Un mot de signification plus spécifique qu'un terme général ou supérieur qui lui est applicable.

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Exemple Niveau (TMF)	fruit vs pomme termSec
Hyperonyme	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://purl.org/trimed/dcr/dc/dc-6 hypernym <termNote> chaîne de caractères L'hyperonyme est le terme qui désigne le générique. fruit vs pomme termSec
Unité phraséologique	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication	Trimed http://www.datcatinfo.net/datcat/DC-339 phraseologicalUnit <termNote> chaîne de caractères Tout groupe de deux ou plus mots qui sont fréquemment exprimés ensemble et qui se composent de plus d'un concept, selon lequel les mots individuels fonctionnent généralement dans plus d'une catégorie grammaticale (partie du discours dans la syntaxe d'une phrase) [Définition fournie par le module TBX Basic]. Bien qu'elles soient composées de plus d'un mot et qu'elles contiennent souvent plus d'un concept, les unités phraséologiques peuvent être traitées comme des unités terminologiques individuelles dans les bases de données.

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Note Exemple Niveau (TMF)	termSec
Collocation	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description	Trimed http://www.datcatinfo.net/datcat/DC-340 Collocation <termNote> chaîne de caractères Combinaison récurrente de mots caractérisée par la cohésion en ce que les composants de la collocation doivent coexister dans un énoncé ou une série d'énoncés, même s'ils ne doivent pas nécessairement maintenir une proximité immédiate les uns avec les autres.
	Explication	Les collocations diffèrent des unités phraséologiques en ce que les composants de ces dernières doivent généralement apparaître dans une séquence fixe. Les combinaisons de mots récurrentes qui forment un terme complexe (par exemple, adjectif + nom, nom + nom, etc.) et qui représentent un concept unique ne sont pas des collocations.
	Note Exemple	(Virus) attraper, contracter, inoculer, porter un ; se protéger contre les . Un se développe, se répand, se propage.
	Niveau (TMF)	termSec

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
Registre	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-423
	Identifiant	Register
	Classification	<termNote>
	Typologie de contenu	liste déroulante
	Valeur(s)	registre familier, registre technique, argot, vulgaire.
	Description	Classification indiquant le niveau de langue individuellement assigné à un terme.
Explication		
Note		
Exemple		
Niveau (TMF)	termSec	
Nom populaire	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-328
	Identifiant	commonName
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Synonyme d'un terme scientifique international utilisé dans le discours général dans une langue donnée.
Explication		
Note		
Exemple	première maladie vs rougeole	
Niveau (TMF)	termSec	
Nom scientifique	Module	Trimed
	PID	http://www.datcatinfo.net/datcat/DC-3754
	Identifiant	scientificName
	Classification	<termNote>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
Description	Terme technique utilisé par les experts du domaine.	
Explication		

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Note Exemple Niveau (TMF)	première maladie vs rougeole termSec
Code ICPC2	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://purl.org/trimed/dcr/dc/dc-2 icpc2Code <xref> chaîne de caractères URL du terme correspondant dans la ressource ICP2. termSec
Code ICD10	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://purl.org/trimed/dcr/dc/dc-3 icd10Code <xref> chaîne de caractères URL du terme correspondant dans la ressource ICD10. termSec
Terme SNOMED CT	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note	Trimed http://purl.org/trimed/dcr/dc/dc-4 Snomed <xref> chaîne de caractères URL du terme correspondant dans la ressource SNOMED.

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Exemple Niveau (TMF)	termSec
Terme MeSh	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Trimed http://purl.org/trimed/dcr/dc/dc-8 Mesh <xref> chaîne de caractères URL du terme correspondant dans la ressource MESH. termSec
Source	Module PID Identifiant Classification Typologie de contenu Valeur(s) Description Explication Note Exemple Niveau (TMF)	Basic http://www.datcatinfo.net/datcat/DC-471 Source <descrip> chaîne de caractères Une citation complète des informations bibliographiques relatives à un document ou à une autre ressource. La catégorie de donnée /source/ fait référence à la /définition/ au niveau de la section de langue et au /contexte/ au niveau de la section de terme. langSec, termSec
Référence croisée	Module PID Identifiant Classification Typologie de contenu Valeur(s)	Basic http://www.datcatinfo.net/datcat/DC-164 CrossReference <ref> chaîne de caractères

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Description	Champ utilisé pour diriger l'utilisateur vers une autre information associée ; par exemple, une autre fiche terminologique.
	Explication Note	La catégorie de donnée /référence croisée/ doit être placée à la fois au niveau du concept et au niveau du terme : cela permettra d'indiquer les associations entre différents concepts ou termes dans la base de données.
	Exemple Niveau (TMF)	conceptEntry, termSec
Référence croisée externe	Module	Basic
	PID	http://www.datcatinfo.net/datcat/DC-226
	Identifiant	externalCrossReference
	Classification	<xref>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	URL ou autre URI vers une ressource externe à la base de données ou au fichier d'échange.
	Explication Note	La catégorie de donnée /référence croisée externe/ fait référence à la / définition/ et au /context/.
	Exemple Niveau (TMF)	langSec, termSec
Responsabilité	Module	Basic
	PID	http://www.datcatinfo.net/datcat/DC-451
	Identifiant	Responsibility
	Classification	<transacGrp>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	

Suite à la page suivante

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Description	Un identifiant attribué à l'individu associé à la transaction concernant la gestion de la base de données.
	Explication Note	La catégorie de donnée / responsabilité/ devrait être placée à la fois au niveau du concept et au niveau du terme : cela permettra de se référer à la personne qui a formulé la fiche terminologique (au niveau du concept) et à la personne qui a compilé la section du terme.
	Exemple Niveau (TMF)	conceptEntry, termSec
Type de transaction	Module PID	Basic http://www.datcatinfo.net/datcat/DC-1689
	Identifiant Classification Typologie de contenu Valeur(s) Description	Transaction <transacGrp> liste déroulante origine, modification L'une des étapes de la formulation, de l'approbation et de l'utilisation d'une entrée terminologique.
	Explication Note	Dans la fiche TriMED, la catégorie de donnée /type de transaction/ doit être placée à la fois au niveau du concept et au niveau du terme : cela permet d'identifier le type de transaction (origine ou modification) effectuée par l'utilisateur.
	Exemple Niveau (TMF)	conceptEntry, termSec
Date	Module PID	Core http://www.datcatinfo.net/datcat/DC-4335
	Identifiant Classification Typologie de contenu Valeur(s)	date <date> chaîne de caractères

Tableau 6.1 suite de la page précédente

Catégorie de données	Spécification	Description
	Description	Moment auquel une transaction ou un événement a lieu.
	Explication	Selon ISO 8601, la disposition AAAAMM- JJ, avec possibilité d'extension à la date et à l'heure, par exemple AAAA-MM-JJ hh : mm : ss, doit être utilisée pour représenter les dates.
	Note	Dans la fiche de Tri- MED, il devrait y avoir une date se référant au concept et une date se référant au terme : c'est-à-dire qu'il faut déclarer la date de création de la fiche terminologique au niveau du concept et la date d'entrée des données dans la section consacrée au terme.
	Exemple Niveau (TMF)	2020-02-18 conceptEntry, termSec
Note	Module	Core
	PID	http://www.datcatinfo.net/datcat/DC-382
	Identifiant	Note
	Classification	<note>
	Typologie de contenu	chaîne de caractères
	Valeur(s)	
	Description	Une déclaration qui fournit des informations supplémentaires sur toute partie d'une entrée de la ressource terminologique.
	Explication	
	Note	Dans la fiche de Tri- MED, la catégorie de donnée /note/ doit être placée à chaque niveau du métamodèle TMF (concept, langue, terme). Dans la section langue, la /note/ fait référence aux commentaires sur la définition.
	Exemple Niveau (TMF)	conceptEntry, langSec, termSec

Annexe B

Data Category Module Definition

Name: Trimed

Namespace: http://shiny.dei.unipd.it/TriMED/data_category_repository/

Description: The Trimed Module is a data category module collecting data categories selected from 1) DatCatInfo (<http://www.datcatinfo.net/#/>) and 2) TriMED Data Category Repository

Data Categories:

Image 6.2 – Formalisme complet du module *Trimed*

Name	PID	Classification	Value(s)	Level (TMF)
superordinateConcept	http://www.datcatinfo.net/datcat/DC-495	<descrip>		conceptEntry
subordinateConcept	http://www.datcatinfo.net/datcat/DC-490	<descrip>		conceptEntry
subDomain	http://purl.org/trimed/dcr/dc/dc-7	<descrip>		conceptEntry
conceptIdentifier	http://www.datcatinfo.net/datcat/DC-139	<descrip>		conceptEntry
languageIdent	http://www.datcatinfo.net/datcat/DC-279	<descrip>		langSec
identifier	http://www.datcatinfo.net/datcat/DC-3894	<descrip>		termSec

Suite à la page suivante

Image 6.2 suite de la page précédente

Name	PID	Classification	Value(s)	Level (TMF)
grammaticalNumber	http://www.datcatinfo.net/datcat/DC-251	<termNote>	singular, plural, dual, mass, otherNumber	termSec
derivative	http://www.datcatinfo.net/datcat/DC-4611	<termNote>		termSec
pronunciation	http://www.datcatinfo.net/datcat/DC-407	<termNote>		termSec
etymology	http://www.datcatinfo.net/datcat/DC-221	<termNote>		termSec
variant	http://www.datcatinfo.net/datcat/DC-330	<termNote>		termSec
abbreviation	http://www.datcatinfo.net/datcat/DC-1418	<termNote>		termSec
fullForm	http://www.datcatinfo.net/datcat/DC-321	<termNote>		termSec
acronym	http://www.datcatinfo.net/datcat/DC-334	<termNote>		termSec
semicAnalysis	http://purl.org/trimed/dcr/dc/dc-1	<descrip>		langSec
synonym	http://www.datcatinfo.net/datcat/DC-317	<termNote>		termSec
hyponym	http://purl.org/trimed/dcr/dc/dc-5	<termNote>		termSec
hypernym	http://purl.org/trimed/dcr/dc/dc-6	<termNote>		termSec
phraseologicalUnit	http://www.datcatinfo.net/datcat/DC-339	<termNote>		termSec

Image 6.2 suite de la page précédente

Name	PID	Classification	Value(s)	Level (TMF)
collocation	http://www.datcatinfo.net/ datcat/DC-340	<termNote>		termSec
register	http://www.datcatinfo.net/ datcat/DC-423	<termNote>	colloquialRegister, neutralRegister, technicalRegister, in-houseRegister, benchlevelRegister, slangRegister, vulgarRegister	termSec
commonName	http://www.datcatinfo.net/ datcat/DC-328	<termNote>		termSec
scientificName	http://www.datcatinfo.net/ datcat/DC-3754	<termNote>		termSec
icpc2Code	http://purl.org/trimed/dcr/dc/ dc-2	<xref>		termSec
icd10Code	http://purl.org/trimed/dcr/dc/ dc-3	<xref>		termSec
snomed	http://purl.org/trimed/dcr/dc/ dc-4	<xref>		termSec
mesh	http://purl.org/trimed/dcr/dc/ dc-8	<xref>		termSec
conceptRelation	http://datcatinfo.termweb.se/ datcat/DC-142	<descrip>		conceptEntry
conceptualSphere	http://purl.org/trimed/dcr/dc/ dc-9	<descrip>	A, B, C, D, E, F, G, H	conceptEntry

Bibliographie

- Alarcón-Navío, E., C. López-Rodríguez, et M. Tercedor-Sánchez (2016). Variation dénomminative et familiarité en tant que source d'incertitude en traduction médicale. *Meta : journal des traducteurs/Meta: Translators' Journal* 61 (1), 117–144.
- Amjadian, E., D. Inkpen, T. S. Paribakht, et F. Faez (2018). Distributed specificity for automatic terminology extraction. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 24 (1), 23–40.
- Atzeni, P., S. Ceri, S. Paraboschi, et R. Torlone (1999). *Database systems: concepts, languages & architectures*. London: McGraw-Hill.
- Auger, P. (1989). La terminotique et les industries de la langue. *Meta : journal des traducteurs/Meta: Translators' Journal* 34 (3), 450–456.
- Auger, P. et M.-C. L'Homme (1994). La terminologie selon une approche textuelle : une représentation plus adéquate du lexique dans les langues de spécialité. In *ALFA : Actes de Langue Française et de Linguistique/Symposium on French Language and Linguistics*, 7, Halifax, NS, Canada, 1994, 17–21. Dept. of French, Dalhousie University.
- Balliu, C. (2005). La didactique de la traduction médicale, deux ou trois choses que je sais d'elle. *Meta : journal des traducteurs/Meta: Translators' Journal* 50 (1), 67–77.
- Balliu, C. (2010). Le traducteur, le médecin et le patient. *Meta : journal des traducteurs/Meta: Translators' Journal* 55 (1), 15–22.
- Bengio, Y., R. Ducharme, P. Vincent, et C. Jauvin (2003). A neural probabilistic language model. *Journal of machine learning research* 3, 1137–1155.
- Bergenholtz, H. et U. Kaufmann (1997). Terminography and lexicography. A critical survey of dictionaries from a single specialised

- field. *HERMES-Journal of Language and Communication in Business* (18), 91–125.
- Berners-Lee, T. (2006). *Linked data-design issues*. Date d'accès 24/02/2022 : <https://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, T., J. Hendler, et O. Lassila (2001). The semantic web. *Scientific american* 284 (5), 34–43.
- Berra, A. (2012). Faire des humanités numériques. In Mounier, P. (ed.), *Read/Write Book 2 : Une introduction aux humanités numériques*, Marseille : OpenEdition Press, 25–43.
- Bianco, A. (2002). *Lingue speciali e traduzione specializzata*. Lecce : Adriatica salentina.
- Bizer, C., T. Heath, et T. Berners-Lee (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems* 5 (3), 1–22.
- Boulanger, J.-C. (1995). Présentation : images et parcours de la socio-terminologie. *Meta : journal des traducteurs/Meta: Translators' Journal* 40 (2), 194–205.
- Bourigault, D., C. Jacquemin, et M.-C. L'Homme (2001). (edd.), *Recent advances in computational terminology*, Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Bourigault, D. et M. Slodzian (1999). Pour une terminologie textuelle. *Terminologies nouvelles* 19, 29–32.
- Boutin-Quesnel, R. (1985). *Vocabulaire systématique de la terminologie*. Québec : Publications du Québec.
- Bowker, L. et D. Fisher (2010). Computer-aided translation. In Gambier, Y., L. Van Doorslaer (edd.), *Handbook of translation studies*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 60–65.
- Bowker, L. et S. Hawkins (2006). Variation in the organization of medical terms: Exploring some motivations for term choice. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12 (1), 79–110.
- Broeder, D., I. Schuurman, et M. Windhouwer (2014). Experiences with the ISOcat data category registry. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, 2014, 4565–4568. European Language Resources Association (ELRA).

- Cabré Castellví, M. T. (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona : Editorial Antártida/Empúries.
- Cabré Castellví, M. T. (1999a). *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona : Universidad Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Cabré Castellví, M. T. (1999b). *Terminology: Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Cabré Castellví, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 9 (2), 163–199.
- Calderan, L. et J. Millet (2015). *BIG DATA : nouvelles partitions de l'information : Actes du séminaire IST Inria, octobre 2014*. Louvain-la-Neuve : De Boeck Supérieur.
- Campo, Á. (2013). *The reception of Eugen Wüster's work and the development of terminology*. Ph.D. thesis, Université de Montréal (Faculté des arts et des sciences), Montréal.
- Cardillo, E., L. Serafini, et A. Tamilin (2009). A lexical-ontological resource for consumer healthcare. In Kostkova, P. (ed.), *Electronic Healthcare: Second International ICST Conference, eHealth 2009, Istanbul, Turkey, September 23-15, 2009, Revised Selected Papers*, Berlin, Heidelberg : Springer, 131–138.
- Carvalho, S., R. Costa, et C. Roche (2018). The role of conceptual relations in the drafting of natural language definitions: an example from the biomedical domain. In *Proceedings of the LREC 2018 Workshop "Globalex 2018 – Lexicography & WordNets"*, 10–16. Miyazaki: Globalex.
- Carvalho, S. M. P., R. Costa, et C. Roche (2016). Ontoterminology meets lexicography. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, 2016, 8–15. European Language Resources Association (ELRA).
- Castro, C. M., C. Wilson, F. Wang, et D. Schillinger (2007). Babel babble: physicians' use of unclarified medical jargon with patients. *American journal of health behavior* 31 (1), S85–S95.
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021).

- Shiny: Web Application Framework for R. R package version 1.7.1.*
<https://CRAN.R-project.org/package=shiny>.
- Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transaction on Database Systems* 1 (1), 9–36.
- Chiarcos, C., S. Nordhoff, et S. Hellmann (2012). *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*. Berlin/Heidelberg: Springer.
- Church, K., W. Gale, P. Hanks, et D. Hindle (1991). Using statistics in lexical analysis. In Zernik U. (ed.), *Lexical acquisition: exploiting on-line resources to build a lexicon*, New Jersey: Lawrence Erlbaum, 115-164.
- Church, K. W. et P. Hanks (1990). Word association norms, mutual information, and lexicography. *Computational linguistics* 16 (1), 22– 29.
- Cimiano, P., C. Chiarcos, J. P. McCrae, et J. Gracia (2020). *Linguistic Linked Data: Representation, Generation and Applications*. Cham, Switzerland : Springer.
- Cimiano, P., J. P. McCrae, V. Rodríguez-Doncel, T. Gornostay, A. Gómez-Pérez, B. Siemoneit, et A. Lagzdins (2015). Linked terminologies: applying linked data principles to terminological resources. In *Proceedings of the eLex 2015 Conference*, Sussex, UK, 2015, 504–517.
- Clarke, S. G. D. (2001). Thesaural relationships. In Bean, C.A. et R., Green (edd.), *Relationships in the organization of knowledge. Information Science and Knowledge Management*, Dordrecht : Springer, 37–52.
- Condamines, A. (2005). Linguistique de corpus et terminologie. *Langages* (1), 36–47.
- Condamines, A. (2007). L'interprétation en sémantique de corpus : le cas de la construction de terminologies. *Revue française de linguistique appliquée* 12 (1), 39–52.
- Cortelazzo, M. (1990). *Lingue speciali: la dimensione verticale*. Padova : Unipress.
- Corti, L., V. Van den Eynden, L. Bishop, et M. Woollard (2019). *Managing and sharing research data: a guide to good practice*. Los Angeles : SAGE Publications Limited.
- Coseriu, E. (1969). *Einführung in die strukturelle Linguistik*. Tübingen : Gunter Narr Verlag.

- Costa, H., A. Zaretskaya, G. Corpas Pastor, et M. Seghiri Domínguez (2016). Nine terminology extraction tools: Are they useful for translators? *Multilingual* (27) 3, 14–20.
- Costa, R. (2006). Plurality of theoretical approaches to terminology. In Pitch, H. (ed.), *Modern approaches to terminological theories and applications*, Bern: Peter Lang Publishing, 79–89.
- Costa, R. (2013). Terminology and specialised lexicography: two complementary domains. *Lexicographica* 29 (1), 29–42.
- Costa, R. et R. Silva (2008). De la typologie à l'ontologie de textes. In *Terminologie & Ontologies : Théories et Applications. Actes de la deuxième conférence-TOTh Annecy – 2008*, 3–16, Annecy : Institut Porphyre – Savoir et Connaissance.
- Costa, R., R. Silva, S. Barros, et A. L. Soares (2012). Mediation strategies between terminologists and experts. In *Proceedings of GLAT 2012-Terminologies : textes, discours et accès aux savoirs spécialisés*, Genoa, Italy, 2012, 297–308.
- Da Silva Santos, L., M. D. Wilkinson, A. Kuzniar, R. Kaliyaperumal, M. Thompson, M. Dumontier, et K. Burger (2016). FAIR data points supporting big data interoperability. In Zelm M., G. Doumeingts, et J. P. Mendonça (edd.), *Enterprise Interoperability in the Digitized and Networked Factory of the Future*. London: ISTE Press, 270–279.
- De Bessé, B. (2000). Le domaine. In Béjoint H., et P. Thoiron (edd), *Le sens en terminologie*, Lyon : Presses Universitaires Lyon, 182–197.
- De Jong, F., B. Maegaard, K. De Smedt, D. Fišer, et D. Van Uytvanck (2018). CLARIN: towards FAIR and responsible data science using language resources. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018, 3259–3264. European Language Resources Association (ELRA).
- Delavigne, V. (2017). Term usage and socioterminological variation. In Drouin, P., A. Francoeur, J. Humbley, et A. Picton (edd.), *Multiple Perspectives on Terminological Variation*. Volume 18, Amsterdam/Philadelphia: John Benjamins Publishing Company, 31–56.
- Depecker, L. (2017). *Entre signe et concept : Éléments de terminologie générale*. Paris : Presses Sorbonne Nouvelle.
- Di Buono, M. P., P. Cimiano, M. F. Elahi, et F. Grimm (2020). Terminallod: Simplifying the conversion and hosting of terminological

- resources as linked data. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, 2020, 28–35.
- Di Nunzio, G. M. et F. Vezzani (2018). Using R markdown for replicable experiments in evidence-based medicine. In Bellot, P., C. Trabelsi, J. Mothe, F. Murtagh, J. Nie, L. Soulier, E. SanJuan, L. Cappellato, et N. Ferro (edd.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction – 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10–14, 2018, Proceedings*, Volume 11018, Cham: Springer, 8–39.
- Di Nunzio, G. M. et F. Vezzani (2021). On the reusability of terminological data. In *Proceedings of the 10th AIUCD Conference*, 2021, 183-186. Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD).
- Diki-Kidiri, M. (2000). Une approche culturelle de la terminologie. *Terminologies nouvelles* (21), 27–31.
- Diki-Kidiri, M. (2007). Éléments de terminologie culturelle. *Cahiers du RIFAL* (26), 14–25.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 9 (1), 99–115.
- Drouin, P., A. Francoeur, J. Humbley, et A. Picton (2017). (edd.), *Multiple perspectives on terminological variation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Drouin, P., N. Grabar, T. Hamon, et K. Kageura (2015). Introduction to the special issue: Terminology across languages and domains. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 21 (2), 139–150.
- Drouin, P., N. Grabar, T. Hamon, K. Kageura, et K. Takeuchi (2018). Computational terminology and filtering of terminological information: Introduction to the special issue. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 24 (1), 1–6.
- Dubuc, R. (2002). *Manuel pratique de terminologie*. Montréal (Québec) : Linguatech.
- Dugas, A. (1994). Variation orthographique. *Equivalences* 24 (2), 153–170.

- Durieux, C. (1996). Pseudo-synonymes en langue de spécialité. *Cahier du CIEL* 97, 89–114.
- Elhadad, N. et K. Sutaria (2007). Mining a lexicon of technical terms and lay equivalents. In *ACL'07 workshop on biological, translational, and clinical language processing (BioNLP'07)*, Prague, Czech Republic, 49–56.
- Faber, P. (2015). Frames as a framework for terminology. In Kockaert, J. H., et F. Steurs (edd.), *Handbook of terminology*, Volume 1, Amsterdam/Philadelphie: John Benjamins Publishing Company, 14–33.
- Faber, P. (2002). Oncoterm: sistema bilingüe de información y recursos oncológicos. In Alcina, A. (ed.), *La traducción científico-técnica y la terminología en la sociedad de la información*, Castelló de la Plana : Publicacions de la Universitat Jaume, 177-188.
- Faber, P. (2012). (ed.), *A cognitive linguistics view of terminology and specialized language*. Berlin/Boston: De Gruyter Mouton.
- Felber, H. (1980). International standardization of terminology: Theoretical and methodological aspects. *International Journal of the Sociology of Language* 1980 (23), 65–80.
- Felber, H. (1981). The Vienna school of terminology fundamentals and its theory. In Krommer-Benz, M. (ed.), *Theoretical and Methodological Problems of Terminology. Proceedings of the International Symposium*, Volume 6, Infoterm Series, München-New York-London-Paris: K-G Saur, 69-86.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. In Harnad, S. R., H. D. Steklis, et J. Lancaster (edd.), *Origins and Evolution of Language and Speech*, Volume 280, New York: Annals of the NY Academy of Sciences, 20-32.
- Fischbach, H. (1962). Problems of medical translation. *Bulletin of the Medical Library Association* 50 (3), 462.
- Foo, J. (2012). *Computational terminology: Exploring bilingual and monolingual term extraction*. Ph.D. thesis, University Electronic Press, Linköping.
- Frantzi, K. T., S. Ananiadou, et J. Tsujii (1998). The c-value/nc-value method of automatic recognition for multi-word terms. In *ECDL'98 Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries*, London, UK: Springer-Verlag, 1998, 585–604.

- Freixa, J. (2006). Causes of denominative variation in terminology: A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12 (1), 51–77.
- Gábor, K., D. Buscaldi, A.-K. Schumann, B. QasemiZadeh, H. Zargayouna, et T. Charnois (2018). Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, 679–688. Association for Computational Linguistics.
- Galinski, C. (1982). Ten years of infoterm—activities and achievements. *Journal of Information Science* 5 (2–3), 103–114.
- Galinski, C. (1990). Terminology and phraseology. *Terminology Science and Research: Journal of the International Institute for Terminological Research* 1 (1–2), 70–86.
- Gaudin, F. (1993). *Socioterminologie : des problèmes sémantiques aux pratiques institutionnelles*. Mont-Saint-Aignan : Publ. de l'Université de Rouen.
- Gaudin, F. (2003). *Socioterminologie. Une approche sociolinguistique de la terminologie*. Bruxelles : Duculot.
- Gefen, A. (2015). Les enjeux épistémologiques des humanités numériques. *Socio. La nouvelle revue des sciences sociales* (4), 61–74.
- Gile, D. (1987). La terminotique en interprétation de conférence : un potentiel à exploiter. *Meta : journal des traducteurs/Meta: Translators' Journal* 32 (2), 164–169.
- Gotti, M. (1991). *I linguaggi specialistici: caratteristiche linguistiche e criteri pragmatici*. Scandicci : La Nuova Italia.
- Gouadec, D. (1987). Les horizons de la terminotique. *Meta : journal des traducteurs/Meta : Translators' Journal* 32 (2), 130–138.
- Gouadec, D. (1990). *Terminologie : constitution des données*. Paris : AFNOR.
- Hadden, K., C. Coleman, et A. Scott (2018). The bilingual physician: Seamless switching from medicalese to plain language. *Journal of Graduate Medical Education* 10 (2), 130–133.
- Hébert, L. (2007). *Dispositifs pour l'analyse des textes et des images*. Limoges : Presses de l'Université de Limoges.

- Hendler, J., F. Gandon, et D. Allemang (2020). *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*. New York: Association for Computing Machinery.
- Humbley, J. (2004). La réception de l'œuvre d'eugen wüster dans les pays de langue française. *Cahiers du CIEL 2004–2005*, 33–51.
- Humbley, J. (2014). Écoles de terminologie : mythe ou réalité ? In Martel, P., et J. Maurais (edd.), *Mélanges Jean-Claude CORBEIL*, Volume 8, Niemeyer Verlag, 159–167.
- Humbley, J. (2016). Catherine Resche (dir.), terminologie et domaines spécialisés, approches plurielles. Paris : Classiques Garnier, rencontres 143, série linguistique 2, 2015. *ASp. la revue du GERAS* (70), 127–132.
- Inkpen, D., T. S. Paribakht, F. Faez, et E. Amjadian (2016). Term evaluator: A tool for terminology annotation and evaluation. *International Journal of Computational Linguistics and Applications* 7 (2), 145–165.
- ISO 1087-1 (2000, October). Terminology work – Vocabulary – Part 1: Theory and application. Standard, International Organization for Standardization, Geneva, CH.
- ISO 1087-1 (2019). Terminology work and Terminology Science-Vocabulary – Part 1: Theory and application. Standard, International Organization for Standardization, Geneva, CH.
- ISO-12620 (2019, March). Management of terminology resources – Data category specifications. Standard, International Organization for Standardization, Geneva, CH.
- ISO-16642 (2017, November). Computer applications in terminology – Terminological markup framework. Standard, International Organization for Standardization, Geneva, CH.
- ISO-30042 (2019, April). Management of terminology resources—Term-Base eXchange (TBX). Standard, International Organization for Standardization, Geneva, CH.
- Jaffré, J.-P. (2010). De la variation en orthographe. *Ela. Études de linguistique appliquée* (3), 309–323.
- Jammal, A. (1999). Une méthodologie de la traduction médicale. *Meta : journal des traducteurs/Meta: Translators' Journal* 44 (2), 217–237.
- Jezeq, E. (2005). *Lessico. Classi di parole, strutture, combinazioni*. Bologna : Il Mulino.

- Jucks, R. et R. Bromme (2007). Choice of words in doctor-patient communication: an analysis of health-related internet sites. *Health Communication* 21 (3), 267–277.
- Kageura, K., B. Daille, H. Nakagawa, et L.-F. Chien (2004). Introduction: Recent trends in computational terminology. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 10 (1), 1–21.
- Kageura, K. et B. Umino (1996). Methods of automatic term recognition: A review. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 3 (2), 259–289.
- Kemps-Snijders, M., M. Windhouwer, P. Wittenburg, et S. E. Wright (2008). ISOcat: Corraling data categories in the wild. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 2008. European Language Resources Association (ELRA).
- Khayari, M., S. Schneider, I. Kramer, L. Romary (2006). Unification of multi-lingual scientific terminological resources using the ISO 16642 standard. The TermSciences initiative. In *International Workshop Acquiring and representing multilingual, specialized lexicons: the case of biomedicine*. Date d'accès 25/02/2022 : <https://hal.archives-ouvertes.fr/hal-00022424>.
- Lee, K. et L. Romary (2010). Towards interoperability of iso standards for language resource management. In *Proceedings of Language Resources and Interoperability, The Second International Conference on Global Interoperability for Language Resources (ICGL201)*, Hong Kong, 95–104.
- Lee-Jahnke, H. (2001). L'enseignement de la traduction médicale : un double défi ? *Meta : journal des traducteurs/Meta: Translators' Journal* 46 (1), 145–153.
- Lefever, E., L. Macken, et V. Hoste (2009). Language-independent bilingual terminology extraction from a multilingual parallel corpus. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 496–504.
- Lerat, P. (1995). *Les langues spécialisées*. Paris : Presses Universitaires de France.
- Lethuillier, J. (1989). La synonymie en langue de spécialité. *Meta : journal des traducteurs/Meta: Translators' Journal* 34 (3), 443–449.

- L'Homme, M. (2020). *Lexical Semantics for Terminology: An Introduction*. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- L'Homme, M.-C. (2004). *La terminologie : principes et techniques*. Montreal : Les Presses de l'Université de Montreal.
- L'Homme, M.-C. (2006). A look at some canadian contributions to terminology. In Picht H. (ed.), *Modern approaches to terminological theories and applications*, Bern : Peter Lang, 55–75.
- L'Homme, M.-C. et A. San Martín (2016). Définition terminologique : systématisation de règles de rédaction dans les domaines de l'informatique et de l'environnement. *Cahiers de lexicologie* 109 (2), 147–174.
- Llanos, L. C., D. Bouamor, P. Zweigenbaum, et S. Rosset (2016). Managing linguistic and terminological variation in a medical dialogue system. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, 2016, 3167–3173. European Language Resources Association (ELRA).
- Loginova Clouet, E. (2014). *Traitement automatique des termes composés : segmentation, traduction et variation*. Ph.D. thesis, Université de Nantes, France.
- Lommel, A., A. Melby, N. Glenn, J. Hayes, et T. Snow (2014). TBX-Min: a simplified tbx-based approach to representing bilingual glossaries. In *Terminology and Knowledge Engineering 2014*. Date d'accès 25/02/2022 : <https://hal.archives-ouvertes.fr/hal-01005851>.
- Lurquin, G. (1986). La synonymie en terminologie médicale (1^{ère} partie). *Le Langage et l'homme* (60), 18–26.
- Macken, L., E. Lefever, et V. Hoste (2013). Taxis: Bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 19 (1), 1–30.
- Magris, M. (2012). Variation in terminologie, terminographie und phraseographie. *JoSTrans-The Journal of Specialised Translation* 18, 160–174.
- Magris, M. (2016). Verso un uso consapevole della terminologia nell'informazione medica rivolta al paziente. In Soliman, L. (ed.) *Lingue speciali: dinamiche di ricerca*. Padova : CLUEP.

- Marzá, N. E. (2009). *The specialised lexicographical approach: A step further in dictionary-making*. Bern: Peter Lang.
- McCrae, J., D. Spohr, et P. Cimiano (2011). Linking lexical resources and ontologies on the semantic web with lemon. In *Extended Semantic Web Conference*, Berlin, Heidelberg: Springer, 245–259.
- McCrae, J. P., C. Chiarcos, F. Bond, P. Cimiano, T. Declerck, G. De Melo, J. Gracia, S. Hellmann, B. Klimek, S. Moran, et al. (2016). The open linguistics working group: Developing the linguistic linked open data cloud. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, 2016, 2435–2441. European Language Resources Association (ELRA).
- McCray, A. T. (2005). Promoting health literacy. *Journal of the American Medical Informatics Association* 12 (2), 152–163.
- McLure, M., A. V. Level, C. L. Cranston, B. Oehlerts, et M. Culbertson (2014). Data curation: a study of researcher practices and needs. *Portal: Libraries and the Academy* 14 (2), 139–164.
- Melby, A. K. (1995). Implementing the terminology interchange format. In *Standardizing and Harmonizing Terminology: Theory and Practice*, Philadelphia: ASTM International, 187–199.
- Melby, A. K. (2015). TBX: A terminology exchange format for the translation and localization industry. In Kockaert, J. H., et F. Steurs (edd.), *Handbook of terminology*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 393–424.
- Melby, A. K., K.-D. Schmitz, et S. E. Wright (2001). Terminology interchange. In Wright, S. E., et G. Budin (edd.) *Handbook of Terminology Management*, Volume 2, Amsterdam/Philadelphia: John Benjamins Publishing Company, 613–642.
- Meunier, J.-G. (2014). Humanités numériques ou computationnelles : Enjeux herméneutiques. *Sens public* 12. Date d'accès 25/02/2022: <http://sens-public.org/spip.php?article1121>.
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, et J. Dean (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, London: MIT Press, 3111–3119.
- Monin, S. (1993). La siglaison en langue médicale et problèmes de traduction. *ASp. la revue du GERAS* (2), 29–53.

- Mons, B., C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, et M. D. Wilkinson (2017). Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud. *Information Services & Use* 37 (1), 49–56.
- Montalt, V. (2011). Medical translation and interpreting. In Gambier, Y., et L.V. Doorslaer (edd), *Handbook of translation studies*, Volume 2, Amsterdam/Philadelphia: John Benjamins Publishing Company, 79–83.
- Montalt, V. et M. Gonzalez-Davies (2014). *Medical translation step by step: Learning by drafting*. London : Routledge.
- Nakagawa, H. (2001). Experimental evaluation of ranking and selection methods in term extraction. In J. C. Bourigault D, et M.-C. L’Homme (edd.), *Recent advances in computational terminology*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 303–326.
- Oliver, A. et M. Vázquez (2015). Tbxtools: a free, fast and flexible tool for automatic terminology extraction. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, 473–479.
- Palmer, C., N. M. Weber, A. H. Renear, et T. Muñoz (2013). Foundations of data curation: The pedagogy and practice of “purposeful work” with research data. *Archives Journal* 3. Date d’accès 25/02/2022: <http://www.archivejournal.net/essays/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/>
- Patel, V., T. Branch, et J. Arocha (2002). Errors in interpreting quantities as procedures: The case of pharmaceutical labels. *International journal of medical informatics* 3 (65), 193–211.
- Pavel, S., D. Nolet, et C. Leonhardt (2001). *Précis de terminologie/ The Handbook of Terminology*. Adapted into English by C. L. Hull: Public Works and Government Services of Canada.
- Picht, H. (1982). Reflexions on training in terminology in the 1980s. In *Terminologies for the eighties*, Volume 7 of Infoterm Series, Munich/New York/London/Paris : K G Saur, 172–185.
- Picton, A. et P. Dury (2017). Diastatic variation in language for specific purposes. In Drouin, P., A. Francoeur, J. Humbley, et A. Picton (edd.), *Multiple Perspectives on Terminological Variation*, Volume 18, Amsterdam/Philadelphia: John Benjamins Publishing Company, 57–80.

- Planas, E. (2012). BiTermEx un prototype d'extraction de mots composés à partir de documents comparables via la méthode compositionnelle, In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2 : TALN*, 415–422.
- Polguère, A. (2015). Non-compositionnalité : ce sont toujours les locutions faibles qui trinquent. *Verbum (Presses Universitaires de Nancy)* 37 (2), 257–280.
- Quérin, S. (2013). Variation terminologique en français médical. *Panace* 14 (38), 235–239.
- Raffo, M. (2016). Translation and popularization: Medical research in the communicative continuum. *Meta : journal des traducteurs/ Meta: Translators' Journal* 61, 163–175.
- Rastier, F. (2009). *Sémantique interprétative*. Paris : Presses universitaires de France.
- Reineke, D. (2014). TBX between termbases and ontologies. In *Proceedings of the 11th International Conference on Terminology and Knowledge Engineering, TKE 2014*. Date d'accès 26/02/2022: <http://hal.archives-ouvertes.fr/hal-01005838>.
- Reineke, D. et L. Romary (2019). Bridging the gap between SKOS and TBX. *Edition-Die Fachzeitschrift für Terminologie* 19 (2). Date d'accès 25/02/2022: <https://hal.inria.fr/hal-02398820>.
- Roche, C. (2012). Ontoterminology: How to unify terminology and ontology into a single paradigm. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012, 2626–2630. European Language Resources Association (ELRA).
- Roche, C. (2015). Ontological definition. In Kockaert, J. H., et F. Steurs (edd.), *Handbook of terminology*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 128–152.
- Roche, C., A. Alcina, et R. Costa (2019). Terminological resources in the digital age. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 25 (2), 139–145.
- Roche, C., M. Calberg-Challot, L. Damas, et P. Rouard (2009). Ontoterminology: A new paradigm for terminology. In *International Conference on Knowledge Engineering and Ontology Development*, 321–326.
- Roche, C., R. Costa, S. Carvalho, et B. Almeida (2019). Knowledge-based terminological e-dictionaries: The endoterm and al-andalus

- pottery projects. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 25 (2), 259–290.
- Romary, L. (2001a). An abstract model for the representation of multilingual terminological data: TMF-Terminological Markup Framework. In *Proceedings of the 5th TermNet Symposium, TAMA 2001*, Antwerp, Belgium. Date d'accès : 25/02/2022 <http://hal.inria.fr/inria-00100405>.
- Romary, L. (2001b). Un modèle abstrait pour la représentation de terminologies multilingues informatisées. *Cahiers GUTenberg* (39–40), 81–88.
- Romary, L., I. Kramer, S. Salmon-Alt, et J. Roumier (2006). Gestion de données terminologiques : principes, modèles, méthodes. In *Terminologie et accès à l'information*. Paris : Hermes Science Publisher.
- Rouleau, M. (1995). La langue médicale : une langue de spécialité à emprunter le temps d'une traduction. *TTR : traduction, terminologie, rédaction* 8 (2), 29–49.
- Rouleau, M. (2003). La terminologie médicale et ses problèmes. *Panace* 4 (12), 143–152.
- Rousseau, L.-J. (1993). Terminologie et phraséologie, deux composantes indissociables des langues de spécialités. *Terminologies nouvelles – Phraséologie Actes du séminaire international* (10), 9–11.
- Roy-Hubara, N. et A. Sturm (2020). Design methods for the new database era: a systematic literature review. *Software and Systems Modeling* 19 (2), 297–312.
- Sager, J. C. (1990). *Practical course in terminology processing*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Salton, G. et C.-S. Yang (1973). On the specification of term values in automatic indexing. *Journal of documentation* 29 (4), 351–372.
- Tercedor Sánchez, M. et J. A. Prieto Velasco (2013). Las barreras en la comunicación médico-paciente: el proyecto VARIMED. In Martínez López, A.B, I. Jiménez Gutiérrez, et M. I. Martínez Robledo (edd.), *Traducir la Cultura: De barreras culturales en la traducción científica y técnica*, Granada: Comares, 593–606.

- Sandoval, A. M., J. Díaz, L. C. Llanos, et T. Redondo (2019). Biomedical term extraction: Nlp techniques in computational medicine. *IJIMAI* 5 (4), 51–59.
- Santos, C. et R. Costa (2015). Domain specificity. In Kockaert, J. H., et F. Steurs (edd.), *Handbook of terminology*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 153–179.
- Schmitz, K.-D. (2012). Using international standards for terminology exchange. *Terminologija* 19 (2012), 33–38.
- Sclano, F. et P. Velardi (2007). TermExtractor: A Web application to learn the common terminology of interest groups and research communities. In *Proceedings of the Seventh Terminology and Artificial Intelligence Conference (Sophia Antipolis, France, Oct. 8–9), 2007*, 85–94. Presses Universitaires de Grenoble.
- Serianni, L. (2003). *Italiani scritti*. Bologna : Il mulino.
- Serianni, L. (2005). *Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente*. Milano : Garzanti Libri.
- Silva, R. (2014). *Gestão de terminologia pela qualidade: processos de validação*. Ph.D. thesis, Universidade Nova de Lisboa, Portugal.
- Simon, N. I. et V. Kešelj (2018). Automatic term extraction in technical domain using part-of-speech and common-word features. In *Proceedings of the ACM Symposium on Document Engineering 2018*, Halifax, NS, Canada, 2018, 1–4. Association for Computing Machinery.
- Sparck-Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation* 28 (5), 111–121.
- Studer, R., V. R. Benjamins, et D. Fensel (1998). Knowledge engineering: Principles and methods. *Data & knowledge engineering* 25 (1–2), 161–197.
- Temmerman, R. (1997). The process of (neo-) lexicalization: the case of the life sciences. In *Terminologie et interdisciplinarité. Actes du colloque organisé en avril 1996 par le Centre de terminologie de Bruxelles (Institut libre Marie Haps) et l'Association européenne de professeurs de langues vivantes (AEPLV)*, Louvain-la-Neuve, Belgium, 1997, 47–62. Peeters Publishers.
- Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive-approach*, Amsterdam/Philadelphia : John Benjamins Publishing Company.

- Temmerman, R. et K. Kerremans (2003). Termonotography: Ontology building and the sociocognitive approach to terminology description. In *Proceedings of the International Congress of Linguistics (CIL17)*, Prague, Czech Republic, 2003, 1–10. Matfyzpress.
- Tran, T., H. Chekroud, P. Thiery, et A. Julienne (2009). Internet et soins : un tiers invisible dans la relation médecine/patient. *Health Communication* 3 (21), 267–77.
- Van Hoof, H. (1986). Les éponymes médicaux : essai de classification. *Meta : journal des traducteurs/Meta: Translators' Journal* 31 (1), 59–84.
- Van Hoof, H. (1993). *Dictionnaire des éponymes médicaux : français–anglais*. Leuven : Peeters Publishers.
- Van Hoof, H. (2001). La traduction des éponymes médicaux banalisés de langue anglaise. *Meta : journal des traducteurs/Meta: Translators' Journal* 46 (1), 82–91.
- Vandaele, S. (2001). Noyaux conceptuels et traduction médicale. *Meta : journal des traducteurs/Meta: Translators' Journal* 46 (1), 16–21.
- Vandaele, S. et M. Gingras Harvey (2013). Les nomenclatures anatomiques : histoire et traduction. *Panacea* 14 (38), 222–34.
- Vecchiato, Sara et Gerolimich, S. (2013). Le langage médical est-il ‘trop complexe’ ? *Nouvelles perspectives en sciences sociales* 9 (1), 81–122.
- Vezzani, F. (2021). La ressource FAIRterm : entre pratique pédagogique et professionnalisation en traduction spécialisée. *Synergies Italie* (17), 51–64.
- Vezzani, F. et G. M. Di Nunzio (2019a). Computational terminology in eHealth. In Manghi, P., L. Candela, et G. Silvello (edd.), *Digital Libraries: Supporting Open Science - 15th Italian Research Conference on Digital Libraries, IRCDL 2019, Pisa, Italy, January 31 – February 1, 2019, Proceedings*, Volume 988, Cham: Springer, 72–85.
- Vezzani, F. et G. M. Di Nunzio (2019b). On the use of terminological records in specialised translation. In *Proceedings of the 8th AIUCD Conference*, Udine, Italy, 2019, 151–156. Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD).
- Vezzani, F. et G. M. Di Nunzio (2020a). Methodology for the standardization of terminological resources: design of TriMED database

- to support multi-register medical communication. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 26 (2), 266–298.
- Vezzani, F. et G. M. Di Nunzio (2020b). On the Formal Standardization of Terminology Resources: The Case Study of TriMED. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, 2020, 4903–4910. European Language Resources Association (ELRA).
- Vezzani, F., G. M. Di Nunzio, et G. Henrot (2018). TriMED: A Multilingual Terminological Database. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018, 4367–4371. European Language Resources Association (ELRA).
- Vu, T., A. Aw, et M. Zhang (2008). Term extraction through unithood and termhood unification. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume II*, Hyderabad, India, 2008, 631–636. The Association for Computer Linguistics.
- Wakabayashi, J. (1996). Teaching medical translation. *Meta : Journal des traducteurs/Meta: Translators' Journal* 41 (3), 356–365.
- Wang, L., Y. Zhang, M. Jiang, J. Wang, J. Dong, Y. Liu, C. Tao, G. Jiang, Y. Zhou, et H. Xu (2018). Toward a normalized clinical drug knowledge base in china—applying the rxnorm model to chinese clinical drugs. *Journal of the American Medical Informatics Association* 25 (7), 809–818.
- Warburton, K. (2015). Managing terminology in commercial environments. In Kockaert, J. H., et F. Steurs (edd.), *Handbook of terminology*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 359–391.
- Warburton, K. (2021). *The Corporate Terminologist*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data* 3 (1), 1-9.
- Windhouwer, M. et I. Schuurman (2014). Linguistic resources and cats: how to use ISOcat, RELcat and SCHEMAcat. In *Proceedings of the 9th International Conference on Language Resources*

- and Evaluation (LREC'14)*, Reykjavik, Iceland, 2014, 3806–3810. European Language Resources Association (ELRA).
- Wright, S. E. (2001). Data categories for terminology management. In Wright, S. E., et G. Budin (edd.) *Handbook of Terminology Management*, Volume 2, Amsterdam/Philadelphia: John Benjamins Publishing Company, 552–571.
- Yadav, V. et S. Bethard (2018). 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, New Mexico, USA, 2018, 2145–2158. Association for Computational Linguistics.
- Zanola, M. T. (2018). *Che cos' è la terminologia*. Roma : Carocci editore.
- Zeng, Q. T. et T. Tse (2006). Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association* 13 (1), 24–29.

Index

A

accessibilité 18, 34, 36, 41, 68,
82, 151, 166
actualisation 20, 83, 84, 93–
95, 98
affichage 94, 144, 145, 148, 149,
152–154
analyse des exigences 20, 61,
99, 108
analyse sémique 115, 116, 127,
128, 131, 138, 139, 147,
152, 174
approche complémentaire 54
approche lexico-sémantique 51
approche sociocognitive 49, 52
association 19, 28, 39, 64, 70–
75, 77, 82, 96, 111, 112, 118,
121, 124, 166, 183

B

besoins d'information 20, 61,
109
biunivocité 46

C

C/NC value 30
candidats termes 30, 50, 57
caractéristique distinctive 56
caractéristique essentielle 56
caractéristiques
formelles 115, 147

cardinalités 111, 112
catégories de données 13, 15,
62, 75–82, 86–90, 93, 111,
112, 114, 115, 119, 121, 123–
130, 138–141, 144, 146, 151,
152, 157, 163, 167
changement de terminologie 105
cohérence 74, 78, 81, 86, 94,
95, 128
collection de données
terminologiques 76, 77, 97,
111, 113, 128, 160
collocation 115, 117, 118, 131,
133, 135, 136, 147, 148,
179, 189
compréhensibilité 101
computationnel 14, 19, 23, 24,
26–30, 32, 40, 42
concept 19, 20, 24, 25, 28, 30–
32, 39, 46–50, 52, 52–58,
61, 70–74, 76, 77, 79, 82,
90, 93, 100, 106, 111, 113,
114, 117–124, 126, 130–
132, 137–140, 147, 150,
155–157, 164, 165, 167–170,
172–175, 177–179, 183–185,
187, 189
concept subordonné 120, 121,
130, 138, 147, 168
concept superordonné 120, 138,
139, 147, 167

- conception 13, 15, 17, 20, 25, 26,
 33, 38, 43, 46, 47, 58, 59, 61,
 62, 68–70, 73–75, 81–83,
 88, 95, 98, 99, 109, 110, 111,
 141, 142, 151, 159
 concision 100
 connaissances spécialisées 17,
 25, 30, 54, 56, 61, 100,
 173, 174
 construit 14, 15, 37, 50, 118
 contexte 14, 17, 27–30, 33, 35,
 37, 40, 47, 51, 52, 58, 63, 64,
 66, 67, 84, 99, 101, 102, 107,
 108, 115, 118, 124, 130, 140,
 146–149, 157, 175, 182
 corpus électroniques 50, 51
 curation de données 35
- D**
 décodage 104, 109
 définition 19, 26, 33, 43, 45, 52,
 74, 78, 80, 81, 89, 90, 92,
 103, 108, 109, 113, 115–121,
 123–125, 127, 129, 130,
 138–140, 146, 147, 157,
 164–166, 174, 178, 182,
 183, 185
 Data Category as
 Attribute 87, 138
 Data Category as Tag 87, 138
 DatCatInfo 75, 78–81, 89, 125,
 127, 159, 167–180, 182–185,
 187–189
 dialecte TBX 88, 89, 92, 129–
 131, 137
 dialectes privés 88
 dialectes publics 87, 129, 165
- dimension conceptuelle 14, 52,
 54–56, 58, 69
 dimension linguistique 52, 57
 domaine 17, 18, 20, 23–33, 35,
 37, 46, 47, 49, 52–58, 61, 62,
 69, 70, 84, 94, 97–99, 101–
 104, 106–109, 115, 119, 121,
 127, 130, 138–140, 147, 153,
 154, 157, 158, 169, 170, 173,
 174, 180
 données 13–15, 17–21, 24–27,
 29, 30, 33–43, 45, 50, 56,
 58, 59, 61–70, 72–90, 93–
 99, 103–106, 107, 109–115,
 117, 119–131, 138–141, 144,
 146, 148, 149, 151–154,
 157–185
 données liées 38–40, 43, 159–
 161, 163, 164
 données linguistiques liées
 ouvertes 40, 41
 données ouvertes liées 39, 40
- E**
 entité 15, 19, 20, 36, 39, 48, 53,
 54, 58, 70–74, 77, 79, 82,
 111, 112, 122, 139, 140, 159,
 160
 entité-concept 70–72, 79
 entité-langue 71, 72
 entité-terme 70–72, 79
 entrée terminologique 74, 76,
 77, 93, 107, 111, 113, 120,
 123, 124, 139, 140, 184
 éponymie 105, 106
 extraction automatique de
 termes 27, 30

F

FAIRisation 15, 20, 63, 96, 158, 159, 166
 fiche terminologique 14, 20, 93, 94, 107, 108, 110, 113, 114, 117–121, 123, 125, 137, 146, 148, 152, 160, 167, 169, 183–185
 formalisme TBXMD 90, 130
 format de représentation 15, 82, 84, 160, 161
 formats d'échange 66, 85, 86
 fouille de textes 30
 frame-based Terminology 51, 103

G

gestion des données 27, 65

H

humanités numériques 23
 hyperonyme 115, 117, 121, 127, 131, 147, 178
 hyponyme 115, 117, 121, 127, 131, 147, 177

I

implémentation 15, 18, 20, 25, 26, 33, 38, 49, 52, 58, 61, 64, 69, 74, 75, 82–84, 98, 99, 102, 111, 121, 125, 128, 151, 158, 159, 161, 166
 infrastructures 36, 97
 ingénierie des connaissances 52, 55
 insertion 20, 83, 94, 95

intelligence

artificielle 28, 29, 52
 interaction 87, 129, 141, 145, 196
 interopérabilité 16, 37, 64, 66, 68, 74, 75, 78, 81, 82, 84, 86, 110, 125, 149, 151, 160, 166
 ISO TC/37 SC 3 62, 68

L

langue médicale 100.101, 104, 106, 116
 langues de spécialité 25, 49
 Lexical Markup Framework 74
 lexicographie 25, 29, 46, 51, 55
 lexicologie 25
 linguistique computationnelle 26, 28, 40
 linguistique de corpus 29
 lisibilité 101
 locution 117, 118

M

métadonnées 17, 19, 33, 36, 37, 62–65, 67, 97, 127, 151, 152
 métamodèle structurel 62, 68, 73–75
 médicalese 101
 modélisation conceptuelle 18, 19, 26, 33, 43, 67, 69, 73, 82, 109, 110, 148, 157
 modification 20, 83, 94, 95, 124, 144, 184
 module TBX 89, 129, 178
 mono-référentialité 100
 Mutual Information 30

N

- neutralité émotionnelle 100
- normalisation 46, 48, 108, 110, 125, 140
- numérique 13–15, 17–20, 23, 24, 26, 32–39, 41–43, 45, 46, 54, 55, 58, 59, 61, 62, 68–70, 74, 81–84, 89, 95, 98, 99, 103, 108–110, 120, 121, 140, 158–161, 165, 166
- numérisation 14, 17, 19, 23, 24, 26, 42, 160

O

- onomasiologique 46, 52, 53, 55, 164
- ontologie 33, 47, 52, 66, 149, 160
- ontoterminologie 19, 52, 53

P

- pertinence 31, 51
- PID 36, 80, 82, 90, 126–128, 130, 159, 163, 167–185, 187
- poids 30
- polysémie 46, 48, 49, 105, 118
- précision 100
- principes FAIR 13, 14, 19, 35, 37–39, 43, 62, 65, 67, 68, 151, 159
- propriétés 15, 19, 20, 51, 56, 68–71, 74, 78, 82, 89, 90, 112, 114, 162, 165, 171
- publication 15, 20, 29, 38, 45, 48, 51, 57, 63, 82–86, 96–98, 110, 140, 143, 154, 163, 164

R

- répertoire de catégories de données 75, 79, 81, 125, 128
 - répertoire de données 75, 78, 97, 98
 - réseaux lexicaux 57, 58, 61
 - réutilisabilité 18, 36, 37, 84, 166
 - recherche d'information 29, 30, 145, 154
 - registre 80, 102, 104, 107, 113, 115, 119, 131, 140, 147, 180
 - Relax NG 94
 - représentation de la connaissance *in vitro* 48
 - représentation de la connaissance *in vivo* 48
 - représentation des connaissances 17, 37
 - Resource Description Framework 39, 160, 161
 - ressources lexicographiques 25
 - ressources terminologiques 13, 17, 24, 25, 29, 30, 33, 43, 49, 54, 55, 62, 64, 65, 68, 74, 75, 79, 81, 84, 85, 106–108, 125, 165
- S
- sémantique cognitive 49
 - sémantique des cadres 51
 - sémasiologique 46, 51, 52, 55
 - schéma conceptuel 56, 61
 - schéma entité-association 19, 70, 72, 77, 111, 112
 - science de la terminologie 13, 24

- science ouverte 13, 14, 37,
 62, 64, 97
 section de langue 77, 111, 182
 section de terme 77, 78, 90, 111,
 126, 138, 140, 182
 socioterminologie 19, 47, 48
 spécifications de catégories de
 données 78, 80
 spécificité 14, 15, 20, 31, 84
 sphères conceptuelles 121, 153–
 157, 169
 suppression 20, 83, 94
 synonymie 46, 48, 49, 52, 100,
 105, 106, 117, 140, 177
 système conceptuel 53, 56,
 119, 120
- T**
 T-Score 30
 TBX-Basic 88, 129, 137, 165
 TBX-Core 88, 129, 137, 165
 TBX-Min 88, 137, 165
 techniques d'apprentissage
 profond 32
 Term frequency-Inverse
 Document Frequency 30
 TermBase eXchange 13, 15, 20,
 68, 82, 84, 86, 128
 terme 14, 15, 17, 19, 20, 24–28,
 30–32, 34, 36, 38, 39, 46,
 48–55, 57, 58, 61, 69–73,
 77–79, 81, 82, 90, 93, 100–
 124, 126–131, 135, 137,
 138–141, 145, 146, 148–150,
 153–157, 159–161, 165, 166,
 169–172, 175–185
- termes complexes 77, 93, 157
 termes simples 157
 termhood 31
 termicité 31
 terminographe 25, 33, 50,
 67, 68, 99
 terminographie 14, 19, 24, 25,
 27, 42, 54
 Terminological Markup
 Framework 68, 73, 75
 terminologie 13–15, 17–19,
 23–30, 32–35, 37, 38, 42, 43,
 45–55, 58, 59, 61–64, 67–70,
 73, 74, 81, 82, 84–88, 96,
 99–101, 103–105, 107–110,
 113, 114, 119, 125, 129, 141,
 142, 144, 148, 149, 151, 156,
 158–161, 163, 164, 166
 terminologie accessible 64
 terminologie
 computationnelle 14, 19,
 24, 27–30, 32, 42
 terminologie culturelle 19,
 47, 49, 50
 terminologie FAIR 14, 15, 18,
 19, 38, 62, 68, 69, 74, 81, 82,
 84, 96, 98, 99, 109, 110, 141,
 148, 158, 159
 terminologie interopérable 66
 terminologie médicale 99, 100,
 105, 108, 109
 terminologie numérique 13–15,
 18, 19, 24, 26, 32–34, 43, 45,
 68, 70, 166
 terminologie réutilisable 67
 terminologie textuelle 19, 47, 50

- terminologie trouvable 63
 terminologie 13, 17, 18, 20, 25–27, 32–35, 43, 45, 46, 48, 50, 54–58, 62, 68, 72, 74, 80, 81, 83–85, 95, 99, 106, 141, 161, 165
 terminologie numérique 18, 20, 26, 34, 35, 43, 45, 54, 55, 58, 62, 74, 83, 84, 95, 161, 165
 terminotique 14, 19, 24, 26, 27, 42
 termitude 31
 termontographie 52
 textes spécialisés 27, 57
 théorie communicative de la terminologie 19, 47, 48
 théorie des prototypes 49
 théorie générale de la terminologie 19, 46, 47, 52
 Traduction Assistée par Ordinateur 110, 129, 148, 153
 traduction automatique 30, 31
 traduction médicale 104, 110
 traitement automatique de la langue 27
 transcodage 104, 105, 109
 transparence 36, 100
 travail terminologique 13–15, 19, 24–27, 43, 45, 54, 58, 61, 68
 travail terminologique numérique 24, 54, 68
 trouvabilité 18, 36, 63, 64, 67, 68, 81, 82, 151, 166
- U
 unité terminologique 48
 unités de compréhension 49
 unités phraséologiques 178, 179
 usage 14, 47, 88, 104, 105, 107, 110, 119, 146, 148
 utilisateurs 20, 51, 61–64, 67, 88, 89, 95, 97, 99, 109, 114, 128, 141, 143–145, 153, 154, 157
- V
 variation diastratique 102, 104, 109, 114, 119, 141, 145
 variation terminologique 49, 101–103, 106, 107, 141, 146
- W
 Web de documents 38
 Web de données 37, 38
 Web sémantique 14, 21, 33, 34, 38, 42, 43, 160, 161

Linguistic Insights

Studies in Language and Communication

This series aims to promote specialist language studies in the fields of linguistic theory and applied linguistics, by publishing volumes that focus on specific aspects of language use in one or several languages and provide valuable insights into language and communication research. A cross-disciplinary approach is favoured and most European languages are accepted.

The series includes two types of books:

- Monographs – featuring in-depth studies on special aspects of language theory, language analysis or language teaching.
- Collected papers – assembling papers from workshops, conferences or symposia.

Each volume of the series is subjected to a double peer-reviewing process.

- Vol. 1 Maurizio Gotti & Marina Dossena (eds)
Modality in Specialized Texts. Selected Papers of the 1st CERLIS Conference.
421 pages. 2001. ISBN 3-906767-10-8 · US-ISBN 0-8204-5340-4
- Vol. 2 Giuseppina Cortese & Philip Riley (eds)
Domain-specific English. Textual Practices across Communities
and Classrooms.
420 pages. 2002. ISBN 3-906768-98-8 · US-ISBN 0-8204-5884-6
- Vol. 3 Maurizio Gotti, Dorothee Heller & Marina Dossena (eds)
Conflict and Negotiation in Specialized Texts. Selected Papers
of the 2nd CERLIS Conference.
470 pages. 2002. ISBN 3-906769-12-7 · US-ISBN 0-8204-5887-2
- Vol. 4 Maurizio Gotti, Marina Dossena, Richard Dury, Roberta Facchinetti & Maria Lima
Variation in Central Modals. A Repertoire of Forms and Types of Usage
in Middle English and Early Modern English.
364 pages. 2002. ISBN 3-906769-84-4 · US-ISBN 0-8204-5898-8

Editorial address:

Prof. Maurizio Gotti Università di Bergamo, Dipartimento di Lingue, Letterature e Culture
Straniere Piazza Rosate 2, 24129 Bergamo, Italy
Fax: +39 035 2052789, E-Mail: m.gotti@unibg.it

- Vol. 5 Stefania Nuccorini (ed.)
Phrases and Phraseology. Data and Descriptions.
187 pages. 2002. ISBN 3-906770-08-7 · US-ISBN 0-8204-5933-X
- Vol. 6 Vijay Bhatia, Christopher N. Candlin & Maurizio Gotti (eds)
Legal Discourse in Multilingual and Multicultural Contexts.
Arbitration Texts in Europe.
385 pages. 2003. ISBN 3-906770-85-0 · US-ISBN 0-8204-6254-3
- Vol. 7 Marina Dossena & Charles Jones (eds)
Insights into Late Modern English. 2nd edition.
378 pages. 2003, 2007.
ISBN 978-3-03911-257-9 · US-ISBN 978-0-8204-8927-8
- Vol. 8 Maurizio Gotti
Specialized Discourse. Linguistic Features and Changing Conventions.
351 pages. 2003, 2005.
ISBN 3-03910-606-6 · US-ISBN 0-8204-7000-7
- Vol. 9 Alan Partington, John Morley & Louann Haarman (eds)
Corpora and Discourse.
420 pages. 2004. ISBN 3-03910-026-2 · US-ISBN 0-8204-6262-4
- Vol. 10 Martina Möllering
The Acquisition of German Modal Particles. A Corpus-Based Approach.
290 pages. 2004. ISBN 3-03910-043-2 · US-ISBN 0-8204-6273-X
- Vol. 11 David Hart (ed.)
English Modality in Context. Diachronic Perspectives.
261 pages. 2003. ISBN 3-03910-046-7 · US-ISBN 0-8204-6852-5
- Vol. 12 Wendy Swanson
Modes of Co-reference as an Indicator of Genre.
430 pages. 2003. ISBN 3-03910-052-1 · US-ISBN 0-8204-6855-X
- Vol. 13 Gina Poncini
Discursive Strategies in Multicultural Business Meetings.
2nd edition. 338 pages. 2004, 2007.
ISBN 978-3-03911-296-8 · US-ISBN 978-0-8204-8937-7
- Vol. 14 Christopher N. Candlin & Maurizio Gotti (eds)
Intercultural Aspects of Specialized Communication.
2nd edition. 369 pages. 2004, 2007.
ISBN 978-3-03911-258-6 · US-ISBN 978-0-8204-8926-1
- Vol. 15 Gabriella Del Lungo Camiciotti & Elena Tognini Bonelli (eds)
Academic Discourse. New Insights into Evaluation.
234 pages. 2004. ISBN 3-03910-353-9 · US-ISBN 0-8204-7016-3
- Vol. 16 Marina Dossena & Roger Lass (eds)
Methods and Data in English Historical Dialectology.
405 pages. 2004. ISBN 3-03910-362-8 · US-ISBN 0-8204-7018-X
- Vol. 17 Judy Noguchi
The Science Review Article. An Opportune Genre in
the Construction of Science.
274 pages. 2006. ISBN 3-03910-426-8 · US-ISBN 0-8204-7034-1

- Vol. 18 Giuseppina Cortese & Anna Duszak (eds)
Identity, Community, Discourse. English in Intercultural Settings.
495 pages. 2005. ISBN 3-03910-632-5 · US-ISBN 0-8204-7163-1
- Vol. 19 Anna Trosborg & Poul Erik Flyvholm Jørgensen (eds)
Business Discourse. Texts and Contexts.
250 pages. 2005. ISBN 3-03910-606-6 · US-ISBN 0-8204-7000-7
- Vol. 20 Christopher Williams
Tradition and Change in Legal English. Verbal Constructions
in Prescriptive Texts.
2nd revised edition. 216 pages. 2005, 2007. ISBN 978-3-03911-444-3.
- Vol. 21 Katarzyna Dziubalska-Kolaczyk & Joanna Przedlacka (eds)
English Pronunciation Models: A Changing Scene.
2nd edition. 476 pages. 2005, 2008. ISBN 978-3-03911-682-9.
- Vol. 22 Cristián Abello-Contesse, Rubén Chacón-Beltrán,
M. Dolores López-Jiménez & M. Mar Torreblanca-López (eds)
Age in L2 Acquisition and Teaching.
214 pages. 2006. ISBN 3-03910-668-6 · US-ISBN 0-8204-7174-7
- Vol. 23 Vijay K. Bhatia, Maurizio Gotti, Jan Engberg & Dorothee Heller (eds)
Vagueness in Normative Texts.
474 pages. 2005. ISBN 3-03910-653-8 · US-ISBN 0-8204-7169-0
- Vol. 24 Paul Gillaerts & Maurizio Gotti (eds)
Genre Variation in Business Letters. 2nd printing.
407 pages. 2008. ISBN 978-3-03911-681-2.
- Vol. 25 Ana María Hornero, María José Luzón & Silvia Murillo (eds)
Corpus Linguistics. Applications for the Study of English.
2nd printing. 526 pages. 2006, 2008. ISBN 978-3-03911-726-0
- Vol. 26 J. Lachlan Mackenzie & María de los Ángeles Gómez-González (eds)
Studies in Functional Discourse Grammar.
259 pages. 2005. ISBN 3-03910-696-1 · US-ISBN 0-8204-7558-0
- Vol. 27 Debbie G. E. Ho
Classroom Talk. Exploring the Sociocultural Structure of Formal ESL Learning.
2nd edition. 254 pages. 2006, 2007. ISBN 978-3-03911-434-4
- Vol. 28 Javier Pérez-Guerra, Dolores González-Álvarez, Jorge L. Bueno-Alonso
& Esperanza Rama-Martínez (eds)
'Of Varying Language and Opposing Creed.' New Insights into Late Modern English.
455 pages. 2007. ISBN 978-3-03910-788-9
- Vol. 29 Francesca Bargiela-Chiappini & Maurizio Gotti (eds)
Asian Business Discourse(s).
350 pages. 2005. ISBN 3-03910-804-2 · US-ISBN 0-8204-7574-2
- Vol. 30 Nicholas Brownlees (ed.)
News Discourse in Early Modern Britain. Selected Papers of CHINED 2004.
300 pages. 2006. ISBN 3-03910-805-0 · US-ISBN 0-8204-8025-8
- Vol. 31 Roberta Facchinetti & Matti Rissanen (eds)
Corpus-based Studies of Diachronic English.
300 pages. 2006. ISBN 3-03910-851-4 · US-ISBN 0-8204-8040-1

- Vol. 32 Marina Dossena & Susan M. Fitzmaurice (eds)
Business and Official Correspondence. Historical Investigations.
209 pages. 2006. ISBN 3-03910-880-8 · US-ISBN 0-8204-8352-4
- Vol. 33 Giuliana Garzone & Srikant Sarangi (eds)
Discourse, Ideology and Specialized Communication.
494 pages. 2007. ISBN 978-3-03910-888-6
- Vol. 34 Giuliana Garzone & Cornelia Ilie (eds)
The Use of English in Institutional and Business Settings.
An Intercultural Perspective.
372 pages. 2007. ISBN 978-3-03910-889-3
- Vol. 35 Vijay K. Bhatia & Maurizio Gotti (eds)
Explorations in Specialized Genres.
316 pages. 2006. ISBN 3-03910-995-2 · US-ISBN 0-8204-8372-9
- Vol. 36 Heribert Picht (ed.)
Modern Approaches to Terminological Theories and Applications.
432 pages. 2006. ISBN 3-03911-156-6 · US-ISBN 0-8204-8380-X
- Vol. 37 Anne Wagner & Sophie Cacciaguidi-Fahy (eds)
Legal Language and the Search for Clarity / Le langage juridique et la quête de clarté. Practice
and Tools / Pratiques et instruments.
487 pages. 2006. ISBN 3-03911-169-8 · US-ISBN 0-8204-8388-5
- Vol. 38 Juan Carlos Palmer-Silveira, Miguel F. Ruiz-Garrido &
Inmaculada Fortanet-Gómez (eds)
Intercultural and International Business Communication.
Theory, Research and Teaching.
2nd edition. 343 pages. 2006, 2008. ISBN 978-3-03911-680-5
- Vol. 39 Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt &
Herbert Schendl (eds)
Syntax, Style and Grammatical Norms. English from 1500–2000.
250 pages. 2006. ISBN 3-03911-181-7 · US-ISBN 0-8204-8394-X
- Vol. 40 Marina Dossena & Irma Taavitsainen (eds)
Diachronic Perspectives on Domain-Specific English.
280 pages. 2006. ISBN 3-03910-176-0 · US-ISBN 0-8204-8391-5
- Vol. 41 John Flowerdew & Maurizio Gotti (eds)
Studies in Specialized Discourse.
293 pages. 2006. ISBN 3-03911-178-7
- Vol. 42 Ken Hyland & Marina Bondi (eds)
Academic Discourse Across Disciplines.
320 pages. 2006. ISBN 3-03911-183-3 · US-ISBN 0-8204-8396-6
- Vol. 43 Paul Gillaerts & Philip Shaw (eds)
The Map and the Landscape. Norms and Practices in Genre.
256 pages. 2006. ISBN 3-03911-182-5 · US-ISBN 0-8204-8395-4
- Vol. 44 Maurizio Gotti & Davide Giannoni (eds)
New Trends in Specialized Discourse Analysis.
301 pages. 2006. ISBN 3-03911-184-1 · US-ISBN 0-8204-8381-8
- Vol. 45 Maurizio Gotti & Françoise Salager-Meyer (eds)
Advances in Medical Discourse Analysis. Oral and Written Contexts.
492 pages. 2006. ISBN 3-03911-185-X · US-ISBN 0-8204-8382-6

- Vol. 46 Maurizio Gotti & Susan Šarčević (eds)
Insights into Specialized Translation.
396 pages. 2006. ISBN 3-03911-186-8 · US-ISBN 0-8204-8383-4
- Vol. 47 Khurshid Ahmad & Margaret Rogers (eds)
Evidence-based LSP. Translation, Text and Terminology.
584 pages. 2007. ISBN 978-3-03911-187-9
- Vol. 48 Hao Sun & Dániel Z. Kádár (eds)
It's the Dragon's Turn. Chinese Institutional Discourses.
262 pages. 2008. ISBN 978-3-03911-175-6
- Vol. 49 Cristina Suárez-Gómez
Relativization in Early English (950-1250). the Position of Relative Clauses.
149 pages. 2006. ISBN 3-03911-203-1 · US-ISBN 0-8204-8904-2
- Vol. 50 Maria Vittoria Calvi & Luisa Chierichetti (eds)
Nuevas tendencias en el discurso de especialidad.
319 pages. 2006. ISBN 978-3-03911-261-6
- Vol. 51 Mari Carmen Campoy & María José Luzón (eds)
Spoken Corpora in Applied Linguistics.
274 pages. 2008. ISBN 978-3-03911-275-3
- Vol. 52 Konrad Ehlich & Dorothee Heller (Hrsg.)
Die Wissenschaft und ihre Sprachen.
323 pages. 2006. ISBN 978-3-03911-272-2
- Vol. 53 Jingyu Zhang
The Semantic Salience Hierarchy Model. The L2 Acquisition of Psych Predicates
273 pages. 2007. ISBN 978-3-03911-300-2
- Vol. 54 Norman Fairclough, Giuseppina Cortese & Patrizia Ardizzone (eds)
Discourse and Contemporary Social Change.
555 pages. 2007. ISBN 978-3-03911-276-0
- Vol. 55 Jan Engberg, Marianne Grove Ditlevsen, Peter Kastberg & Martin Stegu (eds)
New Directions in LSP Teaching.
331 pages. 2007. ISBN 978-3-03911-433-7
- Vol. 56 Dorothee Heller & Konrad Ehlich (Hrsg.)
Studien zur Rechtskommunikation.
322 pages. 2007. ISBN 978-3-03911-436-8
- Vol. 57 Teruhiro Ishiguro & Kang-kwong Luke (eds)
Grammar in Cross-Linguistic Perspective.
The Syntax, Semantics, and Pragmatics of Japanese and Chinese.
304 pages. 2012. ISBN 978-3-03911-445-0
- Vol. 58 Carmen Frehner
Email – SMS – MMS
294 pages. 2008. ISBN 978-3-03911-451-1
- Vol. 59 Isabel Balteiro
The Directionality of Conversion in English. A Dia-Synchronic Study.
276 pages. 2007. ISBN 978-3-03911-241-8
- Vol. 60 Maria Milagros Del Saz Rubio
English Discourse Markers of Reformulation.
237 pages. 2007. ISBN 978-3-03911-196-1

- Vol. 61 Sally Burgess & Pedro Martín-Martín (eds)
English as an Additional Language in Research Publication and Communication.
259 pages. 2008. ISBN 978-3-03911-462-7
- Vol. 62 Sandrine Onillon
Pratiques et représentations de l'écrit.
458 pages. 2008. ISBN 978-3-03911-464-1
- Vol. 63 Hugo Bowles & Paul Seedhouse (eds)
Conversation Analysis and Language for Specific Purposes.
2nd edition. 337 pages. 2007, 2009. ISBN 978-3-0343-0045-2
- Vol. 64 Vijay K. Bhatia, Christopher N. Candlin & Paola Evangelisti Allori (eds)
Language, Culture and the Law.
The Formulation of Legal Concepts across Systems and Cultures.
342 pages. 2008. ISBN 978-3-03911-470-2
- Vol. 65 Jonathan Culpeper & Dániel Z. Kádár (eds)
Historical (Im)politeness.
300 pages. 2010. ISBN 978-3-03911-496-2
- Vol. 66 Linda Lombardo (ed.)
Using Corpora to Learn about Language and Discourse.
237 pages. 2009. ISBN 978-3-03911-522-8
- Vol. 67 Natsumi Wakamoto
Extroversion/Introversion in Foreign Language Learning.
Interactions with Learner Strategy Use.
159 pages. 2009. ISBN 978-3-03911-596-9
- Vol. 68 Eva Alcón-Soler (ed.)
Learning How to Request in an Instructed Language Learning Context.
260 pages. 2008. ISBN 978-3-03911-601-0
- Vol. 69 Domenico Pezzini
The Translation of Religious Texts in the Middle Ages.
428 pages. 2008. ISBN 978-3-03911-600-3
- Vol. 70 Tomoko Tode
Effects of Frequency in Classroom Second Language Learning.
Quasi-experiment and stimulated-recall analysis.
195 pages. 2008. ISBN 978-3-03911-602-7
- Vol. 71 Egor Tsedryk
Fusion symétrique et alternances ditransitives.
211 pages. 2009. ISBN 978-3-03911-609-6
- Vol. 72 Cynthia J. Kellett Bidoli & Elana Ochse (eds)
English in International Deaf Communication.
444 pages. 2008. ISBN 978-3-03911-610-2
- Vol. 73 Joan C. Beal, Carmela Nocera & Massimo Sturiale (eds)
Perspectives on Prescriptivism.
269 pages. 2008. ISBN 978-3-03911-632-4
- Vol. 74 Carol Taylor Torsello, Katherine Ackerley & Erik Castello (eds)
Corpora for University Language Teachers.
308 pages. 2008. ISBN 978-3-03911-639-3

- Vol. 75 María Luisa Pérez Cañado (ed.)
English Language Teaching in the European Credit Transfer System.
Facing the Challenge.
251 pages. 2009. ISBN 978-3-03911-654-6
- Vol. 76 Marina Dossena & Ingrid Tiekens-Boon van Ostade (eds)
Studies in Late Modern English Correspondence. Methodology and Data.
291 pages. 2008. ISBN 978-3-03911-658-4
- Vol. 77 Ingrid Tiekens-Boon van Ostade & Wim van der Wurff (eds)
Current Issues in Late Modern English.
436 pages. 2009. ISBN 978-3-03911-660-7
- Vol. 78 Marta Navarro Coy (ed.)
Practical Approaches to Foreign Language Teaching and Learning.
297 pages. 2009. ISBN 978-3-03911-661-4
- Vol. 79 Qing Ma
Second Language Vocabulary Acquisition.
333 pages. 2009. ISBN 978-3-03911-666-9
- Vol. 80 Martin Solly, Michelangelo Conoscenti & Sandra Campagna (eds)
Verbal/Visual Narrative Texts in Higher Education.
384 pages. 2008. ISBN 978-3-03911-672-0
- Vol. 81 Meiko Matsumoto
From Simple Verbs to Periphrastic Expressions:
The Historical Development of Composite Predicates, Phrasal Verbs,
and Related Constructions in English.
235 pages. 2008. ISBN 978-3-03911-675-1
- Vol. 82 Melinda Dooley
Doing Diversity. Teachers' Construction of Their Classroom Reality.
180 pages. 2009. ISBN 978-3-03911-687-4
- Vol. 83 Victoria Guillén-Nieto, Carmen Marimón-Llorca & Chelo Vargas-Sierra (eds)
Intercultural Business Communication and
Simulation and Gaming Methodology.
392 pages. 2009. ISBN 978-3-03911-688-1
- Vol. 84 Maria Grazia Guido
English as a Lingua Franca in Cross-cultural Immigration Domains.
285 pages. 2008. ISBN 978-3-03911-689-8
- Vol. 85 Erik Castello
Text Complexity and Reading Comprehension Tests.
352 pages. 2008. ISBN 978-3-03911-717-8
- Vol. 86 Maria-Lluïsa Gea-Valor, Isabel García-Izquierdo & Maria-José Esteve (eds)
Linguistic and Translation Studies in Scientific Communication.
317 pages. 2010. ISBN 978-3-0343-0069-8
- Vol. 87 Carmen Navarro, Rosa M^a Rodríguez Abella, Francesca Dalle Pezze
& Renzo Miotti (eds)
La comunicación especializada.
355 pages. 2008. ISBN 978-3-03911-733-8

- Vol. 88 Kiriko Sato
The Development from Case-Forms to Prepositional Constructions
in Old English Prose.
231 pages. 2009. ISBN 978-3-03911-763-5
- Vol. 89 Dorothee Heller (Hrsg.)
Formulierungsmuster in deutscher und italienischer Fachkommunikation.
Intra- und interlinguale Perspektiven.
315 pages. 2008. ISBN 978-3-03911-778-9
- Vol. 90 Henning Bergenholtz, Sandro Nielsen & Sven Tarp (eds)
Lexicography at a Crossroads. Dictionaries and Encyclopedias Today,
Lexicographical Tools Tomorrow.
372 pages. 2009. ISBN 978-3-03911-799-4
- Vol. 91 Manouchehr Moshtagh Khorasani
The Development of Controversies. From the Early Modern Period
to Online Discussion Forums.
317 pages. 2009. ISBN 978-3-3911-711-6
- Vol. 92 María Luisa Carrió-Pastor (ed.)
Content and Language Integrated Learning. Cultural Diversity.
178 pages. 2009. ISBN 978-3-3911-818-2
- Vol. 93 Roger Berry
Terminology in English Language Teaching. Nature and Use.
262 pages. 2010. ISBN 978-3-0343-0013-1
- Vol. 94 Roberto Cagliero & Jennifer Jenkins (eds)
Discourses, Communities, and Global Englishes
240 pages. 2010. ISBN 978-3-0343-0012-4
- Vol. 95 Facchinetti Roberta, Crystal David, Seidlhofer Barbara (eds)
From International to Local English – And Back Again.
268 pages. 2010. ISBN 978-3-0343-0011-7
- Vol. 96 Cesare Gagliardi & Alan Maley (eds)
EIL, ELF, Global English. Teaching and Learning Issues
376 pages. 2010. ISBN 978-3-0343-0010-0
- Vol. 97 Sylvie Hancil (ed.)
The Role of Prosody in Affective Speech.
403 pages. 2009. ISBN 978-3-03911-696-6
- Vol. 98 Marina Dossena & Roger Lass (eds)
Studies in English and European Historical Dialectology.
257 pages. 2009. ISBN 978-3-0343-0024-7
- Vol. 99 Christine Béal
Les interactions quotidiennes en français et en anglais.
De l'approche comparative à l'analyse des situations interculturelles.
424 pages. 2010. ISBN 978-3-0343-0027-8
- Vol. 100 Maurizio Gotti (ed.)
Commonality and Individuality in Academic Discourse.
398 pages. 2009. ISBN 978-3-0343-0023-0
- Vol. 101 Javier E. Díaz Vera & Rosario Caballero (eds)
Textual Healing. Studies in Medieval English Medical, Scientific and Technical Texts.
213 pages. 2009. ISBN 978-3-03911-822-9

- Vol. 102 Nuria Edo Marzá
The Specialised Lexicographical Approach. A Step further in Dictionary-making.
316 pages. 2009. ISBN 978-3-0343-0043-8
- Vol. 103 Carlos Prado-Alonso, Lidia Gómez-García, Iria Pastor-Gómez &
David Tizón-Couto (eds)
New Trends and Methodologies in Applied English Language Research.
Diachronic, Diatopic and Contrastive Studies.
348 pages. 2009. ISBN 978-3-0343-0046-9
- Vol. 104 Françoise Salager-Meyer & Beverly A. Lewin
Crossed Words. Criticism in Scholarly Writing?
371 pages. 2011. ISBN 978-3-0343-0049-0.
- Vol. 105 Javier Ruano-García
Early Modern Northern English Lexis. A Literary Corpus-Based Study.
611 pages. 2010. ISBN 978-3-0343-0058-2
- Vol. 106 Rafael Monroy-Casas
Systems for the Phonetic Transcription of English. Theory and Texts.
280 pages. 2011. ISBN 978-3-0343-0059-9
- Vol. 107 Nicola T. Owtram
The Pragmatics of Academic Writing.
A Relevance Approach to the Analysis of Research Article Introductions.
311 pages. 2009. ISBN 978-3-0343-0060-5
- Vol. 108 Yolanda Ruiz de Zarobe, Juan Manuel Sierra &
Francisco Gallardo del Puerto (eds)
Content and Foreign Language Integrated Learning.
Contributions to Multilingualism in European Contexts
343 pages. 2011. ISBN 978-3-0343-0074-2
- Vol. 109 Ángeles Linde López & Rosalía Crespo Jiménez (eds)
Professional English in the European context. The EHEA challenge.
374 pages. 2010. ISBN 978-3-0343-0088-9
- Vol. 110 Rosalía Rodríguez-Vázquez
The Rhythm of Speech, Verse and Vocal Music. A New Theory.
394 pages. 2010. ISBN 978-3-0343-0309-5
- Vol. 111 Anastasios Tsangalidis & Roberta Facchinetti (eds)
Studies on English Modality. In Honour of Frank Palmer.
392 pages. 2009. ISBN 978-3-0343-0310-1
- Vol. 112 Jing Huang
Autonomy, Agency and Identity in Foreign Language Learning and Teaching.
400 pages. 2013. ISBN 978-3-0343-0370-5
- Vol. 113 Mihhail Lotman & Maria-Kristiina Lotman (eds)
Frontiers in Comparative Prosody. In memoriam: Mikhail Gasparov.
426 pages. 2011. ISBN 978-3-0343-0373-6
- Vol. 114 Merja Kytö, John Scahill & Harumi Tanabe (eds)
Language Change and Variation from Old English to Late Modern English.
A Festschrift for Minoji Akimoto
422 pages. 2010. ISBN 978-3-0343-0372-9
- Vol. 115 Giuliana Garzone & Paola Catenaccio (eds)
Identities across Media and Modes. Discursive Perspectives.
379 pages. 2009. ISBN 978-3-0343-0386-6

- Vol. 116 Elena Landone
Los marcadores del discurso y cortesía verbal en español.
390 pages. 2010. ISBN 978-3-0343-0413-9
- Vol. 117 Maurizio Gotti & Christopher Williams (eds)
Legal Discourse across Languages and Cultures.
339 pages. 2010. ISBN 978-3-0343-0425-2
- Vol. 118 David Hirsh
Academic Vocabulary in Context.
217 pages. 2010. ISBN 978-3-0343-0426-9
- Vol. 119 Yvonne Dröschel
Lingua Franca English. The Role of Simplification and Transfer.
358 pages. 2011. ISBN 978-3-0343-0432-0
- Vol. 120 Tengku Sepora Tengku Mahadi, Helia Vaezian & Mahmoud Akbari
Corpora in Translation. A Practical Guide.
135 pages. 2010. ISBN 978-3-0343-0434-4
- Vol. 121 Davide Simone Giannoni & Celina Frade (eds)
Researching Language and the Law. Textual Features and Translation Issues.
278 pages. 2010. ISBN 978-3-0343-0443-6
- Vol. 122 Daniel Madrid & Stephen Hughes (eds)
Studies in Bilingual Education.
472 pages. 2011. ISBN 978-3-0343-0474-0
- Vol. 123 Vijay K. Bhatia, Christopher N. Candlin & Maurizio Gotti (eds)
The Discourses of Dispute Resolution.
290 pages. 2010. ISBN 978-3-0343-0476-4
- Vol. 124 Davide Simone Giannoni
Mapping Academic Values in the Disciplines. A Corpus-Based Approach.
288 pages. 2010. ISBN 978-3-0343-0488-7
- Vol. 125 Giuliana Garzone & James Archibald (eds)
Discourse, Identities and Roles in Specialized Communication.
419 pages. 2010. ISBN 978-3-0343-0494-8
- Vol. 126 Iria Pastor-Gómez
The Status and Development of N+N Sequences in
Contemporary English Noun Phrases.
216 pages. 2011. ISBN 978-3-0343-0534-1
- Vol. 127 Carlos Prado-Alonso
Full-verb Inversion in Written and Spoken English.
261 pages. 2011. ISBN 978-3-0343-0535-8
- Vol. 128 Tony Harris & María Moreno Jaén (eds)
Corpus Linguistics in Language Teaching.
214 pages. 2010. ISBN 978-3-0343-0524-2
- Vol. 129 Tetsuji Oda & Hiroyuki Eto (eds)
Multiple Perspectives on English Philology and History of Linguistics.
A Festschrift for Shoichi Watanabe on his 80th Birthday.
378 pages. 2010. ISBN 978-3-0343-0480-1

- Vol. 130 Luisa Chierichetti & Giovanni Garofalo (eds)
Lengua y Derecho. líneas de investigación interdisciplinaria.
283 pages. 2010. 978-3-0343-0463-4
- Vol. 131 Paola Evangelisti Allori & Giuliana Garzone (eds)
Discourse, Identities and Genres in Corporate Communication.
Sponsorship, Advertising and Organizational Communication.
324 pages. 2011. 978-3-0343-0591-4
- Vol. 132 Leyre Ruiz de Zarobe & Yolanda Ruiz de Zarobe (eds)
Speech Acts and Politeness across Languages and Cultures.
402 pages. 2012. 978-3-0343-0611-9
- Vol. 133 Thomas Christiansen
Cohesion. A Discourse Perspective.
387 pages. 2011. 978-3-0343-0619-5
- Vol. 134 Giuliana Garzone & Maurizio Gotti
Discourse, Communication and the Enterprise. Genres and Trends.
451 pages. 2011. ISBN 978-3-0343-0620-1
- Vol. 135 Zsuzsa Hoffmann
Ways of the World's Words.
Language Contact in the Age of Globalization.
334 pages 2011. ISBN 978-3-0343-0673-7
- Vol. 136 Cecilia Varcasia (ed.)
Becoming Multilingual.
Language Learning and Language Policy between Attitudes and Identities.
213 pages. 2011. ISBN 978-3-0343-0687-5
- Vol. 137 Susy Macqueen
The Emergence of Patterns in Second Language Writing.
A Sociocognitive Exploration of Lexical Trails.
325 pages. 2012. ISBN 978-3-0343-1010-9
- Vol. 138 Maria Vittoria Calvi & Giovanna Mapelli (eds)
La lengua del turismo. Géneros discursivos y terminología.
365 pages. 2011. ISBN 978-3-0343-1011-6
- Vol. 139 Ken Lau
Learning to Become a Professional in a Textually-Mediated World.
A Text-Oriented Study of Placement Practices.
261 pages. 2012. ISBN 978-3-0343-1016-1
- Vol. 140 Sandra Campagna, Giuliana Garzone, Cornelia Ilie & Elizabeth Rowley-Jolivet (eds)
Evolving Genres in Web-mediated Communication.
337 pages. 2012. ISBN 978-3-0343-1013-0
- Vol. 141 Edith Esch & Martin Solly (eds)
The Sociolinguistics of Language Education in International Contexts.
263 pages. 2012. ISBN 978-3-0343-1009-3
- Vol. 142 Forthcoming.
- Vol. 143 David Tizón-Couto
Left Dislocation in English. A Functional-Discoursal Approach.
416 pages. 2012. ISBN 978-3-0343-1037-6

- Vol. 144 Margrethe Petersen & Jan Engberg (eds)
Current Trends in LSP Research. Aims and Methods.
323 pages. 2011. ISBN 978-3-0343-1054-3
- Vol. 145 David Tizón-Couto, Beatriz Tizón-Couto, Iria Pastor-Gómez & Paula Rodríguez-Puente (eds)
New Trends and Methodologies in Applied English Language Research II.
Studies in Language Variation, Meaning and Learning.
283 pages. 2012. ISBN 978-3-0343-1061-1
- Vol. 146 Rita Salvi & Hiromasa Tanaka (eds)
Intercultural Interactions in Business and Management.
306 pages. 2011. ISBN 978-3-0343-1039-0
- Vol. 147 Francesco Straniero Sergio & Caterina Falbo (eds)
Breaking Ground in Corpus-based Interpreting Studies.
254 pages. 2012. ISBN 978-3-0343-1071-0
- Vol. 148 Forthcoming.
- Vol. 149 Vijay K. Bhatia & Paola Evangelisti Allori (eds)
Discourse and Identity in the Professions. Legal, Corporate and Institutional Citizenship.
352 pages. 2011. ISBN 978-3-0343-1079-6
- Vol. 150 Maurizio Gotti (ed.)
Academic Identity Traits. A Corpus-Based Investigation.
363 pages. 2012. ISBN 978-3-0343-1141-0
- Vol. 151 Priscilla Heynderickx, Sylvain Dieltjens, Geert Jacobs, Paul Gillaerts & Elizabeth de Groot (eds)
The Language Factor in International Business.
New Perspectives on Research, Teaching and Practice.
320 pages. 2012. ISBN 978-3-0343-1090-1
- Vol. 152 Paul Gillaerts, Elizabeth de Groot, Sylvain Dieltjens, Priscilla Heynderickx & Geert Jacobs (eds)
Researching Discourse in Business Genres. Cases and Corpora.
215 pages. 2012. ISBN 978-3-0343-1092-5
- Vol. 153 Yongyan Zheng
Dynamic Vocabulary Development in a Foreign Language.
262 pages. 2012. ISBN 978-3-0343-1106-9
- Vol. 154 Carmen Argondizzo (ed.)
Creativity and Innovation in Language Education.
357 pages. 2012. ISBN 978-3-0343-1080-2
- Vol. 155 David Hirsh (ed.)
Current Perspectives in Second Language Vocabulary Research.
180 pages. 2012. ISBN 978-3-0343-1108-3
- Vol. 156 Seiji Shinkawa
Unhistorical Gender Assignment in Lahamon's *Brut*. A Case Study of a Late Stage in the Development of Grammatical Gender toward its Ultimate Loss.
186 pages. 2012. ISBN 978-3-0343-1124-3
- Vol. 157 Yeonkwon Jung
Basics of Organizational Writing: A Critical Reading Approach.
151 pages. 2014. ISBN 978-3-0343-1137-3.

- Vol. 158 Bárbara Eizaga Rebolgar (ed.)
Studies in Linguistics and Cognition.
301 pages. 2012. ISBN 978-3-0343-1138-0
- Vol. 159 Giuliana Garzone, Paola Catenaccio, Chiara Degano (eds)
Genre Change in the Contemporary World. Short-term Diachronic Perspectives.
329 pages. 2012. ISBN 978-3-0343-1214-1
- Vol. 160 Carol Berkenkotter, Vijay K. Bhatia & Maurizio Gotti (eds)
Insights into Academic Genres.
468 pages. 2012. ISBN 978-3-0343-1211-0
- Vol. 161 Beatriz Tizón-Couto
Clausal Complements in Native and Learner Spoken English. A corpus-based study
with Lindsei and Vicolse. 357 pages. 2013. ISBN 978-3-0343-1184-7
- Vol. 162 Patrizia Anesa
Jury Trials and the Popularization of Legal Language. A Discourse Analytical Approach.
247 pages. 2012. ISBN 978-3-0343-1231-8
- Vol. 163 David Hirsh
Endangered Languages, Knowledge Systems and Belief Systems.
153 pages. 2013. ISBN 978-3-0343-1232-5
- Vol. 164 Eugenia Sainz (ed.)
De la estructura de la frase al tejido del discurso. Estudios contrastivos español/italiano.
305 pages. 2014. ISBN 978-3-0343-1253-0
- Vol. 165 Julia Bamford, Franca Poppi & Davide Mazzi (eds)
Space, Place and the Discursive Construction of Identity.
367 pages. 2014. ISBN 978-3-0343-1249-3
- Vol. 166 Rita Salvi & Janet Bowker (eds)
Space, Time and the Construction of Identity.
Discursive Indexicality in Cultural, Institutional and Professional Fields.
324 pages. 2013. ISBN 978-3-0343-1254-7
- Vol. 167 Shunji Yamazaki & Robert Sigley (eds)
Approaching Language Variation through Corpora. A Festschrift in Honour of Toshio Saito.
421 pages. 2013. ISBN 978-3-0343-1264-6
- Vol. 168 Franca Poppi
Global Interactions in English as a Lingua Franca. How written communication is changing
under the influence of electronic media and new contexts of use.
249 pages. 2012. ISBN 978-3-0343-1276-9
- Vol. 169 Miguel A. Aijón Oliva & María José Serrano
Style in syntax. Investigating variation in Spanish pronoun subjects.
239 pages. 2013. ISBN 978-3-0343-1244-8
- Vol. 170 Inés Olza, Óscar Loureda & Manuel Casado-Velarde (eds)
Language Use in the Public Sphere. Methodological Perspectives and Empirical Applications
564 pages. 2014. ISBN 978-3-0343-1286-8
- Vol. 171 Aleksandra Matulewska
Legilinguistic Translatology. A Parametric Approach to Legal Translation.
279 pages. 2013. ISBN 978-3-0343-1287-5

- Vol. 172 Maurizio Gotti & Carmen Sancho Guinda (eds)
Narratives in Academic and Professional Genres.
513 pages. 2013. ISBN 978-3-0343-1371-1
- Vol. 173 Madalina Chitez
Learner corpus profiles. The case of Romanian Learner English.
244 pages. 2014. ISBN 978-3-0343-1410-7
- Vol. 174 Chihiro Inoue
Task Equivalence in Speaking Tests.
251 pages. 2013. ISBN 978-3-0343-1417-6
- Vol. 175 Gabriel Quiroz & Pedro Patiño (eds.)
LSP in Colombia: advances and challenges.
339 pages. 2014. ISBN 978-3-0343-1434-3
- Vol. 176 Catherine Resche
Economic Terms and Beyond: Capitalising on the Wealth of Notions.
How Researchers in Specialised Varieties of English Can Benefit from Focusing on Terms.
332 pages. 2013. ISBN 978-3-0343-1435-0
- Vol. 177 Wei Wang
Media representation of migrant workers in China. Identities and stances
198 pages. 2018. 978-3-0343-1436-7
- Vol. 178 Cécile Desoutter & Caroline Mellet (dir.)
Le discours rapporté: approches linguistiques et perspectives didactiques.
270 pages. 2013. ISBN 978-3-0343-1292-9
- Vol. 179 Ana Díaz-Negrillo & Francisco Javier Díaz-Pérez (eds)
Specialisation and Variation in Language Corpora.
341 pages. 2014. ISBN 978-3-0343-1316-2
- Vol. 180 Pilar Alonso
A Multi-dimensional Approach to Discourse Coherence. From Standardness to Creativity.
247 pages. 2014. ISBN 978-3-0343-1325-4
- Vol. 181 Alejandro Alcaraz-Sintes & Salvador Valera-Hernández (eds)
Diachrony and Synchrony in English Corpus Linguistics.
393 pages. 2014. ISBN 978-3-0343-1326-1
- Vol. 182 Runhan Zhang
Investigating Linguistic Knowledge of a Second Language.
207 pages. 2015. ISBN 978-3-0343-1330-8
- Vol. 183 Hajar Abdul Rahim & Shakila Abdul Manan (eds.)
English in Malaysia. Postcolonial and Beyond.
267 pages. 2014. ISBN 978-3-0343-1341-4
- Vol. 184 Virginie Fasel Lauzon
Comprendre et apprendre dans l'interaction. Les séquences d'explication en classe
de français langue seconde.
292 pages. 2014. ISBN 978-3-0343-1451-0
- Vol. 185 Forthcoming.
- Vol. 186 Wei Ren
L2 Pragmatic Development in Study Abroad Contexts
256 pages. 2015. ISBN 978-3-0343-1358-2
- Vol. 187 Marina Bondi & Rosa Lorés Sanz (eds)
Abstracts in Academic Discourse. Variation and Change.
361 pages. 2014. ISBN 978-3-0343-1483-1

- Vol. 188 Giuditta Caliendo
Rethinking Community. Discourse, Identity and Citizenship in the European Union.
240 pages. 2017. ISBN 978-3-0343-1561-6
- Vol. 189 Paola Evangelisti Allori (ed.)
Identities in and across Cultures.
315 pages. 2014. ISBN 978-3-0343-1458-9
- Vol. 190 Erik Castello, Katherine Ackerley & Francesca Coccetta (eds).
Studies in Learner Corpus Linguistics. Research and Applications for Foreign Language
Teaching and Assessment.
358 pages. 2015. ISBN 978-3-0343-1506-7
- Vol. 191 Ruth Breeze, Maurizio Gotti & Carmen Sancho Guinda (eds)
Interpersonality in Legal Genres.
389 pages. 2014. ISBN 978-3-0343-1524-1
- Vol. 192 Paola Evangelisti Allori, John Bateman & Vijay K. Bhatia (eds)
Evolution in Genre. Emergence, Variation, Multimodality.
364 pages. 2014. ISBN 978-3-0343-1533-3
- Vol. 193 Jiyeon Kook
Agency in Arzt-Patient-Gesprächen. Zur interaktionistischen Konzeptualisierung von Agency
271 pages. 2015. ISBN 978-3-0343-1666-8
- Vol. 194 Susana Nicolás Román & Juan José Torres Núñez (eds)
Drama and CLIL. A new challenge for the teaching approaches in bilingual education.
170 pages. 2015. ISBN 978-3-0343-1629-3
- Vol. 195 Alessandra Molino & Serenella Zanotti (eds)
Observing Norm, Observing Usage. Lexis in Dictionaries and in the Media.
430 pages. 2015. ISBN 978-3-0343-1584-5
- Vol. 196 Begoña Soneira
A Lexical Description of English for Architecture. A Corpus-based Approach.
267 pages. 2015. ISBN 978-3-0343-1602-6
- Vol. 197 M Luisa Roca-Varela
False Friends in Learner Corpora. A corpus-based study of English false friends
in the written and spoken production of Spanish learners.
348 pages. 2015. ISBN 978-3-0343-1620-0
- Vol. 198 Rahma Al-Mahrooqi & Christopher Denman
Bridging the Gap between Education and Employment. English Language Instruction in
EFL Contexts.
416 pages. 2015. ISBN 978-3-0343-1681-1
- Vol. 199 Rita Salvi & Janet Bowker (eds)
The Dissemination of Contemporary Knowledge in English. Genres, discourse strategies
and professional practices.
171 pages. 2015. ISBN 978-3-0343-1679-8
- Vol. 200 Maurizio Gotti & Davide S. Giannoni (eds)
Corpus Analysis for Descriptive and Pedagogical Purposes. ESP Perspectives.
432 pages. 2014. ISBN 978-3-0343-1516-6
- Vol. 201 Ida Ruffolo
The Perception of Nature in Travel Promotion Texts. A Corpus-based Discourse Analysis.
148 pages. 2015. ISBN 978-3-0343-1521-0

- Vol. 202 Ives Trevian
English suffixes. Stress-assignment properties, productivity, selection and combinatorial processes.
471 pages. 2015. ISBN 978-3-0343-1576-0
- Vol. 203 Maurizio Gotti, Stefania Maci & Michele Sala (eds)
Insights into Medical Communication.
422 pages. 2015. ISBN 978-3-0343-1694-1
- Vol. 204 Carmen Argondizzo (ed.)
European Projects in University Language Centres. Creativity, Dynamics, Best Practice.
371 pages. 2015. ISBN 978-3-0343-1696-5
- Vol. 205 Aura Luz Duffé Montalván (ed.)
Estudios sobre el léxico. Puntos y contrapuntos.
502 pages. 2016. ISBN 978-3-0343-2011-5
- Vol. 206 Maria Pavesi, Maicol Formentelli & Elisa Ghia (eds)
The Languages of Dubbing. Mainstream Audiovisual Translation in Italy.
275 pages. 2014. ISBN 978-3-0343-1646-0
- Vol. 207 Ruth Breeze & Inés Olza (eds)
Evaluation in media discourse. European perspectives.
268 pages. 2017. ISBN 978-3-0343-2014-6
- Vol. 208 Vijay K. Bhatia & Maurizio Gotti (eds)
Arbitration Discourse in Asia.
331 pages. 2015. ISBN 978-3-0343-2032-0
- Vol. 209 Sofia Bemposta-Rivas, Carla Bouzada-Jabois, Yolanda Fernández-Pena, Tamara Bouso, Yolanda J. Calvo-Benzies, Iván Tamaredo (eds)
New trends and methodologies in applied English language research III. Synchronic and diachronic studies on discourse, lexis and grammar processing.
280 pages. 2017. ISBN 978-3-0343-2039-9
- Vol. 210 Francisco Alonso Almeida, Laura Cruz García & Víctor González Ruiz (eds)
Corpus-based studies on language varieties.
285 pages. 2016. ISBN 978-3-0343-2044-3
- Vol. 211 Juan Pedro Rica Peromingo
Aspectos lingüísticos y técnicos de la traducción audiovisual (TAV).
177 pages. 2016. ISBN 978-3-0343-2055-9
- Vol. 212 Maria Vender
Disentangling Dyslexia. VenderPhonological and Processing Deficit in Developmental Dyslexia.
338 pages. 2017. ISBN 978-3-0343-2064-1
- Vol. 213 Zhilong Xie
Bilingual Advantages. Contributions of Different Bilingual Experiences to Cognitive Control Differences Among Young-adult Bilinguals.
221 pages. 2016. ISBN 978-3-0343-2081-8
- Vol. 214 Larissa D'Angelo
Academic posters. A textual and visual metadiscourse analysis.
367 pages. 2016. ISBN 978-3-0343-2083-2
- Vol. 215 Evelyne Berger
Prendre la parole en L2. Regard sur la compétence d'interaction en classe de langue.
246 pages. 2016. ISBN 978-3-0343-2084-9

- Vol. 216 David Lasagabaster and Aintzane Doiz (eds)
CLIL experiences in secondary and tertiary education: In search of good practices.
262 pages. 2016. ISBN 978-3-0343-2104-4
- Vol. 217 Elena Kkese
Identifying Plosives in L2 English: The Case of L1 Cypriot Greek Speakers.
317 pages. 2016. ISBN 978-3-0343-2060-3
- Vol. 218 Sandra Campagna, Elana Ochse, Virginia Pulcini & Martin Solly (eds)
Languaging in and across Communities: New Voices, New Identities. Studies in Honour
of Giuseppina Cortese.
507 pages. 2016. ISBN 978-3-0343-2073-3
- Vol. 219 Adriana Orlandi & Laura Giacomini (ed.)
Defining collocation for lexicographic purposes. From linguistic theory to lexicographic
practice.
328 pages. 2016. ISBN 978-3-0343-2054-2
- Vol. 220 Pietro Luigi Iaia
Analysing English as a Lingua Franca in Video Games. Linguistic Features,
Experiential and Functional Dimensions of Online and Scripted Interactions.
139 pages. 2016. ISBN 978-3-0343-2138-9
- Vol. 221 Dimitrinka G. Nikleva (ed.)
La formación de los docentes de español para inmigrantes en distintos contextos educativos.
390 pages. 2017. ISBN 978-3-0343-2135-8
- Vol. 222 Katherine Ackerley, Marta Guarda & Francesca Helm (eds)
Sharing Perspectives on English-Medium Instruction.
308 pages. 2017. ISBN 978-3-0343-2537-0
- Vol. 223 Juana I. Marín-Arrese, Julia Lavid-López, Marta Carretero, Elena Domínguez Romero,
M^a Victoria Martín de la Rosa & María Pérez Blanco (eds)
Evidentiality and Modality in European Languages. Discourse-pragmatic perspectives.
427 pages. 2017. ISBN 978-3-0343-2437-3
- Vol. 224 Gilles Col
Construction du sens : un modèle instructionnel pour la sémantique.
292 pages. 2017. ISBN 978-3-0343-2572-1
- Vol. 225 Ana Chiquito & Gabriel Quiroz (eds)
Pobreza, Lenguaje y Medios en América Latina.
362 pages. 2017. ISBN 978-3-0343-2142-6
- Vol. 226 Xu Zhang
English Quasi-Numeral Classifiers. A Corpus-Based Cognitive-Typological Study.
360 pages. 2017. ISBN 978-3-0343-2818-0
- Vol. 227 María Ángeles Orts, Ruth Breeze & Maurizio Gotti (eds)
Power, Persuasion and Manipulation in Specialised Genres. Providing Keys to the
Rhetoric of Professional Communities.
368 pages. 2017. ISBN 978-3-0343-3010-7
- Vol. 228 Maurizio Gotti, Stefania Maci & Michele Sala (eds)
Ways of Seeing, Ways of Being: Representing the Voices of Tourism.
453 pages. 2017. ISBN 978-3-0343-3031-2
- Vol. 229 Dino Selvaggi
Plurilingual Code-Switching between Standard and Local Varieties.
A Socio-Psycholinguistic Approach
371 pages. 2018. ISBN 978-3-0343-2663-6

- Vol. 230 Anca-Cristina Sterie
Interprofessional interactions at the hospital. Nurses' requests and reports of problems in calls with physicians.
371 pages. 2017. ISBN 978-3-0343-2734-3
- Vol. 231 Xiaodong Zhang
Understanding Chinese EFL Teachers' Beliefs and Practices in the Textbook-Based Classroom.
189 pages. 2017. ISBN 978-3-0343-3053-4
- Vol. 232 Manuela Caterina Moroni & Federica Ricci Garotti (Hrsg.)
Brücken schlagen zwischen Sprachwissenschaft und DaF-Didaktik.
345 pages. 2017. ISBN 978-3-0343-2667-4
- Vol. 233 Dimitrinka Georgieva Nikleva
Necesidades y tendencias en la formación del profesorado de español como lengua extranjera
401 pages. 2017. ISBN 978-3-0343-2946-0
- Vol. 234 Juan Santana-Lario & Salvador Valera (Hrsg.)
Competing patterns in English affixation.
272 pages. 2017. ISBN 978-3-0343-2701-5
- Vol. 235 Francisco Salgado-Robles
Desarrollo de la competencia sociolingüística por aprendices de español en un contexto de inmersión en el extranjero
241 pages. 2018. ISBN 978-3-0343-2323-9
- Vol. 236 Maria Chiara Janner
Sguardi linguistici sulla marca. Analisi morfosintattica dei nomi commerciali in italiano
345 pages. 2017. ISBN 978-3-0343-2667-4
- Vol. 237 Bárbara Herrero Muñoz-Cobo & Otman El Azami Zalachi
La primavera del árabe marroquí.
192 pages. 2017. ISBN 978-3-0343-3104-3
- Vol. 238 Consuelo Pascual Escagedo
El papel del oyente en la construcción de la conversación espontánea de estudiantes italianos en su interlengua y en su lengua materna
295 pages. 2017. ISBN 978-3-0343-3186-9
- Vol. 239 Stefania M. Maci
The MS Digby 133 *Mary Magdalene*. Beyond scribal practices: language, discourse, values and attitudes.
336 pages. 2017. ISBN 978-3-0343-3256-9
- Vol. 240 Eliecer Crespo-Fernández
Taboo in Discourse. Studies on Attenuation and Offence in Communication.
326 pages. 2018. ISBN 978-3-0343-3018-3
- Vol. 241 Jana Altmanova, Maria Centrella, Katherine E. Russo (eds)
Terminology & Discourse / Terminologie et discours.
424 pages. 2018. ISBN 978-3-0343-2417-5
- Vol. 242 Xavier Blanco et Inès Sfar (dir.)
Lexicologie(s) : approches croisées en sémantique lexicale.
442 pages. 2018. ISBN 978-3-0343-3056-5

- Vol. 243 Yunfeng Ge
Resolution of Conflict of Interest in Chinese Civil Court Hearings.
A Perspective of Discourse Information Theory.
302 pages. 2018. ISBN 978-3-0343-3313-9
- Vol. 244 Carla Vergaro
Illocutionary Shell Nouns in English
322 pages. 2018. ISBN 978-3-0343-3069-5
- Vol. 245 Paolo Frassi
L'adjectif en français et sa définition lexicographique.
270 pages. 2018. ISBN 978-3-0343-3394-8
- Vol. 246 Suwilai Premsrirat and David Hirsh (eds)
Language Revitalization. Insights from Thailand
328 pages. 2018. ISBN 978-3-0343-3497-6
- Vol. 247 Wei Wang
Researching Learning and Learners in Genre-based Academic Writing Instruction
282 pages. 2018. ISBN 978-3-0343-3297-2
- Vol. 248 Isusi Alabarte, Alberto & Lahuerta Martínez, Ana Cristina (eds)
La comprensión lectora de lengua extranjera
Estudio de los factores de familiaridad, interés, género y métodos de evaluación
336 pages. 2018. ISBN 978-3-0343-3493-8
- Vol. 249 Mercedes Eurrutia Cavero
Approche didactique du langage techno-scientifique
Terminologie et discours
374 pages. 2018. ISBN 978-3-0343-3512-6
- Vol. 250 Aurora Ruiz Mezcuca (ed.)
Approaches to Telephone Interpretation
Research, Innovation, Teaching and Transference
268 pages. 2018. ISBN 978-3-0343-3330-6
- Vol. 251 Morini Massimiliano
A Day in the News
A Stylistic Analysis of Newsspeak
188 pages. 2018. ISBN 978-3-0343-3507-2
- Vol. 252 Ignacio Guillén-Galve & Ignacio Vázquez-Orta (eds.)
English as a Lingua Franca and Intercultural Communication
Implications and Applications in the Field of English Language Teaching
414 pages. 2018. ISBN 978-3-0343-2763-3
- Vol. 253 Bianca Del Villano
Using the Devil with Courtesy
Shakespeare and the Language of (Im)Politeness
216 pages. 2018. ISBN 978-3-0343-2315-4
- Vol. 254 David Hirsh (ed.)
Explorations in Second Language Vocabulary Research
252 pages. 2018. ISBN 978-3-0343-2940-8

- Vol. 255 Tania Baumann (ed.)
Reiseführer - Sprach- und Kulturmittlung im Tourismus / Le guide turistiche - mediazione
linguistica e culturale in ambito turistico
270 pages. 2018. ISBN 978-3-0343-3402-0
- Vol. 256 Ariadna Sánchez-Hernández & Ana Herraiz-Martínez (eds)
Learning second language pragmatics beyond traditional contexts
376 pages. 2018. ISBN 978-3-0343-3437-2
- Vol. 257 Albert Bastardas-Boada, Emili Boix-Fuster, Rosa Maria Torrens (eds)
Family Multilingualism in Medium-Sized Linguistic Communities
336 pages. 2019. ISBN 978-3-0343-2536-3
- Vol. 258 Yuyang Cai
Examining the Interaction among Components of English for Specific Purposes Ability in
Reading. The Triple-Decker Model
296 pages. 2020. 978-3-0343-2913-2
- Vol. 259 Catia Nannoni
Participe présent et gérondif dans la presse française contemporaine
176 pages. 2019. ISBN 978-3-0343-3631-4
- Vol. 260 Nieves Rodríguez Pérez & Bárbara Heinsch (eds.)
Contextos multilingües. Mediadores interculturales, formación del profesorado de
lenguas extranjeras
289 pages. 2019. ISBN 978-3-0343-3768-7
- Vol. 261 Giuliana Elena Garzone, Mara Logaldo, Francesca Santulli (eds.)
Investigating Conflict Discourses in the Periodical Press
244 pages. 2020. ISBN 978-3-0343-3668-0
- Vol. 262 Laura Nadal
Lingüística experimental y contraargumentación
233 pages. 2019. ISBN 978-3-0343-3791-5
- Vol. 263 Claudia Claridge & Merja Kytö (eds.)
Punctuation in Context – Past and Present Perspectives
288 pages. 2019. ISBN 978-3-0343-3790-8
- Vol. 264 Maurizio Gotti, Stefania Maci, Michele Sala (eds.)
Scholarly Pathways
530 pages. 2020. ISBN 978-3-0343-3860-8
- Vol. 265 Ruth Breeze, Ana M. Fernández Vallejo (eds.)
Politics and populism across modes and media
350 pages. 2020. ISBN 978-3-0343-3707-6
- Vol. 266 Jean Marguerite Jimenez
Understanding the Effects of Immediate Electronic Corrective Feedback on Second
Language Development
252 pages. 2020. ISBN 978-3-0343-3815-8
- Vol. 267 Sergio Rodríguez-Tapia, Adela González-Fernández (eds)
Lenguas y turismo: estudios en torno al discurso, la didáctica y la divulgación
380 pages. 2020. ISBN 978-3-0343-3881-3

- Vol. 268 Ana Bocanegra-Valle (ed.)
Applied Linguistics and Knowledge Transfer. Employability, Internationalisation and Social Challenges
344 pages. 2020. ISBN 978-3-0343-3714-4
- Vol. 269 Beatrice Garzelli
La traducción audiovisual español-italiano. Películas y cortos entre humor y habla soez
202 pages. 2020. ISBN 978-3-0343-4013-7
- Vol. 270 Iván Tamaredo
Complexity, Efficiency, and Language Contact. Pronoun Omission in World Englishes
292 pages. 2020. ISBN 978-3-0343-3902-5
- Vol. 271 Silvia Domenica Zollo,
Origine et histoire du vocabulaire des arts de la table. Analyse lexicale et exploitation de corpus textuels
239 pages. 2020. ISBN 978-3-0343-3890-5
- Vol. 272 Paola Paissa, Michelangelo Conoscenti, Ruggero Druetta, Martin Solly (eds.)
Metaphor and Conflict / Métaphore et conflit
385 pages. 2021. 978-3-0343-4068-7
- Vol. 273 María Martínez-Atienza de Dios,
Entre el léxico y la sintaxis: las fases de los eventos
178 pages. 2021. 978-3-0343-4173-8
- Vol. 274 Noelia Castro-Chao,
Argument Structure in Flux. The Development of Impersonal Constructions in Middle and Early Modern English, with Special Reference to Verbs of Desire.
300 pages. 2021. 978-3-0343-4189-9
- Vol. 275 Gabriella Carobbio, Cécile Desoutter, Aurora Fragonara (eds.)
Macht, Ratio und Emotion: Diskurse im digitalen Zeitalter / Pouvoir, raison et émotion: les discours à l'ère du numérique
246 pages. 2020. 978-3-0343-4184-4
- Vol. 276 Miguel Fuster-Márquez, José Santaemilia, Carmen Gregori-Signes, Paula Rodríguez-Abrunheiras (eds.)
Exploring discourse and ideology through corpora
293 pages. 2021. 978-3-0343-3969-8
- Vol. 277 Tamara Bouso
Changes in Argument Structure
The Transitivity Reaction Object Construction
392. 2021. 978-3-0343-4095-3
- Vol. 278 Maria Luisa Maggioni, Amanda Murphy (eds.)
Back to the Future. English from Past to Present
250 pages. 2021. 978-3-0343-4273-5

- Vol. 279 Luisa Chierichetti
Diálogos de serie. Una aproximación a la construcción discursiva de personajes basada en corpus
246 pages. 2021. 978-3-0343-4274-2
- Vol. 280 Cristina Lastres-López
From subordination to insubordination. A functional-pragmatic approach to if/si-constructions in English, French and Spanish spoken discourse
258 pages. 2021. 978-3-0343-4220-9
- Vol. 281 Eleonora Federici, Stefania Maci (eds.)
Gender issues. Translating and mediating languages, cultures and societies
500 pages. 2021. 978-3-0343-4022-9
- Vol. 282 José Mateo, Francisco Yus (eds.)
Metaphor in Economics and specialised Discourse
354 pages. 2021. 978-3-0343-4048-9
- Vol. 283 Nicholas Brownlee (ed.)
The Role of Context in the Production and Reception of Historical News Discourse
376 pages. 2021. 978-3-0343-4181-3
- Vol. 284 Catalina Fuentes Rodríguez, María Ester Brenes Peña, Víctor Pérez Béjar (eds.)
Sintaxis discursiva: construcciones y operadores en español
396 pages. 2021. 978-3-0343-4306-0
- Vol. 285 Dominic Stewart
Frequency in the dictionary. A corpus-assisted contrastive analysis of English and Italian
176 pages. 2021. 978-3-0343-4368-8
- Vol. 286 Carla Bouzada-Jabois
Nonfinite supplements in the recent history of English
320 pages. 2021. 978-3-0343-4226-1
- Vol. 287 Sofia Bemposta Rivas
Verb-governed infinitival complementation in the recent history of English
320 pages. 2021. 978-3-0343-4227-8
- Vol. 288 Mirella Agorni
Translating Italy for the Nineteenth Century. Translators and an Imagined Nation in the Early Romantic Period 1816–1830s
180 pages. 2021. 978-3-0343-3612-3
- Vol. 289 David Hirsh (ed.)
Research Perspectives in Language and Education
276 pages. 2022. 978-3-0343-4219-3
- Vol. 290 Federica Vezzani
Terminologie numérique : conception, représentation et gestion
238 pages. 2022. 978-3-0343-4394-7