

Numéro 17 / Année 2021

Synergies Italie

Revue du GERFLINT

Industries des langues France-Italie

Coordonné par Annick Farina
et Valeria Zotti



Synergies Italie

Numéro 17 / Année 2021

Industries des langues France-Italie

Coordonné par Annick Farina
et Valeria Zotti



REVUE DU GERFLINT
2021

POLITIQUE EDITORIALE

Synergies Italie est une revue francophone de recherche en sciences humaines et sociales, particulièrement ouverte à l'ensemble des sciences du langage et de la communication.

Sa vocation est de mettre en œuvre, en Italie, le *Programme Mondial de Diffusion Scientifique Francophone en Réseau* du GERFLINT, Groupe d'Études et de Recherches pour le Français Langue Internationale. C'est pourquoi elle publie des articles dans cette langue, mais sans exclusivité linguistique et accueille, de façon majoritaire, les travaux issus de la pensée scientifique des chercheurs francophones de son espace géographique dont le français n'est pas la langue première. Comme toutes les revues du GERFLINT, elle poursuit les objectifs suivants : défense de la recherche scientifique francophone dans l'ensemble des sciences humaines, promotion du dialogue entre les disciplines, les langues et les cultures, ouverture sur l'ensemble de la communauté scientifique, adoption d'une large couverture disciplinaire, aide aux jeunes chercheurs, formation à l'écriture scientifique francophone, veille sur la qualité scientifique des travaux.

Libre Accès et Copyright : © **Synergies Italie** est une revue française éditée par le GERFLINT qui se situe dans le cadre du libre accès à l'information scientifique et technique. Sa commercialisation est interdite. Sa politique éditoriale et ses articles peuvent être directement consultés et étudiés dans leur intégralité en ligne. Le mode de citation doit être conforme au Code français de la Propriété Intellectuelle. La Rédaction de *Synergies Italie*, partenaire de coopération scientifique du GERFLINT, travaille selon les dispositions de la Charte éthique, éditoriale et de confidentialité du Groupe et de ses normes les plus strictes. Les propos tenus dans ses articles sont conformes au débat scientifique et n'engagent que la responsabilité de l'auteur. Conformément aux règles déontologiques et éthiques du domaine de la Recherche, toute fraude scientifique (plagiat, auto-plagiat, retrait inopiné de proposition d'article sans en informer dûment la Rédaction) sera communiquée à l'entourage universitaire et professionnel du signataire de la proposition d'article. Toute procédure irrégulière entraîne refus systématique du texte et annulation de la collaboration.

Périodicité : annuelle
ISSN 1724-0700 / ISSN en ligne 2260-8087

Directeur de publication

Jacques Cortès, Professeur émérite, Université de Rouen Normandie, France

Coordination éditoriale générale et révision du numéro

Sophie Aubin, Universitat de València, Espagne

Présidente d'Honneur

Marie-Berthe Vittoz, Université de Turin, Italie

Rédactrice en chef

Rachele Raus, Université de Turin, Italie

Rédactrice en chef adjointe

Maria Margherita Mattioda, Université de Turin, Italie

Titulaire et Éditeur : GERFLINT

Siège en France

GERFLINT

17, rue de la Ronde mare

Le Buisson Chevalier

27240 Sylvains-les-Moulins - France

www.gerflint.fr

gerflint.edition@gmail.com

Siège de la rédaction en Italie

Université de Turin - Dipartimento di Lingue e Letterature straniere e Cultura moderna, Via S. Ottavio, 20 – 10124 Turin (Italie).

Tél : 011.6702153

Contact de la Rédaction :

synergies.italie@gmail.com

Comité scientifique

Ruth Amosy (Université de Tel-Aviv, Israël) ; Michel Berré (Université de Mons, Belgique), Josiane Boutet (Université de Paris-Sorbonne, France), Sergio Cappello (Université de Udine, Italie), Melita Cataldi (Université de Turin, Italie), Nadine Celotti (Université de Trieste, Italie), Giovanni Dotoli (Université de Bari, Italie), Marie-Marthe Gervais-le Garff (Université de Plymouth, Royaume-Uni), Douglas A. Kibbee (Université de l'Illinois, USA), Alessandro Martini (Université de Lyon III Jean Moulin, France), Eni Orlandi (Université de Campinas, Brésil), Sandrine Reboul-Touré (Université de Paris III, France), Leandro Schena (Université de Modène, Italie).

Comité de lecture permanent

Gerardo Acerenza (Université de Trente, Italie), Giovanni Agresti (Université de Bordeaux Montaigne, France), Maria Cristina Caimotto (Université de Turin, Italie), Roberto Dapavo (Université de Turin, Italie), Annick Farina (Université de Florence, Italie), Patricia Kottelat (Université de Turin, Italie), Gabriella Laffaille (Université de Turin, Italie), Nadia Minerva (Université de Catane, Italie), Benoît Monginot (Université de Turin, Italie), Paola Paissa (Université de Turin, Italie), Daniela Puolato (Université de Naples « Federico II »), Elisa Ravazzolo (Université de Trente, Italie), Giorgio Sale (Université de Sassari, Italie), Mario Squartini (Université de Turin, Italie), Valeria Zotti (Université de Bologne, Italie), Angela Zottola (Université de Turin, Italie).

Évaluateurs invités pour ce numéro

Jana Altmanova (Université de Naples « L'Orientale », Italie), Mathilde Anquetil (Université de Macerata, Italie), Elio Ballardini (Université de Bologne, Italie), Genevieve Bordet (Université Paris-Diderot, France), Dardo De Vecchi (Kedge Business School, France), Paolo Frassi (Université de Vérone, Italie), Nicolas Froeliger (Université Paris-Diderot, France), Yannick Hamon (Université de Venise, Italie), Daniel Henkel (Université Paris 8, France), John Humbley (Université Paris-Diderot, France), Dario Maldussi (Université de Bologne, Italie), Chiara Preite (Université de Modène et de Reggio d'Émilie, Italie), Giovanni Luca Tallarico (Université de Vérone, Italie), Maria Teresa Zanolà (Université Catholique du Sacré-Cœur de Milan, Italie).

Patronages et partenariats

Fondation Maison des Sciences de l'Homme de Paris (FMSH, Pôle *Recherche & prospective*), Sciences Po Lyon (Partenariat institutionnel pour Mir@bel), EBSCO Publishing (EDS), ProQuest.

Numéro financé par le GERFLINT.

PROGRAMME MONDIAL DE DIFFUSION SCIENTIFIQUE FRANCOPHONE EN RÉSEAU

Synergies Italie n° 17 / 2021
<https://gerflint.fr/synergies-italie>



Indexations et références

ABES (SUDOC)
ANVUR
DOAJ
EBSCOhost (Communication Source)
Ent'revues
ERIH Plus
Héloïse
Index Copernicus
ISSN Portal / ROAD
JournalBase (CNRS)
JournalSeek
LISEO (France Education International)
MIAR
Mir@bel
MLA
ProQuest central
Scopus
Scopus Sources
SHERPA-RoMEO
SJR, SCImago Journal & Country Rank
Ulrichsweb
ZDB
Zenodo



Disciplines couvertes par la revue

- Ensemble des Sciences Humaines et Sociales
- Culture et communication internationales
- Sciences du langage
- Littératures francophones
- Didactologie-didactique de la langue-culture française et des langues-cultures
- Éthique et théorie de la complexité

Industries des langues France-Italie

Coordonné par Annick Farina
et Valeria Zotti

Sommaire

Annick Farina, Valeria Zotti	7
Présentation	
Industrie de la traduction	
Eleonora Marzi	19
La traduction automatique neuronale et les biais de genre : le cas des noms de métiers entre l'italien et le français	
Francesca Bisiani	37
Les écarts dénominatifs dans la normative européenne sur la protection de données en matière pénale : une étude entre traduction institutionnelle, dispositifs terminologiques et traduction automatique	
Federica Vezzani	51
La ressource FAIRterm : entre pratique pédagogique et professionnalisation en traduction spécialisée	
Industries de la langue et traitement automatique des langues (TAL)	
Cosimo De Giovanni	67
Corpus, <i>Web</i> et recherche terminologique : le cas des termes du domaine du bien-être animal	
Klara Dankova	79
Les corpus et la traduction spécialisée. Proposition d'un parcours didactique centré sur la terminologie pour les étudiants en traduction (FR - IT)	
Adriana Orlandi, Marco Fasciolo	91
<i>Corpus Pattern Analysis</i> et <i>Classes</i> d'objets : différences théoriques et retombées pratiques de deux approches de description du lexique	
Varia	
Silvia Calvi	109
L'enseignement/apprentissage des unités phraséologiques en terminologie : un parcours didactique en français L2 sur le commerce international	

Comptes rendus de lecture

Annalisa Buonocore	123
Sabina Gola, Michel Pierrard, Evie Tops et Dan Van Raemdonck (éds). <i>Enseigner et apprendre les langues au XXI^e siècle</i> , Bruxelles, Peter Lang, 2019.	
Manon Hayette	127
Christophe Rey. <i>Dictionnaire et Société</i> , Paris, Honoré Champion, 2020.	
Maria Margherita Mattioda	131
Geneviève Felten, <i>Langue commune, cultures distinctes. Les illusions du globish</i> , Laval, Les Presses de l'Université Laval, 2020.	
Giovanni Tallarico	135
Michela Tonti. <i>Le nom de marque dans le discours au quotidien</i> . <i>Prisme lexiculturel et linguistique</i> , Paris, L'Harmattan, 2020	

Annexes

Profils des contributeurs	141
Projet pour le n° 18 / 2022	145
Consignes aux auteurs.....	149
Publications du GERFLINT.....	153



ISSN 1724-0700

ISSN en ligne 2260-8087

Présentation

Annick Farina

Université de Florence, Italie

annick.farina@unifi.it

Valeria Zotti

Université de Bologne, Italie

valeria.zotti@unibo.it



Les industries de la langue ont connu une forte croissance depuis les années quatre-vingt-dix du siècle dernier, aussi bien au sein des institutions européennes et internationales, dans le cadre de programmes de promotion de la diversité linguistique, que dans le secteur privé qui a vu se multiplier le nombre de sociétés actives dans le périmètre de l'ingénierie linguistique, dégagant des chiffres d'affaires de plusieurs millions d'euros (Technolangue, 2007). Les industries de la langue constituent aujourd'hui une grappe industrielle en plein essor et sont des facteurs de croissance économique de plus en plus importants. L'usage du « langage naturel » au sein des systèmes d'information a été reconnu comme une source de productivité certaine, non seulement en termes économiques, mais aussi en termes de quantité de temps investi et de qualité du produit final. Ces vingt dernières années, la frénésie du marché de l'Internet et de la communication multilingue et la progression de l'usage des NTIC ont engendré un élargissement du marché des outils technolinguistiques vers le grand public. Le désir d'utiliser une interface conviviale en langage naturel, qui propose une traduction en ligne, fait désormais partie des exigences banales des internautes (Dabbadie, 2004).

Ce numéro de *Synergies Italie* est l'occasion de présenter quelques recherches en cours en France et en Italie dans le domaine linguistique qui ont des applications concrètes au sein des industries des langues, ainsi que d'analyser les produits (corpus, bases de données terminologiques, outils de traitement automatique) existants, un quart de siècle après l'essor de ces industries. Nous nous sommes intéressées aux retombées que l'emploi généralisé des technologies du traitement automatique des langues (TAL) a eues dans ces deux pays sur le grand public, sur les professionnels de la langue (traducteurs, terminologues, lexicographes) et dans le monde de la recherche universitaire. Nous nous sommes interrogées en particulier sur la présence de la langue française et de la langue italienne (et de la traduction de l'une à l'autre) à l'intérieur des nouveaux instruments développés.

Fortes d'une expérience partagée tant pratique que théorique dans le domaine de la lexicographie monolingue francophone et bilingue franco-italienne et d'un intérêt sans cesse renouvelé pour les nouvelles technologies reliées à l'analyse du lexique, nous nous sommes naturellement tournées vers les méthodes les plus récentes de traitement automatique de la langue. Depuis 2013, nous nous sommes engagées dans un projet interuniversitaire commun, le *Lessico dei Beni Culturali*, visant à la réalisation d'un nouveau dictionnaire plurilingue, entièrement basé sur corpus, qui devrait permettre aux professionnels de la langue et du tourisme d'accéder à une description riche et fiable du lexique du patrimoine artistique. Dans le cadre de ce projet, basé à l'Université de Florence et impliquant plusieurs Universités italiennes et étrangères, nous avons réalisé différents outils qui sont aujourd'hui disponibles en accès libre sur notre site lessicobenculturali.net pour plusieurs langues (allemand, anglais, chinois, français, espagnol, italien, portugais, russe) : des corpus comparables (Billero, Farina, Nicolas, 2020 ; Cetro et Zotti, 2020), des lexiques fondamentaux du patrimoine (Cetro et Zotti, à paraître), et, depuis peu, un nouveau volet qui porte sur la création de corpus parallèles de traduction (voir Zotti, 2017).

Ces différents outils répondent, selon nous, aux besoins actuels des industries de langues, en particulier pour ce qui concerne l'amélioration des outils de TA qui se nourrissent d'énormes quantités de données linguistiques et textuelles et comblent une des lacunes principales que nous avons pu remarquer dans les ressources terminologiques existantes : reliées, comme nous l'avons déjà dit, à des intérêts économiques précis, elles n'ont pas pris en compte certains domaines, comme la terminologie artistique ou la langue littéraire, qui restent pourtant fondamentaux dans les entreprises du tourisme et dans le monde de l'édition internationale.

Les propositions présentées dans ce numéro de *Synergies Italie* sont de différentes natures : les auteurs y exposent de manière réflexive leurs projets de recherche (Marzi, Bisiani), leurs expériences didactiques (Vezzani, Dankova) mais des approches plus théoriques avec des retombées appliquées sont également présentes (De Giovanni, Orlandi et Fasciolo). Les six contributions couvrent les domaines, très féconds dans les industries de la langue, de la traduction automatique, de la linguistique de corpus, de la terminographie, de la lexicographie, ainsi que de l'enseignement des langues et de l'expertise linguistique.

La première partie de ce numéro est consacrée à l'axe thématique « Industrie de la traduction ». Nous avons accueilli trois contributions qui se penchent sur les sujets suivants : les biais de genre dans les outils de traduction automatique neuronale (Marzi), le rôle des dispositifs terminologiques et de traduction automatique dans la diffusion et dans la mise en circulation de la terminologie institutionnelle du droit

pénal (Bisiani), et la conception et à la mise en œuvre d'un nouveau produit linguistique pour la construction de terminologies visant à soutenir les apprenants-traducteurs et les futurs professionnels des langues de spécialité (Vezzani).

La deuxième partie de ce numéro, intitulée « Industries de la langue et traitement automatique des langues (TAL) », est centrée sur deux autres volets des industries des langues qui ont un intérêt théorique et une répercussion pratique à la fois : l'apport des corpus numériques issus du Web pour la recherche terminologique dans le domaine du bien-être animal (De Giovanni), la création de corpus textuels spécialisés pour la traduction de la terminologie de l'urbanisme et l'aménagement du territoire à travers une expérience didactique documentée (Dankova) et, pour finir, l'application d'approches théoriques différentes mais complémentaires, l'une lexicographique, l'autre basée sur corpus, du traitement automatique de la langue (Fasciolo et Orlandi).

Dans la contribution qui ouvre ce numéro, Eleonora Marzi aborde le sujet de l'intelligence artificielle au service de la traduction automatique, un domaine clé de l'industrie de la traduction qui a connu un grand essor dans les dix dernières années. Marzi se penche sur l'analyse de trois outils de traduction automatique neuronale généralistes, disponibles gratuitement en ligne, *Google Translate*, *Microsoft Translator* et *DeepL*, non dans le simple but d'en évaluer et comparer les performances, mais pour éclairer la nature des stéréotypes présents dans les données d'entrées qui nourrissent ces logiciels, un problème majeur dont on a pris conscience récemment. De fait, comme le précise l'auteure, de nombreuses études supposent que l'existence du biais de genre dans la traduction automatique (TA) serait due à une biodiversité insuffisante des données d'entrées. Marzi s'intéresse aux erreurs de traduction en termes de genre qui sont générées par ces systèmes de TA, dans le but de comprendre si ces erreurs sont simplement aléatoires ou bien révélatrices de stéréotypes provenant d'un imaginaire sous-jacent. À partir d'une expérience basée sur un corpus de phrases ayant la même structure syntaxique, traduites en français et en italien dans les deux directions, Marzi dévoile la nature des stéréotypes présents dans les bases de données statistiques des logiciels de TA, en donnant l'exemple d'une liste de 73 noms de métiers déclinés au masculin et au féminin en français et en italien. À travers l'analyse d'un échantillon représentatif, elle confirme son hypothèse de départ, à savoir que la présence dans ces phrases de deux adjectifs appartenant à deux champs sémantiques différents, celui de l'apparence (*beau/belle*) et celui de la compétence (*intelligent/intelligente*), influence l'exactitude des traductions, le taux d'erreurs étant sensiblement plus élevé pour les noms au féminin, ce qui implique l'existence d'une représentation stéréotypée des genres présents dans la langue que les logiciels de TA reproduisent de manière systématique et incontrôlée.

L'Union européenne, qui a fait du multilinguisme l'un de ses principes fondateurs, a commencé à intervenir depuis une dizaine d'années pour faire face au problème de l'inégalité linguistique engendrée par l'intelligence artificielle. Plusieurs mesures ont été adoptées, comme la création du réseau d'excellence META (Alliance de la Technologie pour une Europe multilingue), le lancement du programme ELE (Égalité des Langues en Europe) qui vise à atteindre la complète égalité linguistique d'ici la fin de 2030 et, le 21 avril dernier, la publication d'une proposition pour la régulation de l'approche de l'Union Européenne face à l'intelligence artificielle.

Dans sa contribution, Francesca Bisiani aborde justement la complexité du multilinguisme au sein de l'Union européenne et nous invite à réfléchir sur les écarts interprétatifs qui se produisent lorsque les termes et le discours circulent dans l'espace européen. Elle se penche surtout sur les difficultés liées à l'aménagement linguistique et à la terminologie des concepts du droit européen dans les différentes langues-cultures. Dans sa recherche, l'auteure examine, selon une approche discursive de la terminologie et à travers une démarche à la fois quantitative et qualitative, les variantes dénominatives qui se manifestent dans les versions linguistiques en français, italien, anglais et espagnol des actes européens contraignants qui concernent le traitement des données personnelles en matière pénale (2016-2019). Elle se concentre sur les termes exprimant l'objet de la démarche juridique, c'est-à-dire le champ d'application de la décision-cadre et des directives concernées, en analysant la séquence « prévention + détection + enquêtes + poursuites ».

L'émergence de désalignements conceptuels dans les versions linguistiques des documents étudiés, notamment l'existence de deux séquences terminologiques différentes pour chaque langue, révèle que les décalages conceptuels qui ressortent de la traduction sont des symptômes de discordances idéologiques. La comparaison de ces résultats dans la base de données terminologique de l'UE (*IATE*) et dans deux outils de traduction automatique en libre accès (*DeepL* et *Google Translate*) confirme que les désalignements observés dans les traductions du corpus d'analyse se reproduisent aussi dans les dispositifs terminologiques et dans les outils de traduction automatique, ce qui confirme que ces instruments puisent dans les bases de données documentaires et terminologiques européennes. Ce constat amène l'auteure à souligner le rapport entre la terminologie, la traduction spécialisée et les outils d'aide à la traduction : la masse terminologique qui est employée par les traducteurs, les terminographes, mais aussi par les juristes linguistes et les experts au sein des institutions, alimente les logiciels de traduction automatique ou les concordanciers en ligne. Les termes utilisés par les traducteurs se propagent ainsi

non seulement dans les textes publiés, mais aussi dans les mémoires de traduction, ce qui a pour effet de fabriquer des segments prêts à être réemployés et d'amplifier ce qui a été dit par les instances énonciatrices. Pour cette raison, l'auteure insiste sur la nécessité de revenir de manière critique sur les choix dénominatifs effectués en amont, au moment de la production ou de la traduction de la terminologie institutionnelle juridique et politique.

C'est sur le besoin de développer des ressources plus riches et fiables pour la traduction spécialisée que se penche la contribution de Federica Vezzani. L'auteure propose une ressource, FAIRterm, disponible gratuitement en ligne, qui fournit une méthode pour la formation des futurs traducteurs technico-scientifiques. Ce produit a été conçu comme une base de données multilingue qui collecte des fiches terminologiques structurées et normalisées à des fins didactiques, de traduction spécialisée et de professionnalisation. L'auteure a réalisé un modèle de fiche qui permet aux apprenants-traducteurs de réfléchir au comportement morphosyntaxique, sémantique et phraséologique du terme technique analysé et du candidat terme équivalent. Cette ressource permet en outre d'importer les données structurées obtenues dans les différents systèmes de traduction assistée par ordinateur (TAO) permettant de bénéficier, pendant le processus de traduction spécialisée, des analyses menées pour chaque terme. L'auteure détaille une expérience en traduction spécialisée active dans le domaine œnologique, pour le couple de langues italien-français, qui a permis de valider la méthodologie didactique sous-jacente à l'utilisation de l'application FAIRterm. La perspective est de valider également la fonctionnalité de cette ressource dans d'autres domaines de spécialisation pour pouvoir la mettre à la disposition d'un public large, ce qui répond à un besoin croissant de la communauté scientifique de disposer de données de la recherche accessibles, interopérables et réutilisables.

La deuxième partie de ce numéro s'ouvre avec la contribution de Cosimo De Giovanni, qui se propose de démontrer l'apport des corpus numériques à la recherche terminologique par un cas d'étude. L'auteur analyse la circulation de deux termes, *abattoir mobile* pour le français et *unità mobile di macellazione* pour l'italien, dans deux différents corpus : un corpus parallèle, formé de textes émanant de l'Union européenne et repérés dans le site *EUR-Lex*, qui réglementent le domaine du bien-être animal (BEA), et un corpus composé des textes législatifs promulgués en France et en Italie, les deux analysés entre 1991 et 2009. En inscrivant son étude dans la perspective d'une terminologie communicationnelle, l'auteur montre que l'interprétation des contextes d'apparition des candidats termes se révèle fondamentale pour l'analyse de deux différents points de vue (PdV), un premier concernant les choix faits par le terminologue et le législateur, et un second qui

concerne les choix linguistiques opérés par un groupe de locuteurs sur la scène discursive. L'auteur démontre que l'utilisation du corpus numérique, à la fois du *corpus-web* (un corpus constitué à partir des données repérées sur le *Web* à l'aide de logiciels) et du *web-corpus* (un corpus constitué des données brutes extraites du *Web* à usages linguistiques), s'est avéré essentielle pour préciser la description des variantes dénominatives et conceptuelles des deux termes dans le domaine examiné et pour vérifier le comportement des termes en contexte. La comparaison des deux types de corpus numériques montre que le *web-corpus*, qui se prête bien à des analyses en termes quantitatifs et qualitatifs, pose des problèmes concernant la volatilité des données collectées et la rapidité avec laquelle elles peuvent varier. Il est évident, cependant, que les corpus numériques, sources d'enrichissement et d'intégration de données déjà collectées à partir de corpus traditionnels, peuvent être utiles pour la construction de nouvelles banques de terminologie.

Une constante dans toutes les contributions est la réflexion sur la nature et la qualité des sources employées par les outils technologiques et l'absence de grands corpus textuels informatisés pour certaines langues. Ces sources ne sont pas représentatives de certains domaines ni de certaines langues, et la qualité des données n'est pas toujours fiable, ce qui fait que les traducteurs et terminologues sont amenés à construire leurs propres corpus spécialisés. C'est justement la question abordée par Klara Dankova, qui propose un parcours didactique destiné à des étudiants italophones en traduction FR-IT de niveau de français avancé, en attirant l'attention sur les avantages que l'emploi de corpus peut apporter pour la traduction spécialisée. Des études récentes montrent en effet que, dans l'espace francophone, les corpus ne représentent pas encore des outils d'aide à la traduction utilisés fréquemment et de façon adéquate par les traducteurs professionnels (Loock, 2016 : 1-2). Le parcours proposé par l'auteure se base précisément sur la construction et l'exploitation, à l'aide de l'outil *Sketch Engine*, de corpus FR-IT comparables et parallèles qui concernent l'urbanisme et l'aménagement du territoire. Les sources de documentation sélectionnées couvrent la période 2015-2020 et contiennent de la sorte la terminologie de référence la plus actuelle possible de ce domaine. À l'heure actuelle, la taille modeste des corpus ne permet que des recherches limitées, mais la base textuelle est destinée à être enrichie dans l'avenir. Le parcours didactique, qui accorde une attention particulière à la traduction de la terminologie, est complété par une série d'exercices proposés à partir des besoins de traduction de textes concrets. Les étudiants et futurs traducteurs sont ainsi sensibilisés à l'importance de s'en servir dans la traduction de textes relevant d'un domaine spécialisé, tout en étant informés de leurs limites. La prudence s'impose notamment au niveau terminologique, car, comme l'auteure

le souligne, les termes utilisés dans les documents de l'UE ne sont pas nécessairement pertinents pour la traduction des textes rédigés dans un autre contexte.

La dernière contribution contenue dans ce numéro prolonge la réflexion sur l'apport du domaine de la linguistique de corpus mais en le comparant avec un autre domaine central dans les industries des langues, celui de la lexicographie, et cela avec une approche innovante qui combine la réflexion théorique à l'application de ces théories dans le cadre du traitement automatique des langues. Les deux auteurs, Adriana Orlandi et Marco Fasciolo, proposent de comparer la Théorie des Classes d'Objets (TCO), une approche lexicographique française conçue par Michel Mathieu-Colas et Gaston Gross comme un développement de l'idée de lexique-grammaire de Maurice Gross, avec la *Corpus Pattern Analysis* (CPA), une méthode d'analyse des patrons basée sur les corpus qui est à la base de la ressource linguistique T-Pas. En partant du constat que, dans les deux approches, les notions de *classes d'objets* et de *types sémantiques* d'une part, et les notions de *schéma prédicatif* et de *pattern* de l'autre semblent très proches, les auteurs démontrent qu'il existe des différences non marginales entre la *TCO* et la *CPA*. Ces différences ne concernent pas véritablement les résultats pratiques des deux approches, mais plutôt leurs présupposés théoriques et leurs implications méthodologiques. Les auteurs opposent *TCO* et *CPA* sur deux points spécifiques. Ils examinent tout d'abord la différence entre les notions de *schéma prédicatif* et de *pattern*, et observent qu'elles peuvent être apparentées à la dichotomie saussurienne *langue/parole*. En effet, la notion de *schéma prédicatif* relèverait du niveau de la *phrase-modèle*, alors que la notion de *pattern* relèverait du niveau de l'*énoncé*. Dans un second temps, les auteurs analysent une portion de la polysémie du verbe prédicatif *suivre/seguire*, notamment les acceptions de ce verbe en tant qu'« activité », et comparent les schémas prédicatifs du verbe *suivre* isolés suivant la *TCO*, et les *patterns* du verbe *seguire* dans la ressource T-Pas, isolés moyennant la *CPA*. Ils remarquent que les schémas prédicatifs sont plus nombreux que les T-Pas (*patterns*) et qu'ils sont souvent plus précis et plus spécifiques que ces derniers. Ils constatent ainsi que, sur le plan applicatif, la *CPA* pourrait tirer profit des travaux menés sur la *TCO* au sein du laboratoire LDI, la description des schémas prédicatifs étant plus fine dans la *TCO* que dans la *CPA*, ce qui augmenterait le pouvoir de désambiguïsation des *patterns*. Étant donné qu'au-delà des différences qui caractérisent les deux approches sur le plan théorique et méthodologique, la *CPA* et la *TCO* partagent une même visée applicative lexicographique, c'est-à-dire la réalisation de dictionnaires électroniques pour le traitement automatique du langage, les auteurs souhaitent en dernière instance que les approches et les méthodes existantes en matière d'analyse de corpus sortent des frontières nationales et entament un processus de

connaissance réciproque. L'adoption d'une approche « mixte » pourrait améliorer sensiblement les résultats obtenus et avoir des applications concrètes dans de nombreux domaines de l'industries des langues.

En conclusion, ce numéro de *Synergies Italie* lance une réflexion sur plusieurs aspects des industries des langues sous le signe de l'interdisciplinarité et de la rencontre entre l'informatique et la linguistique. Les études exploratoires présentées ici par des chercheurs du monde universitaire touchent des questions qui sont d'une grande actualité et qui vont à l'encontre des préoccupations des professionnels de l'ingénierie linguistique, des terminographes, des lexicographes, des experts de traduction automatique ainsi que des traducteurs et des enseignants en langues étrangères. Nous espérons que cette réflexion sera enrichie progressivement par leur expertise pratique en la matière.

Bibliographie

- Association Enrages 1986. *Les industries de la langue : enjeux pour l'Europe : actes du colloque de Tours*. Paris : Université de Paris VIII-Vincennes.
- Auger, P. 1989. « Informatique et terminologie : revue des technologies nouvelles ». In : *Actes du colloque terminologie et industries de la langue. Meta : journal des traducteurs*, n° 34 (3), p. 485-492.
- Baker, M., Francis, G., Tognini-Bonelli, E. 1993, *Text & Technology*, Amsterdam/Philadelphie : John Benjamins.
- Bédard, C. 2000. « Mémoire de traduction cherche traducteur de phrases ». *Traduire*, n° 186, p. 41-49.
- Bernardini, S., Ferraresi, A. 2013. *I corpora nella didattica della traduzione: Corpus Use and Learning to Translate*. Bologne : CLUEB.
- Billero, R., Farina, A., Nicolas, C. (éds.) 2020. *I corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*. Florence : Firenze University Press.
- Cetro, R., Zotti, V. 2020. « Les corpus et la base terminologique LBC. Des ressources pour la traduction du patrimoine artistique ». In : Mangeot, M., Tutin, A. *Lexique(s) et genre(s) textuel(s) : approches sur corpus. Actes de la conférence 11^e Journées du réseau « Lexicologie, Terminologie, Traduction »*. Paris : Éditions des archives contemporaines, p. 81-98.
- Cetro, R., Zotti, V. (à paraître). « Le corpus LBC français : bases, développement et applications ». <https://www.lessicobeniculturali.net/>
- Cresti, E., Panunzi A. 2013. *Introduzione ai corpora dell'italiano*. Bologne : Il Mulino.
- Dabbadie, M. 2004. « Industrie de la langue, vous êtes plutôt TIL ou TAL ? ». *VEILLE. Le magazine des professionnels de l'information stratégique*, n° 73.
- Délégation Générale à la Langue Française et aux Langues de France (DGLFLF) 2015. *Mieux comprendre les outils d'aide à la traduction*.
- Groupe de Réflexion sur les Industries de l'Information et les Industries de la Langue (GRIIL) 2005. *Livre blanc. Le traitement automatique des langues dans les industries de l'information*.
- Kubler, N. (éd.) 2011. *Language Corpora, Teaching, and Resources: from Theory to Practice*. Berne : Peter Lang.
- L'Homme, M.-C., Jacquemin, Ch., Bourigault, D. 2001. *Recent Advances in Computational Terminology*. Amsterdam/Philadelphie : John Benjamins.

Loock, R. 2016. *La traductologie de corpus*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

Technolangue 2007. *Technologies de la langue en Europe : marché et tendances* réalisée par le Bureau Van Dijk, à la demande du Ministère de la recherche dans le cadre du programme Technolangue. <http://www.technolangue.net>

Zotti, V. 2017. « L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle Vite di Vasari ». In: Zotti, V., Pano Alaman, A. (éds.). *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*. Florence : Firenze University Press, p. 105-134.

Synergies Italie n° 17 / 2021



Industrie
de la traduction





ISSN 1724-0700

ISSN en ligne 2260-8087

La traduction automatique neuronale et les biais de genre : le cas des noms de métiers entre l'italien et le français

Eleonora Marzi

Université de Bologne, Italie

eleonora.marzi@unibo.it

<https://orcid.org/0000-0002-4558-9483>

Reçu le 10-02-2021 / Évalué le 15-03-2021 / Accepté le 18-04-2021

Résumé

Au cours des dernières années, la traduction automatique (TA) a fait des progrès remarquables grâce à l'utilisation de réseaux de neurones artificiels ; toutefois, l'intérêt et le succès des logiciels de TA - rendus disponibles gratuitement pour le grand public - ont mis en évidence la présence de biais de genre dans les résultats obtenus. Ce biais pourrait dépendre d'une biodiversité insuffisante de données qui servent à l'entraînement des algorithmes d'intelligence artificielle. Prenant pour exemple les noms de métiers, cette étude entend vérifier si la présence de deux différents types d'adjectifs (d'apparence et de compétence) peut influencer la traduction automatique qui concerne le genre en allant du français vers l'italien ou vice-versa. La liste des noms de métiers analysés, qui a été rédigée à partir du système de classification des professions européennes ESCO (*European Skills, Competences, Occupation*), a permis de créer des phrases que nous avons fait traduire par trois moteurs de traduction automatique neuronale : *Google Translate*, *DeepL*, *Microsoft Translator*.

Mots-clés : études de genre, écart entre les sexes, apprentissage profond, ESCO (*European Skills, Competences*), traduction automatique

La traduzione automatica neurale e il divario di genere nel linguaggio: il caso delle professioni tra italiano e francese

Riassunto

Durante gli ultimi anni, la traduzione automatica (TA) ha fatto dei progressi notevoli grazie all'approccio delle reti neurali; l'interesse e il successo degli strumenti di TA - resi disponibili gratuitamente al grande pubblico - hanno messo in evidenza la presenza di un divario di genere nei risultati. Tale divario potrebbe dipendere da una biodiversità insufficiente dei dati che servono ad allenare l'algoritmo d'intelligenza artificiale. Prendendo in considerazione il caso dei nomi delle professioni, il presente studio vuole verificare in che modo la presenza di due diversi tipi di aggettivi (d'apparenza e di competenza) possa influenzare la traduzione automatica del genere dal francese all'italiano o viceversa. L'analisi è stata condotta utilizzando una lista di nomi di professione che abbiamo selezionato a partire dal sistema di classificazione delle professioni europee ESCO (*European Skills, Competences*,

Occupation) e che è servita per creare delle frasi che abbiamo fatto tradurre da tre motori di traduzione neuronale: *Google Translate*, *DeepL*, *Microsoft Translator*.

Parole chiave: studi di genere, *gender data gap*, *deep learning*, ESCO (*European Skills, Competences, Occupation*), traduzione automatica

Neural machine translation and gender gap in language: the case profession names in Italian and French

Abstract

In recent years, machine translation (MT) has made considerable progress thanks to the neural network approach. The interest and the success of MT tools - made freely available to the general public - have highlighted the presence of a gender gap in the results. This gap could be due to an insufficient biodiversity of the data used to train the artificial intelligence algorithm. Considering as a case study the names of professions, the present article aims to verify how the presence of two different types of adjectives (appearance and competence) can influence the automatic translation of gender from French to Italian or vice versa. The analysis was carried out using a list of profession names that we selected from the ESCO Project (*European Skills, Competences, Occupation*) and which was used to create sentences that we had translated by three neural translation engines: *Google Translate*, *DeepL*, *Microsoft Translator*.

Keywords: gender studies, *gender data gap*, *deep learning*, ESCO (*European Skills, Competences, Occupation*), automatic translation

1. La traduction automatique : fonctionnement et typologies

Dans son essai *Dire presque la même chose*, Umberto Eco donne l'exemple de textes traduits de manière automatique qui deviennent presque anecdotiques. La réflexion d'Eco exemplifie les deux sentiments opposés que la traduction automatique suscite depuis toujours. D'un côté les technophobes (Eco, 1964), convaincus d'une distance infranchissable entre la performance humaine et celle de la machine, et de l'autre, les technophiles qui rêvent d'un instrument capable d'imiter les capacités humaines.

L'histoire de la traduction automatique oscille entre ces deux positions, scandée par l'introduction de nouvelles technologies, notamment les corpus numériques et les techniques d'apprentissage liées à l'intelligence artificielle. Ce mouvement incessant entre la confiance envers une traduction automatique comparable à ce que ferait l'être humain, et le découragement qui la voit condamnée à un état d'imperfection, n'a jamais cessé et a produit diverses générations d'élaboration dans le domaine de la traduction, toutes basées chacune sur une idée différente de la traduction.

Les tout premiers systèmes de traduction automatique (TA) naissent autour des années 1950. Ils sont fondés sur une approche de la traduction entendue comme décryptage d'un code (Weaver, 1949), traduisant donc mot à mot. Les éléments qui composaient ces systèmes étaient un compilateur (*parser*), permettant d'identifier la relation entre les mots de la langue source et ceux de la langue d'arrivée, un dictionnaire monolingue, avec les informations morphosyntaxiques de la langue source, et un dictionnaire bilingue, utilisé pour permettre au système de trouver l'équivalent dans la langue cible. Cette traduction mot à mot rendait évidemment difficile la traduction des expressions métaphoriques ou ambiguës.

L'attention portée à la grammaire a permis de développer une deuxième génération de systèmes informatiques à base de règles (*rule-based*). Ces systèmes partent d'une conception théorique de la traduction pour laquelle, grâce à l'analyse de la représentation du signifié dans la langue source, il est possible de créer une représentation équivalente dans la langue cible¹ ; au niveau technologique, elles sont construites à travers une série de règles morphologiques, syntaxiques et sémantiques. Si ce type de système a l'avantage de traduire correctement les structures syntaxiquement non parallèles entre deux langues, il présente encore des défauts. Tout d'abord, l'impossibilité d'incorporer dans une liste toutes les règles grammaticales d'une langue. Ensuite, pour une interprétation correcte des ambiguïtés liées aux termes homographes et polysémiques, il est nécessaire de prendre en considération le cotexte et le contexte. Les travaux successifs ont emprunté cette direction et, aux alentours des années 1980, l'apparition de corpus de grande taille en format numérique a permis de créer des systèmes basés sur les corpus (*corpus-based*). Ce modèle s'appuie sur une grande quantité d'exemples corrects et alignés, qui sont identifiés grâce à des calculs statistiques. Il s'agit, en quelque sorte, d'une tentative de résoudre le problème lié à la connaissance du contexte : par exemple, l'expression idiomatique italienne « *in bocca al lupo* » est traduite en français par « bonne chance » en raison de la récurrence de ces séquences, et de leur co-occurrence régulière dans des segments alignés correspondants. Toutefois, cette solution présente une limite : se basant sur des statistiques tirées d'exemples du corpus, le système n'est pas en mesure de traduire des expressions linguistiques inhabituelles ou poétiques et plus généralement les expressions qui présentent un certain niveau de créativité ou qui sont peu représentées dans le corpus.

Au début des années 2000, un tournant important est marqué par l'introduction des techniques issues de l'intelligence artificielle. Le système de TA devient alors capable d'apprendre et d'améliorer ses performances grâce à des algorithmes entraînés sur des quantités énormes de données non structurées. En 2014, les réseaux de neurones artificiels apparaissent : cette architecture entend imiter

les réseaux biologiques des neurones du cerveau humain grâce à une organisation par niveaux qui interagissent entre eux à travers des nœuds contenant des centres de connaissance. L'élaboration des données grâce aux réseaux de neurones artificiels permet de créer des modèles de connaissance qui s'améliorent sans cesse. Cette technologie, qui s'appuie sur des systèmes définis à base de connaissance (*knowledge-based*), est utilisée pour la traduction automatique neuronale (*Neural Machine Translation - NMT*). Bien qu'elle exploite divers outils - tels que dictionnaires, règles grammaticales et sémantiques, glossaires spécialisés et mémoires de traductions - elle représente en quelque sorte une évolution de la traduction basée sur les corpus (Wilks, 2009).

2. Le biais de genre et la traduction automatique

En 2016, *Google* a lancé son traducteur automatique neuronal ; en 2018, *Microsoft*, *Systran* et *Linguee* en ont fait de même. La mise à disposition gratuite des outils de TA a modifié l'interface des logiciels suggérant par exemple à l'utilisateur une traduction correcte ou lui permettant d'évaluer la traduction faite par le moteur. Cette pratique ne découle pas d'une sensibilité démocratique envers la connaissance, mais elle vise à tirer profit de la compétence linguistique des utilisateurs qui, de cette manière, deviennent des experts linguistiques bénévoles.

La visibilité majeure acquise récemment par la traduction automatique ainsi que l'usage massif de ce type de traduction ont conduit à une prise de conscience à propos des stéréotypes de genre dans les traductions². Cette tendance s'est accrue grâce à un intérêt grandissant et à une sensibilité renouvelée pour la question du genre dans le langage, dont témoigne, entre autres, le débat récent sur le langage épïcène³.

De nombreuses études, que nous précisons dans cette section, supposent que l'existence du biais de genre dans les traductions automatiques serait due à une biodiversité insuffisante des données d'entrées. En effet, les architectures neuronales constituent l'évolution des architectures à base de corpus, qui sont alors unies aux approches d'apprentissage profond (*deep learning*) : cela veut dire que les données de sortie sont grandement affectées par la qualité des données d'entrées qui entraînent les algorithmes d'intelligence artificielle. Bien que l'on obtienne de meilleurs résultats lorsque le moteur de traduction automatique est entraîné sur un corpus représentant le même type de discours spécifique que celui du texte à traduire, il faut aussi considérer que la langue (et par conséquent le corpus) est le reflet de la société : longtemps, certains noms de métiers n'ont pas disposé d'équivalents féminins, puisque ces métiers n'étaient jamais exercés par

des femmes. Par exemple, si le mot « infirmière » est normalement bien attesté, c'est aussi parce que cette profession, qui a subi une forte croissance depuis la première et la seconde guerre mondiale, était exercée habituellement par des femmes ; il n'en va pas de même pour le mot « ingénieure », qui est apparu bien plus tard parce que ce métier a été longtemps réservé aux hommes.

Bien que les stéréotypes de genre présents dans le langage varient au fil des années dans le sens d'un rééquilibrage (Jones *et al.*, 2020), l'écart est encore présent. Caroline Criado Perez (2019) parle de décalage dans les données sexospécifiques (*gender data gap*), en faisant référence à l'absence d'une représentation adéquate des femmes dans les bases de données statistiques ; de manière similaire, Ricardo Baeza-Yates (2018) parle de biais algorithmique (*algorithmic bias*) en signalant que le bas pourcentage de femmes dans le secteur de la programmation par rapport à celui des hommes aurait des conséquences sur la conception des algorithmes.

À la fin de l'année 2018, Google a officiellement admis la présence des stéréotypes de genre au sein de ses applications automatiques⁴ (traduction, suggestion de rédaction dans *Gmail*, etc.) et a également publié un article sur les mesures adoptées pour résoudre cet écart⁵.

Si l'existence de stéréotypes de genre est donc prouvée, les méthodes utilisées pour les détecter peuvent être variées. Dans une étude sur le moteur de TA de Google, Marcelo Prates (Prates *et al.*, 2019) a sélectionné plus de 1 000 professions, une liste des adjectifs les plus fréquents selon le *Corpus of Contemporary American English* (COCA) et une structure syntaxique à double possibilité, sans ou avec l'adjectif. Les résultats montrent que le système traduit préférentiellement les pronoms personnels au masculin plutôt qu'au féminin et que la présence de certains adjectifs comme « *Arrogant, Cruel and Guilty* » (« arrogant, cruel et coupable », *ibidem* : 24) porte l'algorithme à privilégier la traduction au masculin, alors que d'autres adjectifs comme « *Shy, Attractive, Happy, Kind and Ashamed* » (« timide, séduisant, heureux, gentil, honteux/qui a honte », *idem*) facilitent la traduction au féminin, suggérant une représentation stéréotypée des genres. Stanovsky (Stanovsky, G. *et al.* 2019) parvient à la même conclusion en constatant que la traduction devient plus précise en présence d'une sorte d'amalgame entre certaines professions et le genre, comme dans « *female nurse* » (« infirmière »). De même, la structure syntaxique a fait l'objet d'études qui ont essayé d'identifier des éléments critiques ou de faciliter une traduction correcte. Costa Jussà (2019), qui propose d'intégrer des mots pour mitiger l'impact des stéréotypes de genre (*words embeddings*), constate que les noms propres contribuent à réduire l'ambiguïté. Enfin, Monti (2017) identifie des constructions comme les anaphores,

les appositions et l'accord sujet/objet avec le complément prédicatif qui seraient problématiques. Toutes ces études nous ont servi de point de départ théorique pour sélectionner une liste de noms de métiers que nous allons présenter dans la section suivante.

3. Méthodologie d'analyse : les noms de métiers et la structure syntaxique

En partant de la prémisse que les systèmes de TA génèrent des traductions erronées en termes de genre, dues entre autres à un manque de biodiversité dans les données d'entrée, nous allons nous intéresser au type d'erreurs de TA, dans la conviction que cette analyse pourrait éclairer la nature des stéréotypes présents. Notre objectif n'est pas d'évaluer la performance des logiciels de TA, mais de savoir si les erreurs sont simplement aléatoires, ou bien si elles sont révélatrices de stéréotypes provenant d'un imaginaire caché.

Tout en considérant les travaux précédemment mentionnés, nous avons tout d'abord sélectionné une liste de noms de métiers, extraite de celle du projet ESCO⁶ (*European Skills Competences and Occupations*), cette dernière comptant 2 942 noms de professions en 27 langues européennes. La liste ESCO s'appuie à son tour sur un code ISCO 08⁷ qui divise les professions en 10 groupes : Professions militaires ; Directeurs, cadres de direction et gérants ; Professions intellectuelles et scientifiques ; Professions intermédiaires ; Employés de type administratif ; Personnel des services directs aux particuliers, commerçants et vendeurs ; Agriculteurs et ouvriers qualifiés de l'agriculture, de la sylviculture et de la pêche ; Métiers qualifiés de l'industrie et de l'artisanat ; Conducteurs d'installations et de machines, et ouvriers de l'assemblage ; Professions élémentaires.

Nous avons sélectionné 73 noms de métiers équivalents en français et en italien (tableau 1), en suivant un critère de représentativité générale en termes de typologie ; nous avons également préféré retenir les seuls hyperonymes plutôt que de multiplier les hyponymes. Par exemple, en ce qui concerne le secteur médical, nous avons préféré le mot « docteur » aux hyponymes « dermatologue », « neurologue », « pharmacologue », « pneumologue », « psychiatre », « pédiatre », etc.

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
directeur	directrice	direttore	direttrice
editeur	editrice	editore	editrice
physicien	physicienne	fisico	fisica
docteur	docteure	dottore	dottoressa

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
infirmier	infirmière	infermiere	infermiera
professeur	professeure	professore	professoressa
programmeur d'applications	programmeuse d'applications	sviluppatore di applicazioni	sviluppatrice di applicazioni
avocat	avocate	avvocato	avvocata
écrivain	écrivaine	scrittore	scrittrice
traducteur	traductrice	traduttore	traduttrice
compositeur	compositrice	compositore	compositrice
danseur	danseuse	ballerino	ballerina
acteur	actrice	attore	attrice
technicien	technicienne	tecnico	tecnica
opticien	opticienne	ottico	ottica
décorateur	décoratrice	decoratore	decoratrice
cuisinier	cuisinière	cuoco	cuoca
employé	employée	impiegato	impiegata
serveur	serveuse	cameriere	cameriera
coiffeur	coiffeuse	parruchiere	parrucchiera
commerçant	commerçante	commerciante	commerciante
vendeur	vendeuse	commesso	commessa
pompier	pompière	vigile del fuoco	vigilessa del fuoco
agent de police	agente de police	agente di polizia	agente di polizia
agriculteur	agricultrice	agricoltore	agricoltrice
ouvrier	ouvrière	operaio	operaia
éleveur	éleveuse	allevatore	allevatrice
pêcheur	pêcheuse	pescatore	pescatrice
chasseur	chasseuse	cacciatore	cacciatrice
potier	potière	vasaio	vasaia
artisan	artisane	artigiano	artigiana
boulangier	boulangère	panettiere	panettiera
pâtissier	pâtissière	pasticcere	pasticcera
barman	barmaid	barman	barmaid
ingénieur	ingénieure	ingegnere	ingegnere
pharmacien	pharmacienne	farmacista	farmacista
esthéticien	esthéticienne	estetista	estetista
magistrat	magistrat	magistrato	magistrato

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
musicien	musicienne	musicista	musicista
chanteur	chanteuse	cantante	cantante
metteur en scène	metteuse en scène	regista	regista
chauffeur de taxi	chauffeuse de taxi	autista di taxi	autista di taxi
astronome	atronome	astronomo	astronoma
météorologue	météorologue	meteorologo	meteorologa
chimiste	chimiste	chimico	chimica
géologue	géologue	geologo	geologa
biologiste	biologiste	biologo	biologa
botaniste	botaniste	botanico	botanica
zoologiste	zoologiste	zoologo	zoologa
architecte	architecte	architetto	architetta
graphiste	graphiste	grafico	grafica
vétérinaire	vétérinaire	veterinario	veterinaria
bibliothécaire	bibliothécaire	bibliotecario	bibliotecaria
sociologue	sociologue	sociologo	sociologa
anthropologue	anthropologue	antropologo	antropologa
peintre	peintre	pittore	pittrice
psychologue	psychologue	psicologo	psicologa
chorégraphe	chorégraphe	coreografo	coreografa
photographe	photographe	fotografo	fotografa
secrétaire	Secrétaire	segretario	segretaria
concierge	concierge	custode d'immobile	custode d'immobile
orfèvre	orfèvre	orafo	orafa
cadre	cadre	dirigente	dirigente
urbaniste	urbaniste	urbanista	urbanista
médecin	médecin	medico	medico
physiothérapeute	physiothérapeute	fisioterapista	fisioterapista
comptable	comptable	contabile	contabile
économiste	économiste	economista	economista
journaliste	journaliste	giornalista	giornalista
interprète	interprète	interprete	interprete
pilote d'avion	Pilote d'avion	pilota di aeromobili	pilota di aeromobili
athlète	athlète	atleta	atleta

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
conducteur d'autobus et de tramways	conductrice d'autobus et de tramways	conducente di autobus e tram	conducente di autobus e tram

Tableau 1 : Liste des noms de métiers

En nous appuyant sur les études de TA mentionnées précédemment, nous avons inséré ces noms dans des phrases dont la structure syntaxique est la suivante : article indéfini, nom du métier, adverbe + adjectif, passé composé (auxiliaire + participe passé), complément de lieu. Cette structure fournit de nombreux indices d'accord de genre même dans le cas d'une profession ayant un nom épïcène (comme « secrétaire » en français ou « *commerciant*e » en italien). Pour ce qui concerne les adjectifs, nous avons distingué l'adjectif d'apparence, tels « beau/belle » en français ou les équivalents italiens « *bello/bella* », de l'adjectif de compétence, comme « intelligent/intelligente » en français ou les équivalents italiens « *intelligente/intelligente* ». Nous avons sélectionné les 500 adjectifs les plus utilisés dans les corpus frWac et itWac (M. Baroni *et al.*, 2008) et avons construit une base de données d'un total de 584 phrases (en tableau 2).

		TOT
apparence m/f (fr)	<i>Un <PROFESSION> très <ADJ. APPARENCE> est entré dans la salle</i> <i>Une <PROFESSION> très <ADJ. APPARENCE> est entrée dans la salle</i>	146 phrases (m/f)
apparence m/f (it)	<i>Un <PROFESSION> molto <ADJ. APPARENCE> è entrato nella stanza</i> <i>Una <PROFESSION> molto <ADJ. APPARENCE> è entrata nella stanza</i>	146 phrases (m/f)
compé- tence m/f (fr)	<i>Un <PROFESSION> très <ADJ. COMPETENCE> est entré dans la salle</i> <i>Une <PROFESSION> très <ADJ. COMPETENCE> est entrée dans la salle</i>	146 phrases (m/f)
compé- tence m/f (it)	<i>Un <PROFESSION> molto <ADJ. COMPETENCE> è entrato nella stanza</i> <i>Una <PROFESSION> molto <ADJ. COMPETENCE> è entrata nella stanza</i>	146 phrases (m/f)
		584 phrases

Tableau 2 : Composition des bases de données.

Ensuite, nous avons fait traduire ces phrases du français vers l'italien ou vice-versa par trois moteurs de TA neuronales : *Google Translate* (GT), *Microsoft Translator* (MT), *DeepL* (DT). Les résultats obtenus sont analysés dans la section suivante.

4. Analyses et résultats

Après avoir fait élaborer les phrases par les moteurs, nous avons analysé manuellement les résultats obtenus, en les annotant selon leur exactitude. Elles ont été jugées « fausses » en ne considérant que le nom de métier, c'est-à-dire que, même si d'autres éléments grammaticaux (accord du participe, article, etc.) ont été correctement traduits au féminin, le nom de métier ne l'était pas. Dans ce cas, la phrase a été annotée comme « incorrecte » puisque le concept a été bien traduit mais que le nom n'a pas été accordé au genre de manière correcte. Dans le cas des noms de métiers épiciènes, l'erreur n'est considérée que lorsque la totalité de la phrase est traduite avec l'accord de l'autre genre par rapport à l'original ou lorsque le nom de métier est traduit par un autre mot.

Comme on peut l'observer dans les tableaux 3 et 4, les taux de traductions correctes des phrases au masculin de la langue source (qu'elle soit en français ou en italien) ont un pourcentage qui va de 93 % à 97% et les taux des traductions correctes pour le féminin oscillent entre 8% et 86%. Plus précisément le féminin ne présente que deux sous-groupes qui ont un taux de réponses correctes assez élevé tandis que tous les autres ont un niveau très bas. Le taux de réponses correctes plus élevé au masculin par rapport au féminin confirme la présence d'un déséquilibre en faveur du masculin, comme l'indiquaient aussi les études mentionnées.

Moteur de TA	masculin			féminin		
	GT	DT	MT	GT	DT	MT
fr>it	96%	96%	97%	33%	86%	11%
it>fr	96%	95%	96%	21%	75%	16%

Tableau 3 : Pourcentages totaux des traductions correctes avec l'adjectif d'apparence, différenciés par moteurs de traduction

Moteur de TA	masculin			féminin		
	GT	DT	MT	GT	DT	MT
fr>it	96%	96%	93%	11%	9%	7%
it>fr	96%	95%	97%	18%	8%	11%

Tableau 4 : Pourcentages totaux des traductions correctes avec l'adjectif de compétence, différenciés par moteurs de traduction

Nous observons encore que le résultat lié au sous-groupe des noms de métiers au féminin montre que la féminisation est moins fréquente dans des phrases avec l'adjectif de compétence (de 7% à 18%) que dans les phrases avec l'adjectif de

l'apparence (11%-86%), ce qui confirme l'hypothèse de départ selon laquelle deux adjectifs appartenant à deux champs sémantiques différents influencent l'exactitude des traductions.

En revanche, la présence d'adjectifs d'apparence et de compétence n'affecte nullement les performances de traduction dans le cas du masculin et les noms des métiers sur lesquels se concentrent les erreurs sont toujours les mêmes : « *infermiere*/infirmier » est traduit à 100% au féminin, suivi de « *estetista*/esthéticien » avec un taux d'erreur de 83 %, de « *segretario*/secrétaire » avec un taux d'erreur de 75%. Certains cas qui semblent isolés comme « potier », « orfèvre », « *ballerino* », « *coreografo* » présentent un taux d'erreur de 8%.

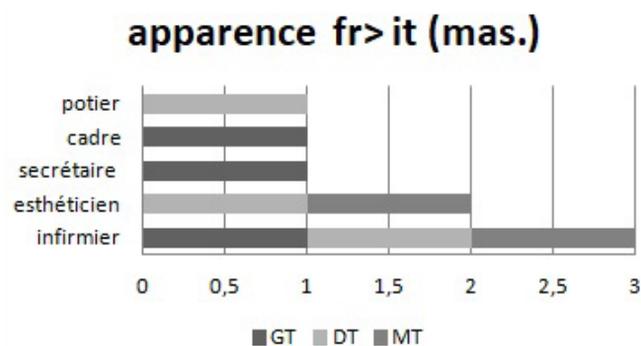


Tableau 5 : Pourcentage des traductions erronées du français vers l'italien pour les noms masculins avec l'adjectif d'apparence

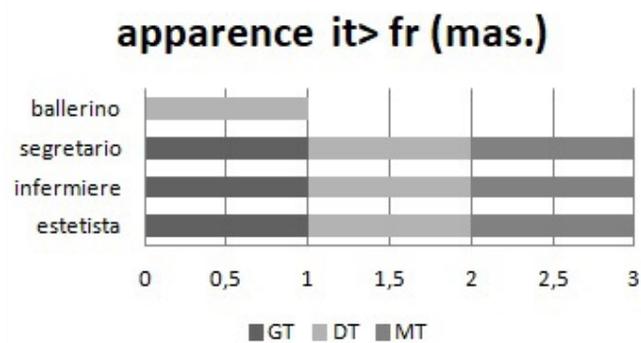


Tableau 6 : Pourcentage des traductions erronées de l'italien vers le français pour les noms masculins avec l'adjectif d'apparence

competence fr>it (mas.)

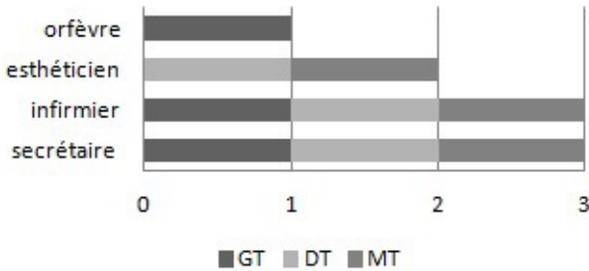
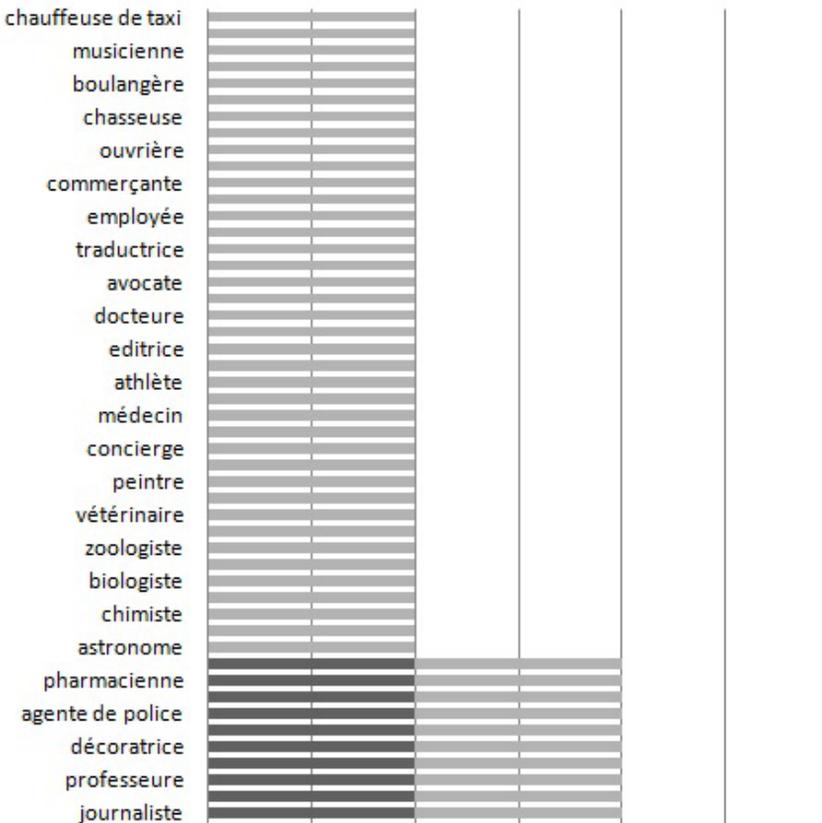


Tableau 7 : Pourcentage des traductions erronées du français vers l’italien pour les noms masculins avec l’adjectif de compétence

adjectif apparence fr > it (fém.)



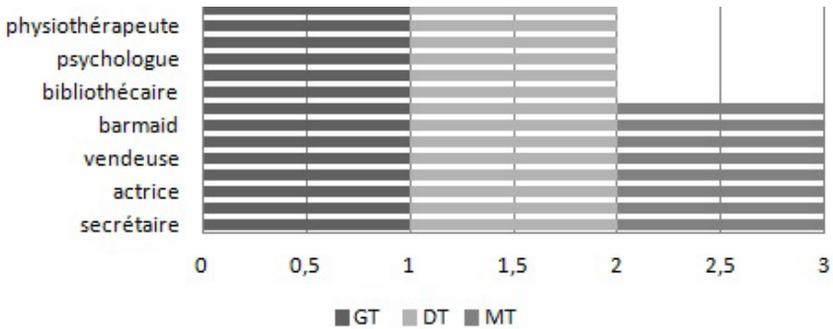


Tableau 8 : Pourcentage des traductions erronées de l’italien vers le français pour les noms masculins avec l’adjectif de compétence

En analysant la typologie d’erreurs, nous observons qu’il s’agit là de noms de métiers traduits au féminin. Nous trouvons donc, en correspondance de « Un infirmier très beau est entré dans la salle », la traduction « *Un’infermiera molto bella è entrata nella stanza* ». Dans ces cas, le nom du métier l’emporte sur la structure syntaxique, qui pourtant possède plusieurs éléments comme les accords ou les adjectifs pouvant conduire à une interprétation correcte de la part du moteur de traduction automatique. La traduction est donc forcée au féminin, et, quelle que soit la langue source, les trois moteurs de traductions montrent toujours les mêmes noms de métiers ayant le plus haut taux d’erreurs.

Dans le cas du masculin, nous avons vu que les noms de métiers traduits de manière erronée par les trois moteurs de traduction sont toujours les mêmes, que ce soit de l’italien vers le français ou du français vers l’italien. En comparant les mêmes noms de métiers avec les résultats des traductions des phrases au féminin, nous observons qu’ils donnent lieu aux rares phrases qui sont traduites correctement.

Comme on peut le voir dans les tableaux 9, 10, 11, 12, « *estetista/esthéticienne* », « *segretario/secrétaire* », « *infermiere/infermiera* » ont des taux de réponses correctes très élevés.



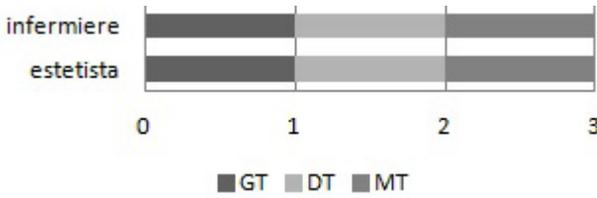
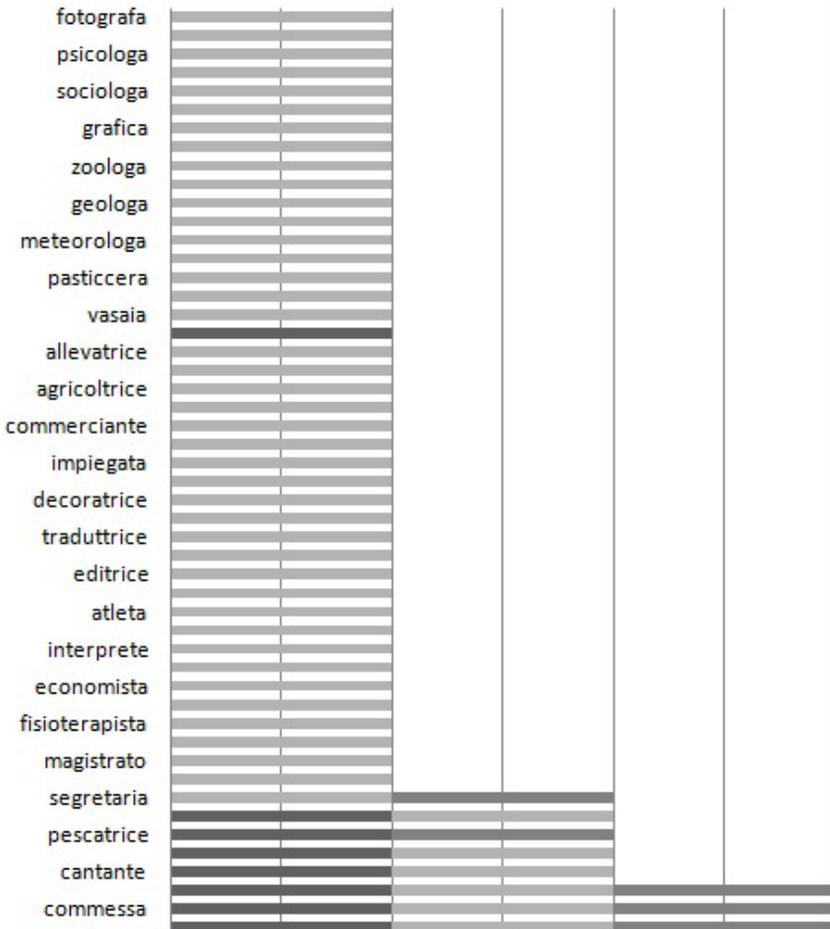


Tableau 9 : Pourcentage des traductions correctes du français vers l’italien pour les noms féminins avec l’adjectif d’apparence

adjectif apparence it > fr (fém.)



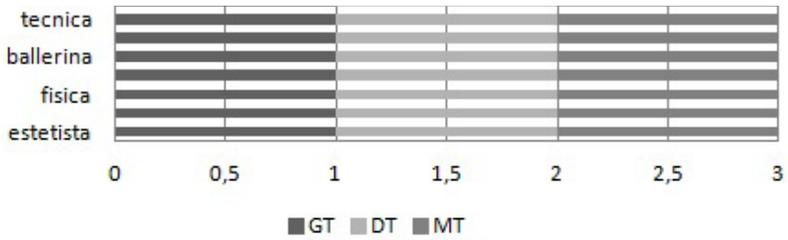


Tableau 10 : Pourcentage des traductions correctes de l'italien vers le français pour les noms féminins avec l'adjectif d'apparence

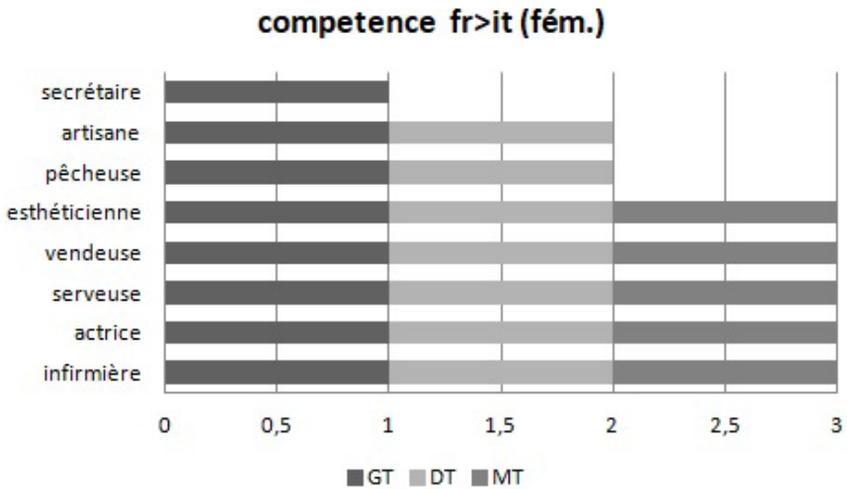
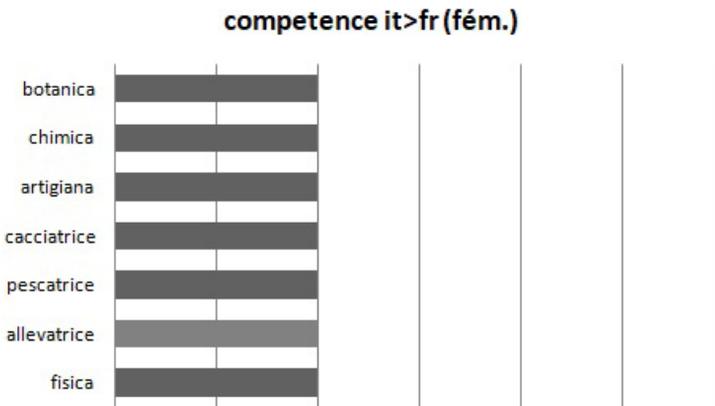


Tableau 11 : Pourcentage des traductions correctes du français vers l'italien pour les noms féminins avec l'adjectif de compétence



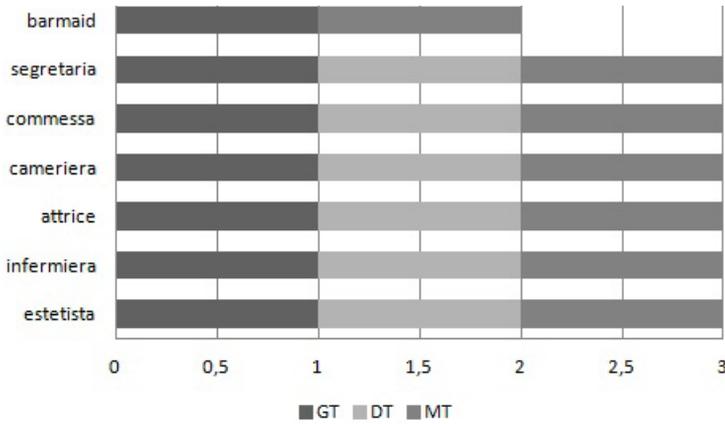


Tableau 12 : Pourcentage des traductions correctes de l’italien vers le français pour les noms féminins avec l’adjectif de compétence

Le choix des mots n’est pas aléatoire : en effet, en comparant les noms de métiers traduits de manière incorrecte au masculin et de manière correcte au féminin, quelle que ce soit la langue source, les mots qui ont un taux très élevé d’erreurs ou de réponses correctes sont toujours les mêmes. De plus, on peut regrouper les noms de métiers qui sont traduits correctement au féminin en utilisant les 10 catégories de la taxonomie du projet ESCO qui, comme nous l’avons dit auparavant, s’inspire de la classification ISCO-08 : nous avons pu ainsi identifier que le secteur ayant une couverture majeure est celui des « Professions intellectuelles et scientifiques » (tableau 13).

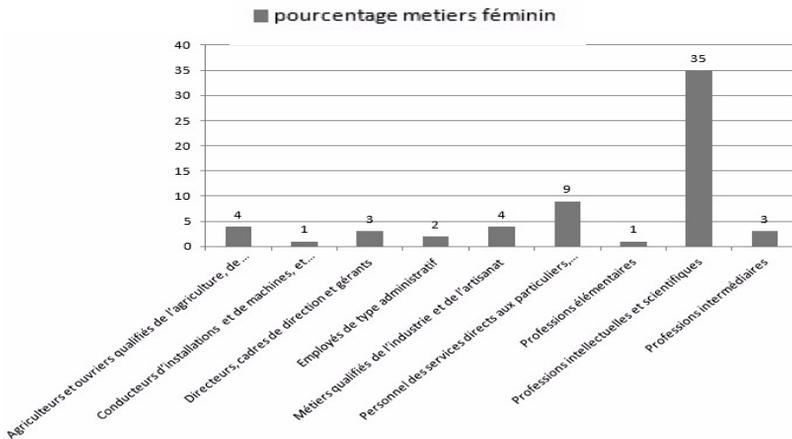


Tableau 13 : Couverture des secteurs professionnels des noms de métiers au féminin selon les catégories ESCO qui sont traduits correctement

Conclusions

L'hypothèse d'une influence des adjectifs est partiellement confirmée : les taux de résultats corrects sont clairement déséquilibrés en faveur des noms au masculin pour lesquelles la présence des adjectifs de compétence et d'apparence ne semblent jouer aucun rôle lors de la traduction. En revanche, pour les noms au féminin le taux d'erreur est bien plus élevé et le rôle des adjectifs d'apparence et de compétence semble jouer un rôle déterminant. Cela impliquerait que dans le cas d'une performance déjà basse - due aussi à une biodiversité insuffisante des données dans le corpus d'entraînement de l'algorithme d'intelligence artificielle - la présence d'adjectifs liés à l'apparence et à la compétence rend la représentation encore moins stable en raison de l'existence d'un stéréotype.

Des études futures en ce domaine pourraient contribuer à approfondir cette question, en élargissant la liste des adjectifs pour couvrir un spectre sémantique plus ample.

Bibliographie

- Baeza-Yates, R. 2018. « Bias on the web ». *Communications of the ACM*, n° 61 (6), p. 54-61.
- Baroni M. et al. 2008. *The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora*. Pays-Bas : Kluwer Academic Publishers.
- Berger, A.-E. 2019, « Le genre de la traduction : introduction ». *Revue De Genere*, n° 5, p. I-XII.
- Cardon, D. et al. 2018. « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle ». *Réseaux*, n° 211 (5), p. 173-220.
- Costa-Jussà, M. R. 2019. « An analysis of gender bias studies in natural language processing ». *Nature Machine Intelligence*, n° 1(11), p. 495-496.
- Criado-Perez, C. 2019. *Invisible Women: Exposing Data Bias in a World Designed for Men*. Londres : Chatto & Windus.
- Eco, U. 1964, *Apocalittici e integrati*. Milan : Bompiani.
- Eco, U. 2003, *Dire presque la même chose*. Traduction de M Bouzaher. Paris : Grasset [or. *Dire quasi la stessa cosa*, Milan : Bompiani].
- Federici, E. (dir.) 2011. *Translating Gender*. Berne : Peter Lang.
- Greco L. 2014. « Les recherches linguistiques sur le genre : un état de l'art ». *Langage et société*, n° 148 (2), p.11-29.
- Jones, J. J. et al. 2020. « Stereotypical Gender Associations in Language Have Decreased Over Time ». *Sociological Science*, n° 7, p. 1-35.
- Loffler-Laurian, A-M. 1996, *La traduction automatique*, Presses universitaires du Septentrion : Villeneuve d'Ascq.
- Monti, J. 2017. « Questioni di genere in traduzione automatica ». In : De Meo, A., Di Pace, L., Manco, A. *Al femminile, scritti linguistici in onore di Cristina Vallini*. Florence : Franco Cesati editore, p. 411-431.
- Morel, H. 2020 « Au-delà de la langue, les femmes invisibles ». *Cahiers d'histoire. Revue d'histoire critique*, p. 147-148.
- Nida, E., Taber, C.R. 1969. *The Theory and Practice of Translation*. Leiden : E.J. Brill.

- Prates, M. *et al.* 2019. « Assessing Gender Bias in Machine Translation - A Case Study with Google Translate ». *Neural Computing and Application*. Londres : Springer. [En ligne] : <https://arxiv.org/pdf/1809.02208.pdf> [consulté le 07 février 2021].
- Nirenburg, S., *et al.* 2003. *Readings in Machine Translation*. Cambridge, Massachussets : The MIT Press.
- Paresh, D. 2018. « Fearful of bias, Google blocks gender-based pronouns from new AI tool ». [En ligne] : <https://news.trust.org/item/20181127055628-022s5> [consulté le 07 février 2021].
- Stanovsky, G. *et al.* 2019. « Evaluating Gender Bias in Machine Translation ». *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Firenze.
- Viennot, E. 2014. *Non, le masculin ne l'emporte pas sur le féminin*. Paris : Editions iXe.
- Von Flotow, L. 2011. *Gender in Translation*. Amsterdam-Philadelphia : John Benjamins Publishing Company.
- Weaver, W. 1949. *Memorandum*, New York : The Rockefeller Foundation.
- Wilks, Y. 2009. *Machine Translation, Its Scope and Limits*. Londres : Springer.

Sitographie⁸

DeepL

<https://www.deepl.com/translator>

ESCO - European Skills Competences and Occupations

<https://ec.europa.eu/esco/portal>

Google Translate (GT)

<https://translate.google.com>

Microsoft Translator (MT) : <https://www.microsoft.com/it-it/translator>

Notes

1. Un des principaux supporteurs de cette approche et aussi fondateur de la moderne « Science de la traduction » fut le linguiste américain Eugene Nida (1969).
2. Voir, entre autres, Luise Von Flotow, L. (2011) et Anne-Emmanuelle Berger (2019).
3. Le langage épïcène ou rédaction épïcène vise à éviter le sexisme du langage, grâce à l'usage de techniques de reformulation lexicale, syntaxique, l'attention à l'accord morphologique et à la typographie. Le langage épïcène permet d'éviter la discrimination de genre et rejette l'idée d'un androcentrisme marqué par l'utilisation du masculin comme universel générique. À ce sujet, voir, entre autres, Éliane Viennot (2014) et Luca Greco (2014).
4. <https://blog.google/products/translate/reducing-gender-bias-google-translate>
5. <https://ai.googleblog.com/2018/12/providing-gender-specific-translations.html> Voir aussi Dave Paresh (2018).
6. ESCO est un projet de la Commission Européenne qui organise une ontologie multilingue en trois piliers : professions, compétences et certifications.
7. À chaque métier équivaut un code ISCO 08 qui est superposable à la catégorie ESCO. Voir : <https://ec.europa.eu/esco/portal/occupation?resetLanguage=true&newLanguage=fr>
8. Date de dernière consultation des sites : 7 février 2021.



ISSN 1724-0700

ISSN en ligne 2260-8087

Les écarts dénominatifs dans la normative européenne sur la protection de données en matière pénale : une étude entre traduction institutionnelle, dispositifs terminologiques et traduction automatique

Francesca Bisiani

Université Catholique de Lille, France

francesca.bisiani@univ-catholille.fr

<https://orcid.org/0000-0003-0204-4130>

Reçu le 10-02-2021 / Évalué le 08-03-2021 / Accepté le 15-04-2021

Résumé

Cette étude se propose d'étudier certains équivalents plurivoques en plusieurs langues qui ressortent des textes contraignants européens sur le traitement de données en matière pénale, selon une approche discursive de la terminologie. Nous entendons d'abord identifier les divergences des points de vue des énonciateurs de plusieurs langues et, à partir de ce corpus, mettre en évidence le rôle des dispositifs terminologiques et de traduction automatique dans la diffusion et dans la mise en circulation de la terminologie institutionnelle.

Mots-clés : terminologie, traduction automatique, traduction, analyse du discours, intelligence artificielle

I disallineamenti terminologici nella normativa europea sulla protezione dei dati personali in ambito penale: uno studio fra traduzione istituzionale, dispositivi terminologici e traduzione automatica

Riassunto

L'articolo si propone di analizzare, tramite un approccio discorsivo della terminologia, alcuni disallineamenti terminologici in francese, italiano, inglese e spagnolo che emergono dai testi vincolanti europei in materia di protezione dei dati personali in ambito penale. Il nostro contributo intende identificare innanzitutto le divergenze che caratterizzano i punti di vista degli enunciatori in più lingue, per poi proporre una riflessione sul ruolo dei dispositivi terminologici e di traduzione automatica nella diffusione e circolazione della terminologia istituzionale.

Parole chiave: terminologia, traduzione automatica, traduzione, analisi del discorso, intelligenza artificiale

Terminological discrepancies in European legislation on data protection in criminal matters: a study between institutional translation, terminology and machine translation devices

Abstract

The present study, based on a discursive approach to terminology, aims to analyse multilingual equivalents in several languages that emerge from the European binding

texts on the protection of personal data in the field of criminal matters. We first intend to identify the different points of view of enunciators in several languages and, on the basis of this, to highlight the role of terminology and machine translation devices in the dissemination and circulation of institutional terminology.

Keywords: terminology, automatic translation, translation, discourse analysis, artificial intelligence

Introduction

L'Union européenne (UE) a fait du multilinguisme l'un de ses principes fondateurs. Toutefois, si la diversité linguistique, en tant que point d'ancrage de l'UE, représente une richesse pour l'espace européen, elle peut se révéler également une « arme à double tranchant¹ » (Cinato, 2010 : 86). Les difficultés liées à l'aménagement linguistique et terminologique des concepts du droit européen dans les différentes langues-cultures révèlent les éléments de complexité du multilinguisme et invitent à s'interroger sur les écarts interprétatifs qui se produisent lorsque les termes et le discours circulent dans l'espace européen.

À partir de ces considérations, notre recherche se propose d'examiner, selon une approche discursive de la terminologie (Raus, 2013 ; Bisiani, 2020), les variantes dénominatives qui se manifestent dans les versions linguistiques en français, italien, anglais et espagnol, des actes contraignants européens concernant le traitement des données personnelles en matière pénale (2016-2019). Nous avançons l'hypothèse que l'étude des équivalents dans un contexte multilingue permet de mettre en évidence des divergences parmi les versions linguistiques des textes contraignants de l'UE qui méritent d'être évaluées sous une perspective juridico-politique. Notre réflexion souhaite également mettre en avant le rôle central des dispositifs terminologiques et de traduction automatique non seulement comme domaine d'observation, mais aussi en tant que « relayeur[s] interlinguistique[s] » (Guidère, 2009 : 48) de la mise en circulation du discours sur la sécurité.

Après avoir introduit le corpus d'étude dans la première partie, nous analyserons selon une démarche quantitative et qualitative à la fois, les désalignements que nous avons remarqués dans les versions linguistiques des documents étudiés. Nous passerons ensuite à la comparaison des résultats obtenus concernant la base de données terminologique de l'UE (*IATE*), et dans deux outils de traduction automatique en libre accès (*DeepL* et *Google Translate*), ce qui nous amènera, avant de passer aux conclusions, à formuler quelques observations sur le rapport entre terminologie, traduction spécialisée et traduction automatique.

1. « Protection des données personnelles » en matière pénale comme observatoire d'étude et la méthode d'analyse

Les textes européens sur le traitement de données en matière pénale sont le point de départ empirique de notre étude. La source de ce choix nous a été suggérée par l'intention d'aborder une terminologie exprimant des débats publics récents sur la protection des données personnelles dans le domaine du renseignement. Les techniques de collecte d'informations à des fins policières et sécuritaires ont été profondément bouleversées par l'évolution de la technologie et par l'émergence de la question éthique de l'usage des données dans le débat public. Comme l'expliquent Chopin et Oudet (2019 : 222), les implications morales des pratiques de renseignement, étudiées traditionnellement par les historiens, ont « longtemps été secondaires » dans la littérature sur ce sujet. Or, au cours des dernières années, cette question est devenue centrale dans les démocraties occidentales à tel point que la régulation des techniques de renseignement représente un critère de légitimité démocratique pour les politiques actuelles (Chopin, Oudet 2019 : 223).

C'est donc à partir de cet arrière-plan que nous avons décidé de recueillir les versions linguistiques des textes-cadre sur l'utilisation et la protection de données en matière pénale et sécuritaire², à savoir la décision-cadre 2008/977/JAI, la directive 2016/680 et la directive 2016/681.

Dans ces actes, que nous allons examiner de manière contrastive en français, italien, anglais et espagnol, notre attention se concentrera sur les termes exprimant l'objet de la démarche juridique, c'est-à-dire le champ d'application de la décision-cadre et des directives concernées. En effet, dans les trois documents en français nous retrouvons une séquence identique, et qui limite les finalités pour lesquelles les dispositions peuvent être appliquées. Dans les trois documents, cette séquence est composée des termes suivants : « prévention », « détection », « enquêtes » « poursuites », comme nous pouvons l'observer dans les extraits suivants :

La présente décision-cadre ne s'applique qu'aux données collectées ou traitées par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales.

(Décision-cadre 2008/977/JAI du 27 novembre 2008)³

Pour être licite, le traitement des données à caractère personnel au titre de la présente directive devrait être nécessaire à l'exécution d'une mission d'intérêt général par une autorité compétente (...) à des fins de prévention et

de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales (...).

(Directive 2016/680 du 27 avril 2016)

Les données PNR recueillies conformément à la présente directive ne peuvent être traitées qu'à des fins de prévention et de détection des infractions terroristes et des formes graves de criminalité ainsi que d'enquêtes et de poursuites en la matière.

(Directive 2016/681 du 27 avril 2016)

Nous voyons d'emblée que la séquence « prévention + détection + enquêtes + poursuites » de la décision-cadre de 2008 est reprise dans les deux directives de 2016. Ce constat nous invite à étudier les équivalents des termes dans les autres versions linguistiques afin d'analyser les éventuels désalignements conceptuels (Raus, 2013 :18). Il faut considérer que l'objectif de ces textes est d'harmoniser les règles des États membres tout en sachant que cet objectif se heurte, comme l'explique l'Agence des droits fondamentaux de l'Union européenne (2017 : 9), avec les législations nationales qui, en matière de renseignement, demeurent encore différentes et complexes. L'observation des désalignements nous permet donc de mettre en évidence les écarts de signification et de considérer d'une part les différentes conceptualisations qui risquent d'entraver l'harmonisation de la législation, d'autre part la dimension politique qui entoure le choix d'une dénomination (Guidère, 2008 : 90).

Ce type de démarche s'inscrit dans les approches discursives (Raus,2013; Bisiani, 2020) de la terminologie et invite à considérer les décalages conceptuels qui ressortent de la traduction comme des symptômes de discordances idéologiques. Il s'agit donc d'étudier les termes, à savoir les mots engagés dans un domaine spécialisé (L'Homme, 2004 : 22) et de les évaluer par rapport au contexte de l'énoncé et de l'usage par le biais de notions de l'analyse du discours à la française (Dufour, Rosier, 2012). La valeur qu'acquiert une unité lexicale dépend donc de l'instance énonciative et de son usage dans un discours donné. Cette approche permet également d'évaluer la capacité de la traduction et des outils terminologiques à doubler l'instance énonciative (Raus, 2013 : 84-86), c'est-à-dire à relayer ou à estomper un type de discours produit par un énonciateur. En effet, les choix dénominatifs, dans la procédure de traduction, qu'elle soit automatique ou humaine, participent au foisonnement de la terminologie et à la diffusion des énoncés produits sur le plan européen.

C'est dans cette optique que nous avons voulu combiner la recherche des termes dans les documents européens avec la comparaison des équivalents dans la base de

données principale de l'UE IATE et dans certains outils de traduction automatique en libre accès. Pour l'exploration du corpus, nous avons utilisé le logiciel *Sketch Engine* qui permet d'afficher les cooccurrences et de repérer les fréquences de la séquence étudiée.

Nous allons présenter les résultats de l'analyse quantitative avant de passer à une analyse qualitative des résultats et à l'observation des termes étudiés dans les dispositifs terminologiques et de traduction.

2. L'analyse contrastive et l'émergence des désalignements

La première étape de la recherche quantitative nous a permis de retracer la présence de la structure « prévention + détection + enquêtes + poursuites » dans les documents en français pour ensuite comparer les résultats avec les autres versions linguistiques. Nous soulignons toutefois que notre analyse quantitative est limitée au repérage de la séquence étudiée dans les documents en français et de ses équivalents dans les autres versions linguistiques. Nous privilégions en effet une démarche qualitative qui s'intéresse aux relations sémantiques qu'entretiennent les termes entre eux, au contexte linguistique et aux reprises intertextuelles et qui invite à évaluer les variantes dénominatives en tant qu'indices de positionnements différents des énonciateurs.

Pour attester la présence ainsi que la stabilité de la séquence « prévention + détection + enquêtes + poursuites » dans notre corpus, nous avons lancé une recherche de la cooccurrence du lemme « prévention » dans *Sketch Engine*. Les résultats nous ont permis d'attester que la séquence étudiée est relativement stable. Le lemme « prévention » apparaît 77 fois dont 38 fois dans la structure étudiée. Les termes apparaissent dans la totalité des occurrences dans le même ordre : « prévention », « détection », « enquêtes » et « poursuites ». Selon une perspective terminologique, ils semblent entamer une relation ontologique d'enchaînement à l'intérieur du même champ conceptuel. Cette relation repose donc sur une « succession d'événements dans le temps » (Cabré, 1998 [1992] : 179) qui est donc séquentielle. Chaque terme désignant une action marque l'étape d'un processus qui prévoit une approche allant d'une stratégie préventive à une réponse réactive.

Or, nous avons procédé au même type de recherches dans les corpus italien, anglais et espagnol. Les résultats nous montrent d'emblée que la stabilité attestée pour la version française n'est pas préservée dans les autres versions linguistiques. En d'autres termes, nous avons trouvé des différences dans la structure et dans la séquence des termes désignant la finalité du traitement. Nous rapportons dans le tableau 1 les structures trouvées et leur fréquence après avoir lancé une recherche des équivalents du lemme « prévention » en italien, anglais et espagnol :

Italien	Anglais	Espagnol
Prevenzione, indagine, accertamento e/o perseguimento (36 occurrences) Prevenzione, accertamento e indagine e/o azione penale (12 occurrences)	Prevention, investigation, detection or/and prosecution (31 occurrences) Prevention, detection investigation or/and prosecution (11 occurrences)	Prevención, investigación, detección y/o el enjuiciamiento (32 occurrences) Prevención, detección, investigación y/o enjuiciamiento (15 occurrences)

Tableau 1 : Cooccurrences et fréquences des équivalents du lemme « prévention ».
Source : *Sketch Engine*

Il faut remarquer que, contrairement à ce que nous avons repéré dans les documents en français, il existe deux séquences terminologiques différentes pour chaque langue. Ce basculement de l'ordre séquentiel est imputable aux équivalents du terme « détection », à savoir « *accertamento* » pour l'italien, « *detection* » pour l'anglais et « *detección* » pour l'espagnol qui se trouvent dans certains cas après le terme « prévention » et dans d'autres cas après le terme « enquêtes ». Il s'avère que l'action désignant la « détection » n'est pas clairement située à l'intérieur de la relation ontologique et nous ne pouvons donc pas savoir s'il s'agit d'une action menée en amont ou après une infraction. Si nous insistons sur l'ordre séquentiel des termes, c'est parce que les concepts terminologiques ne sont pas des « unités isolées » (Cabré, 1998 [1992] : 174). Les termes entretiennent entre eux des relations qui peuvent fournir des informations sur leur conceptualisation.

Quant au terme « détection » dans le corpus français, une étude intertextuelle des documents étudiés nous a amené à supposer que ce concept désigne une action qui s'inscrit dans une approche de type préventive. Nous avons cherché la définition de ce terme dans les documents du corpus et dans les documents préparatoires. Bien que nous n'ayons pas trouvé de définitions, nous constatons que la Commission européenne utilise l'expression « prévention et détection » pour décrire les actions à entreprendre afin d'évaluer la situation avant que l'infraction ait été commise :

Premièrement, les données PNR permettent aux services répressifs d'identifier des personnes auparavant « inconnues » d'eux, c'est-à-dire jusque-là soupçonnées de participation à une infraction grave ou à un acte de terrorisme (...). L'identification de ces personnes aide les services répressifs à prévenir et à détecter les infractions graves, y compris les actes de terrorisme. À cet effet, lesdits services doivent utiliser les données PNR, d'une part, en temps réel, pour les analyser au regard de critères d'évaluation préétablis, qui indiquent les personnes jusque-là « inconnues » devant faire l'objet d'un examen

approfondi et, d'autre part, d'une manière proactive aux fins de l'analyse et de la définition de critères d'évaluation.

(Proposition de directive 2016/681)

En résumé, il ressort de l'analyse contrastive qu'il existe des interprétations différentes de l'action de « détecter » une infraction. Par exemple, dans la séquence en italien « *prevenzione, indagine, accertamento e/o perseguimento* », nous voyons que le terme « *accertamento* », contrairement au français, est placé 36 fois après le terme « *indagine* » (« enquête ») et donc dans un espace d'action réactive plutôt que préventive. Ces divergences découlent sans doute du rapport colingue, à savoir le rapport et les interférences entre langues instituées (Branca-Rosoff, 2001 : 5, citée par Raus, 2013 : 24) et surtout de la perméabilité entre l'anglais et les autres langues officielles. Il est donc possible que les documents des directives en espagnol et en italien aient été traduits à partir de l'anglais⁴. Nous remarquons en tout cas que le français affiche toujours le même ordre séquentiel.

Ces désalignements nous invitent à examiner les variantes en termes de positionnements aussi. Les énonciateurs semblent révéler une volonté d'encadrer l'action de « détecter » dans des approches différentes. Dans les documents français, le terme désignerait une démarche de type proactif tandis que dans la plupart des séquences extraites des documents en italien, anglais et espagnol, il s'agirait plutôt d'une action à entreprendre après l'« enquête ».

Ces hésitations sur le concept de « détection » nous invitent à poursuivre notre enquête par l'étude de ce terme dans les dispositifs terminologiques et de traduction. Notre objectif consiste tout d'abord à cerner la définition de « détection » dans la base de données terminologique *IATE*, ensuite à mieux comprendre le rôle des outils de traduction automatique dans la diffusion des variantes.

3. La « détection » dans les dispositifs terminologiques et de traduction automatique

Nous allons observer à présent les définitions du terme « détection » dans la banque de données multilingue de l'UE *IATE* pour passer ensuite à l'analyse de ce terme dans la traduction automatique. Nous affichons dans le tableau 2 les définitions et les sources repérées dans *IATE* dans le contexte des infractions pénales tout en indiquant que nous n'avons pas trouvé de définitions en langue espagnole :

	Français	Italien	Anglais
Dénomination	Détection	Accertamento/ individuazione	Detection
Définition	action de déceler l'existence d'un phénomène caché <i>Source</i> Le Petit Robert de la langue française, 2020	azione volta a scoprire e delineare gli elementi essenziali di un fatto penalmente rilevante e dei soggetti che ne sono penalmente responsabili <i>Source</i> Luigi Delpino e Rocco Pezzano, Manuale di diritto penale - Parte generale, XXX edizione, Edizioni giuridiche Simone, 2020	action of finding or making a discovery, finding a criminal and or revealing any facts that have been hidden <i>Source</i> The Law Dictionary -What is detection? (7.7.2020)

Tableau 2 : Entrée du terme « détection » et de ses équivalents. Source : *IATE*

À la lumière de ces données, il est intéressant de formuler trois types d'observations : premièrement les définitions en italien et en anglais sont terminologiques, elles décrivent donc la notion par référence exclusive au domaine de spécialité (Cabré, 1998 [1992] : 182). Il s'avère que seule l'entrée française présente une définition tirée d'un dictionnaire de la langue générale qui pourrait donc s'adapter à d'autres contextes, notamment scientifique et informatique. Deuxièmement, la définition française ne contient pas dans ces caractères un nom animé, alors que les autres entrées inscrivent dans leurs traits des noms qui dénotent des êtres vivants tels que « *soggetti* », « *criminal* ». Troisièmement, les définitions en italien et en anglais nous fournissent des indications sur la « temporalité » de la dénomination, autrement dit sur l'espace d'action des pratiques de détection. Dans ces deux langues, la « détection » se réfère à des pratiques qui visent à découvrir une infraction ou un infracteur qui peuvent être punis pénalement, ce qui inscrit ce terme dans une approche réactive plutôt que préventive.

En résumé, nous constatons que les désalignements que nous avons trouvés dans notre corpus se réitèrent dans *IATE*. Les choix dénommatifs ainsi que les définitions en anglais et en italien insistent sur le trait humain et sur l'action de découvrir un fait pénal. Ces entrées ne semblent pas prendre en considération la possibilité d'identifier une personne avant qu'un fait délictueux ait été commis. Cela pourrait

révéler une certaine prudence ou résistance de l'énonciateur quant aux termes désignant des actions à mener en amont de l'infraction. En revanche, la définition française ne se limite pas au domaine pénal, elle propose une définition qui vient de la langue générale ou de la terminologie scientifico-informatique. Le fait de ne pas expliciter le trait humain et le domaine laisse la possibilité d'associer aux personnes un terme qui, traditionnellement, ne s'adapte pas aux noms animés, ni ne précise s'il s'agit d'une action pénale à mener avant ou après l'infraction. Nous allons enfin comparer l'ensemble de nos résultats dans deux dispositifs de traduction automatique en libre accès : *Google Translate* et *DeepL*.

Nous avons d'abord concentré notre analyse sur les équivalents de l'expression « prévention, détection des infractions pénales, enquêtes et poursuites ». Nous rapportons les données dans le tableau 3 :

	Italien	Anglais	Espagnol
<i>DeepL</i>	prevenzione, <i>individuazione</i> , investigazione e perseguimento dei reati penali	prevention, detection, investigation and prosecution of criminal offences	prevención, detección, investigación y persecución de delitos
<i>Google Translate</i>	prevenzione, <i>accertamento</i> di reati, indagini e azioni penali	prevention, detection of criminal offenses, investigation and prosecution	prevención, detección de delitos, investigación y enjuiciamiento

Tableau 3 : Traduction automatique (FR>IT>EN>ES) de l'expression « prévention, détection des infractions pénales, enquêtes et poursuites ». Sources : *DeepL* et *Google Translate*

Le résultat le plus significatif est la présence des termes « *individuazione* » et « *accertamento* » en italien respectivement dans *DeepL* et *Google Translate*. Ces dénominations correspondent à celles qui sont proposées dans les entrées de *IATE*. Bien que la relation séquentielle de l'expression en français soit maintenue en italien, nous soulignons que ces dénominations, selon nos recherches des étapes précédentes, désignent des pratiques qui s'inscrivent dans une approche de type réactif de l'action criminelle. À ce propos, il est intéressant de noter que sur *Google Translate* si nous comparons en italien et en français d'abord le terme « *accertamento* » (tableau 4) et ensuite le terme « détection » (tableau 5) nous trouvons les résultats suivants :

Moteur de TA	masculin			féminin		
	GT	DT	MT	GT	DT	MT
fr>it	96%	96%	93%	11%	9%	7%
it>fr	96%	95%	97%	18%	8%	11%

Tableau 4 : Pourcentages totaux des traductions correctes avec l'adjectif de compétence, différenciés par moteurs de traduction

Français	Italien
Détection	Rilevamento
Définition : Détection électromagnétique par radar.	

Tableau 5 : Traduction automatique (FR>IT) du terme « détection ».
Source *Google Translate*

Encore une fois, comme nous l'avons remarqué dans la base de données *IATE*, nous voyons que l'italien propose une définition terminologique et le français une définition tirée de la langue générale ou scientifique.

Enfin, nous attirons l'attention sur un autre dernier élément significatif. Le résultat de la traduction automatique en *DeepL* du terme « *accertamento* » de l'italien au français donne le même résultat que *GoogleTranslate* : « évaluation » (tableau 6). Toutefois, dès que ce terme s'associe aux termes « *prevenzione* », « *indagine* » et « *perseguitamento* » dans l'expression que nous avons étudiée dans notre analyse, l'équivalent que l'on obtient en français est « détection », tel que nous pouvons le voir dans le tableau 7 :

Italien	Français
Accertamento	Évaluation

Tableau 6 : Traduction automatique (IT>FR) du terme « *accertamento* ».
Source *DeepL*

Italien	Français
Prevenzione, indagine accertamento e perseguitamento	Prévention, enquête, détection et poursuites

Tableau 7 : Traduction automatique (IT>FR) du terme « *accertamento* » en contexte.
Source *DeepL*

Ce constat nous invite à supposer que la traduction automatique de *DeepL* puise dans les bases de données documentaires et terminologiques européennes. Un détour par le concordancier bilingue de *DeepL*, nommé *Linguee* renforce notre hypothèse. Nous remarquons en effet que, si nous cherchons la traduction de l'expression « *prevenzione, indagine, accertamento e perseguitamento* » tous les résultats en italien et en français sont tirés du site de la législation européenne *Eur-Lex* ou de son site officiel *Europa* et la plupart des extraits renvoient à la normative relative au traitement de données en matière pénale.

L'ensemble de ces résultats nous amène à formuler quelques considérations sur l'articulation entre les dispositifs terminologiques et de traduction automatique et à revenir sur notre hypothèse principale afin de tirer les conclusions de notre analyse.

Observations sur les dispositifs et conclusions

Nous souhaitons tout d'abord souligner avec notre recherche le rapport entre la terminologie, la traduction spécialisée et les outils d'aide à la traduction. Les outils proposés par la terminologie sur le plan institutionnel (banques terminologiques, glossaires) sont constamment alimentés. Les traducteurs, les terminographes, mais aussi les juristes linguistes, les experts ou les étudiants universitaires sont invités à effectuer des recherches terminologiques qui puissent aider les traducteurs confrontés à la multiplication de la terminologie spécialisée. Ces termes utilisés par les traducteurs se propagent ensuite non seulement dans les textes publiés, mais aussi dans les mémoires de traduction, avec pour effet de fabriquer des segments prêts à être réemployés. Comme l'explique Bowker (2001 [1998]: 289) « *the popularity of translation memories means that translators are increasingly turning to translated material as a resource for conducting terminological research* ». À son tour, cette masse terminologique alimente les logiciels de traduction automatique ou les concordanciers en ligne qui sont disponibles en libre accès en ligne et amplifie ce qui a été dit par les instances énonciatrices. Le traducteur automatique devient donc un « relayeur interlinguistique de toutes sortes de messages et de discours » (Guidère, 2009 : 48) et notamment du discours institutionnel européen. À ce propos, il paraît encore important de souligner que le traducteur automatique ne s'appuie que sur des techniques algorithmiques et de *machine learning*. Comme l'explique Cardon (2018 : 67) « les règles de calcul des algorithmes sont procédurales et non substantielles. Les artefacts de calcul n'ont pas un accès sémantique aux informations qu'ils manipulent - c'est-à-dire qu'ils n'en ont pas une compréhension symbolique ». En d'autres termes, notre regard ne vise pas à critiquer l'outil en tant que machine, mais invite plutôt à se tourner vers les choix dénominatifs effectués en amont, au moment de la production ou de la traduction de la terminologie institutionnelle juridique et politique.

En revenant sur notre hypothèse de départ, notre approche qui articule l'étude discursive de la terminologie avec l'observation des dispositifs terminologiques et de traduction, nous a permis de constater des désalignements terminologiques entre les versions linguistiques en français, italien, anglais et espagnol des actes contraignants relatifs au traitement de données dans le champ d'application des enquêtes pénales. À partir de l'exploration de la structure séquentielle en français « prévention + détection + enquêtes + poursuites », nous avons vu que la relation

ontologique d'enchaînement des termes n'est pas toujours maintenue dans les autres langues d'étude. Ce résultat nous a permis de focaliser notre recherche sur le terme « détection » à la fois dans les documents d'analyse, dans les bases de données et dans la traduction automatique. Il en ressort que dans le contexte français ce terme semble désigner une pratique pénale à entreprendre de manière préventive avant l'infraction alors que dans les autres langues nous remarquons, de la part des énonciateurs, des hésitations qui tendent à inscrire le même concept dans une approche de type réactive. Il existe donc de différentes « temporalités » selon la version linguistique des mêmes documents.

Bien que ces désalignements puissent être imputables aux conditions de production des directives européennes et à l'interférence entre langues, le principe constitutif de la co-rédaction (Cosmai, 2014 [2007] : 42) des textes juridiques européens ne permet pas de remonter au document source. Par conséquent, dans l'espace multilingue européen, l'interprétation d'un concept juridique et politique se heurte à la difficulté d'identifier le premier usage du terme dans le texte d'origine. Or, les choix dénommatifs opérés dans les directives sont susceptibles de déclencher des actions juridiques ou d'être porteurs de contenu idéologique. Dans ce contexte, l'analyse contrastive des désalignements, dans les documents et dans les dispositifs terminologiques et de traduction, devient une démarche qui permet de mettre en lumière des écarts interprétatifs et qui invite à évaluer en termes politiques et juridiques l'usage des termes, notamment dans la normative récente sur le traitement des données en matière pénale.

Bibliographie

- Agence des droits fondamentaux de l'Union européenne, 2017. *Surveillance by intelligence services: fundamental rights safeguards and remedies in the EU*. Luxembourg : Publications Office of the EU.
- Bisiani, F. 2020. *Pour une approche discursive de la terminologie européenne et nationale, en français et en espagnol, des textes non contraignants sur les politiques sécuritaires (2001-2018)*. Thèse de doctorat : Université degli Studi di Udine, Université de Paris.
- Bowker, L. 2001 [1998]. « Terminology ». In: Baker, M., Saldhana, G. (éds.), *Routledge encyclopedia of Translation studies*. Londres, New York : Routledge, p. 286-290.
- Branca-Rosoff, S. 2001. *L'institution des langues. Autour de Renée Balibar*. Paris: Maison de Sciences de l'Homme.
- Cabré, M. T. 1998 [1992]. *La terminologie. Théorie, méthode et applications*. Ottawa, Paris : Presses de l'Université d'Ottawa, Armand Colin.
- Cardon, D. 2018. « Le pouvoir des algorithmes ». *Pouvoirs*, vol° 164, n° 1, p. 63-73.
- Chopin, O., Oudet, B. 2019 [2016]. *Renseignement et sécurité*. Paris : Armand Colin.
- Cinato, L. 2010. «Terminologia italiana e tedesca nei documenti COM ». In : Raus, R. (éd.), *Multilinguismo e terminologia nell'Unione europea: problematiche e prospettive*. Milan : Hoepli, p. 85-114.
- Cosmai, D. 2014 [2007]. *The Languages of Europe. Multilingualism and Translation in the EU Institutions: Practice, Problems and Perspectives*. Traduit et édité par D. A. Best. Bruxelles : Éditions de l'Université de Bruxelles.

- Dufour, F., Rosier, L. 2012. « Introduction. Héritages et reconfigurations conceptuelles de l'analyse du discours 'à la française' : perte ou profit ? ». *Langage et Société*, n° 140, p. 5-13.
- Guidère, M. 2008. *La communication multilingue. Traduction commerciale et institutionnelle*. Bruxelles : De Boeck.
- Guidère, M. 2009. « Qu'est-ce que la communication orientée ? ». In : Guidère, M. (éd.). *Traduction et Communication orientée*. Paris : Le Manuscrit, p. 13-59.
- L'Homme, M. C. 2004. *La terminologie : principes et techniques*. Montréal : PUM.
- Raus, R. 2013. *La terminologie multilingue. La traduction des termes de l'égalité H/F dans le discours international*. Bruxelles : De Boeck.
- Raus, R. 2019. « L'approche d'archive pour accéder aux termes dans la communication multilingue : le cas des concordanciers ». In : Kottelat, P., Mattioda, M.M., Novallet, L., Raus, R. (éds.). *Confluences, liens, rencontres: synergies d'expression française*. Turin : METI, p. 225-240.

Sitographie⁵

- DeepL : <https://www.deepl.com/translator>
- Eur-Lex
- Accès au droit de l'Union européenne : <https://eur-lex.europa.eu/homepage.html>
- Europa
- Site officiel de l'Union européenne : <https://europa.eu>
- Google Translate : <https://translate.google.com/?hl=es>
- IATE - InterActive Terminology for Europe : <https://iate.europa.eu/home>
- Linguee : <https://www.linguee.com>
- Sketch Engine : <https://www.sketchengine.eu>

Notes

1. *Arma a doppio taglio*.
2. Il est bon de préciser qu'un cadre législatif relatif au traitement de données existe depuis 1995 en Europe et que le texte de référence actuel sur ce sujet est le Règlement Général sur la Protection des Données de 2016. Toutefois, notre étude se focalise uniquement sur les textes qui régulent le traitement de données dans le champ d'application des enquêtes pénales.
3. Les textes des décisions et des directives sont tirés du portail du droit de l'Union européenne *Eur-Lex*.
4. À ce propos, il est bon de rappeler que le dispositif législatif de l'Union européenne ne reconnaît pas les procédures de traduction (Cosmai, 2014 [2007] : 42). En d'autres termes, bien que les documents des directives soient souvent des traductions, il n'existe pas de lien de subordination entre le texte source et le texte cible. Ainsi, chaque texte correspond à une version officielle du même document.
5. La date de la dernière consultation des sites est le 30 janvier 2021.



ISSN 1724-0700

ISSN en ligne 2260-8087

La ressource FAIRterm : entre pratique pédagogique et professionnalisation en traduction spécialisée

Federica Vezzani

Université de Padoue, Italie

federica.vezzani@unipd.it

<https://orcid.org/0000-0003-2240-6127>

Reçu le 08-02-2021 / Évalué le 10-03-2021 / Accepté le 18-04-2021

Résumé

Cette étude vise la description d'un nouveau produit linguistique développé en tant que support de la pratique pédagogique et de la professionnalisation en traduction spécialisée. La ressource, dénommée FAIRterm, est configurée comme une collection de fiches terminologiques multilingues visant à prendre en charge le processus de décodage et de transcodage de la terminologie d'un domaine donné. L'outil est conçu dans le respect des normes ISO en vigueur en matière de gestion terminologique. Pour sa validation, nous proposons la description d'une expérience pédagogique menée pour les langues de travail italien-français dans la perspective de professionnalisation des apprenants-traducteurs dans le domaine œnologique.

Mots-clés : ressources terminologiques, didactique de la traduction spécialisée, terminologie multilingue, terminologie de l'œnologie, normalisation terminologique

La risorsa FAIRterm: tra pratica pedagogica e professionalizzazione in traduzione specializzata

Riassunto

Questo studio mira alla descrizione di un nuovo prodotto linguistico sviluppato per supportare la pratica pedagogica e la professionalizzazione in traduzione specializzata. La risorsa, denominata FAIRterm, si configura come una collezione di schede terminologiche multilingue per sostenere il processo di decodifica e transcodifica della terminologia di un dato dominio. Lo strumento è concepito rispettando gli standard ISO in vigore in materia di gestione della terminologia. Per la sua validazione, proponiamo la descrizione di un esperimento pedagogico condotto per le lingue di lavoro italiano-francese nella prospettiva di professionalizzazione di apprendenti traduttori nel dominio enologico.

Parole chiave: risorse terminologiche, didattica della traduzione specializzata, terminologia multilingue, terminologia dell'œnologie, standardizzazione terminologica

The FAIRterm resource: between pedagogical practice and professionalization in specialized translation

Abstract

This study aims at describing a new linguistic product developed in order to support pedagogical practice and professionalization in specialized translation. The resource, named FAIRterm, is configured as a collection of multilingual terminological records to assist the process of decoding and transcoding the terminology of a given domain. The tool is designed in compliance with the ISO standards in force in terms of terminology management. For its validation, we propose the description of a pedagogical experiment conducted for the Italian-French working languages in the perspective of professionalization of translation learners in the oenological domain.

Keywords: terminological resources, didactics of specialized translation, multilingual terminology, oenology terminology, terminological standardization

Introduction

Le développement exponentiel de l'ingénierie linguistique a conduit, ces dernières années, à la conception et à la mise en œuvre de nombreux produits linguistiques ayant différentes formes et fonctions (Bureau Van Dijk, 2007). Dans cette étude, nous nous concentrons sur le domaine agroalimentaire, en tant que secteur commercial qui relie fortement l'Italie et la France dans un réseau dense d'importations et d'exportations. Le panorama de la recherche concernant la terminologie agroalimentaire trouve des espaces de réflexion intéressants sous différents angles. En particulier, un fort intérêt est porté sur la dimension applicative visant à la construction de terminologies, de dictionnaires, d'ontologies et d'applications informatiques pour les ressources terminologiques, les systèmes d'information et la gestion des connaissances (Chessa *et alii*, 2015).

Dans le contexte agroalimentaire, nous délimitons l'objet de notre analyse à la langue de spécialité du vin et, en particulier, à son volet terminologique (Ludi, 2014 ; Tsai, 2014). Le domaine de l'œnologie est peuplé de nombreuses ressources linguistiques qui visent à satisfaire, notamment, le besoin de transmission multilingue de la terminologie technique¹. Que ce soit sous forme d'un simple glossaire ou de collection de fiches terminologiques, ces produits de l'industrie et/ou de la recherche linguistique constituent un support valable d'un point de vue terminologique pour l'activité professionnelle des traducteurs technico-scientifiques. Dans le contexte national italien, l'une des ressources terminologiques multilingues les plus complètes du secteur vitivinicole a été développée par le centre

de recherche en terminologie multilingue Ce.R.Te.M. de l'Université de Gênes (Piccardo, 2020). Cette ressource, disponible en ligne (voir la Sitographie), permet la consultation de fiches accessibles dans différentes langues - dont l'italien et le français - contenant différentes catégories de données terminologiques pour le terme source et le terme cible. Les informations fournies permettent, en ce sens, un décodage et un transcodage des informations techniques par le traducteur de textes spécialisés. D'un point de vue pédagogique, la combinaison de ces deux étapes du processus de traduction spécialisée constitue la pierre angulaire de la recherche terminologique pour les aspirants professionnels de la langue. D'une manière générale, le processus de formation des apprenants-traducteurs de textes spécialisés repose sur la transmission d'une méthode de travail qui permet - quel que soit le domaine d'investigation - le décodage correct, c'est-à-dire la compréhension des informations fournies dans le texte source, et le transcodage fiable, à savoir la transmission d'informations dans le texte cible (Lethuillier, 2003 ; Gouadec, 2005).

Dans ce panorama, cette étude porte sur la description de la ressource FAIRterm (voir la Sitographie) en tant que nouveau produit visant à soutenir 1) la didactique de la traduction spécialisée tant dans le décodage que dans le transcodage de la terminologie technique, et 2) la professionnalisation des apprenants-traducteurs. La réalisation de cet outil s'inscrit dans un cadre plus large qui implique la participation du Département d'études linguistiques et littéraires (DiSLL) et le Département d'ingénierie de l'information (DEI) de l'Université de Padoue. En particulier, la collaboration consiste en la mise en œuvre d'une ressource conçue comme une collection de fiches terminologiques multilingues structurées et normalisées selon les normes ISO en vigueur en matière de gestion terminologique (Vezzani et Di Nunzio, 2020a-b). Ainsi, la ressource FAIRterm n'est pas configurée comme une plateforme d'apprentissage d'une ou plusieurs langues étrangères, comme par exemple l'application Babel qui, parmi ses nombreuses fonctions, propose également un guide du vocabulaire de l'œnologie². L'outil que nous proposons fournit plutôt une méthode guidée et efficace pour la formation des futurs traducteurs technico-scientifiques en suivant une perspective terminologique. En ce sens, la ressource FAIRterm reprend le même principe sous-jacent à la conception de la base de données ARTES (Aide à la Rédaction de Textes Scientifiques) développée par l'Université Paris Diderot qui vise à intensifier la relation étroite entre recherche et enseignement en terminologie (Pecman et Kübler, 2011)³. Comme pour le projet ARTES, les utilisateurs identifiés pour la ressource FAIRterm sont principalement des apprenants-traducteurs et de futurs professionnels des langues de spécialité. Les valeurs ajoutées de l'outil ici proposé s'inscrivent notamment dans la normalisation structurelle et la richesse du modèle de fiche terminologique mis à la disposition de l'utilisateur.

Les sections suivantes seront consacrées à la description de la structure de la ressource FAIRterm et de ses fonctions (section 1), et à la présentation d'une expérience pédagogique menée dans le cadre de la formation en traduction spécialisée active (de l'italien vers le français) en œnologie (section 2). En conclusion, nous avancerons des évaluations finales et définirons les perspectives de recherche.

1. Ressource FAIRterm

La conception et l'implémentation de la ressource FAIRterm résultent de la volonté de mettre à disposition un outil permettant de soutenir le processus de traduction spécialisée des futurs professionnels de la langue. En particulier, l'objectif est de fournir une méthode guidée pour une compréhension correcte, une maîtrise optimale et une transmission efficace de la terminologie technique dans un domaine de travail donné. Afin de répondre à ces critères, la ressource FAIRterm a été conçue - sur la base d'études antérieures (Vezzani et Di Nunzio, 2020a-b) - comme une base de données multilingue qui collecte des fiches terminologiques structurées et normalisées selon les normes ISO TC 37/SC 3 : 1) ISO 16642 : 2017 pour le métamodèle structurel, 2) ISO 12620 : 2019 pour les catégories de données affichées et 3) ISO 30042 : 2019 pour le format d'échange de données terminologiques⁴.

La ressource FAIRterm, gratuitement disponible en ligne, permet la compilation de fiches terminologiques multilingues. L'accès en mode « compilation » nécessite que l'utilisateur s'authentifie avec ses identifiants personnels afin de mener un travail individuel d'insertion, de modification et de stockage de données terminologiques⁵. Une fois authentifié, l'utilisateur doit sélectionner la langue source du terme à analyser, saisir l'unité lexicale objet d'analyse dans la boîte dédiée et l'ajouter à la base de données⁶. Par la suite, l'utilisateur peut avancer à la phase de compilation en cliquant sur la fonction *Compile record*. Le système demande tout d'abord la sélection de deux langues de travail et permet la recherche du terme saisi dans la phase précédente. À ce stade, l'utilisateur disposera d'un modèle de fiche terminologique bilingue avec mise en miroir des mêmes catégories de données pour le terme source à gauche de l'écran, et pour le terme cible à droite de l'écran⁷.

Les informations linguistiques requises - 40 au total - sont regroupées en quatre panneaux : 1) caractéristiques formelles, 2) sémantique, 3) variation et 4) usage. Ce modèle de fiche a été conçu pour permettre aux apprenants-traducteurs de réfléchir au comportement morphosyntaxique, sémantique et phraséologique du terme technique analysé et de son équivalent en favorisant de cette manière le

processus de décodage et de transcodage du texte. Le tableau 1 présente en détail la liste des catégories de données à renseigner pour la paire de termes dans les deux langues de travail :

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

Tableau 1 : Catégories de données de la ressource FAIRterm.

Une fois les informations saisies, l'utilisateur peut les stocker en cliquant sur la fonction *Save*. À notre avis, l'un des aspects les plus avantageux de cette ressource est la possibilité offerte d'exporter les données terminologiques saisies à la fois sous forme tabulaire (*Download TSV - Tab separated values*) et au format standard *TermBase Exchange* (*Download TBX*) conformément à la norme précédemment citée ISO 30042 : 2019. Cet avantage se traduit par la possibilité d'importer des données structurées dans les systèmes de Traduction Assistée par Ordinateur (TAO) afin de bénéficier, pendant le processus de traduction spécialisée, de l'analyse terminologique menée au préalable. Enfin, la ressource FAIRterm a également été conçue dans l'esprit de la « science ouverte » : son nom dérive de l'application des principes FAIR (Wilkinson *et alii*, 2016) promus dans le cadre de la plateforme *European Open Science Cloud* (voir la Sitographie) ayant pour but ultime de fournir à la communauté scientifique des données de la recherche trouvables, accessibles, interopérables et réutilisables. Dans cette perspective, toutes les données générées à partir de la ressource FAIRterm sont intégrées dans des bases de données spécifiquement conçues pour deux projets actuellement en cours en terminologie commerciale et médicale⁸. Les fiches terminologiques issues de l'expérience pédagogique décrite dans la section suivante et élaborées pour le domaine œnologique sont, à cet égard, librement accessibles en ligne⁹.

2. Expérience en traduction spécialisée

Afin de valider la méthodologie didactique qui implique l'utilisation de l'application FAIRterm pour le décodage et le transcodage de la terminologie, nous avons mené une expérience en traduction spécialisée active dans le domaine œnologique pour le couple de langues italien-français. Cette expérience a vu la participation des étudiants de la deuxième année du cours de *Traduzione Specializzata francese 2* (année académique 2020/2021) dispensé dans le cadre du programme de Master en *Lingue Moderne per la Comunicazione e la Cooperazione Internazionale* de l'Université de Padoue ayant une forte orientation vers la formation des futurs professionnels des langues (traducteurs, interprètes, terminologues et rédacteurs techniques). En particulier, le nombre d'étudiants ayant participé à l'expérience a été de 30 sur 36 étudiants inscrits au cours. Les participants ont été invités à adopter une méthodologie de travail systématique afin d'accomplir de manière optimale et satisfaisante le projet requis pour l'évaluation finale consistant en une traduction active, de l'italien vers le français, de documentation technique pour le domaine de l'œnologie. En particulier, les portions de texte à traduire ont été extraites de la deuxième édition de « *Il Manuale del sommelier: principi di viticoltura ed enologia, degustazione, abbinamenti, legislazione, altre bevande* » publié par De Nicola et alii en 2014. Ce manuel offre un panorama vaste et hétérogène qui comprend des notions techniques de viticulture, de vinification, de dégustation, d'œnogastrologie et de législation. Le processus de traduction de cette documentation spécialisée nécessite donc une méthode rigoureuse d'acquisition et de maîtrise de la terminologie pertinente pour le domaine et les sous-domaines d'analyse.

2.1. Démarche adoptée

La méthodologie de travail proposée aux participants est structurée en phases progressives qui visent à faciliter la compréhension de l'information contenue dans le texte source et sa fidèle transmission dans le texte cible :

- 1) compilation de deux corpus de référence pour les deux langues de travail ;
- 2) extraction et sélection des termes pertinents ;
- 3) compilation de fiches terminologiques sur l'application FAIRterm;
- 4) exportation d'un concordancier de travail commun ;
- 5) importation des données sur un système de gestion terminologique;
- 6) traduction assistée par ordinateur (TAO) du lot de texte attribué.

La phase de compilation des corpus de référence vise à familiariser les participants avec la terminologie spécifique du domaine œnologique dans les deux langues

de travail. Afin de couvrir la plupart des sous-domaines traités dans le manuel, sept groupes de travail ont été organisés, auxquels les participants ont adhéré en fonction de leurs préférences thématiques. Le tableau 2 résume l'organisation des groupes, le nombre de participants, les sous-domaines d'analyse et les thématiques spécifiques à chaque sous-domaine :

Groupe	Participants	Sous-domaine	Thématiques spécifiques
1	4	Le métier du sommelier	1) histoire/évolution de la profession, 2) formation professionnelle, 3) outils à l'usage du professionnel, 4) tâches du professionnel.
2	4	Viticulture	1) culture de la vigne, 2) élevage de la vigne, 3) cycle végétatif de la vigne, 4) types de vignes.
3	5	Vinification	1) vendange, 2) pressurage, 3) vinification, 4) fermentation, 5) affinage.
4	5	Vins spéciaux	1) vin mousseux, 2) vin aromatisé, 3) vin de liqueur, 4) vin de paille, 5) vin de glace.
5	5	Dégustation	1) techniques de dégustation, 2) outils de dégustation, 3) examen visuel, 4) examen olfactif, 5) examen gustatif.
6	3	Législation	1) législation nationale, 2) législation européenne, 3) législation mondiale.
7	4	Autres boissons alcoolisées	1) bière, 2) grappa, 3) brandy, 4) whisky.

Tableau 2 : Travail de compilation du corpus

Pour chaque sous-domaine, chaque participant a librement choisi une thématique de prédilection pour laquelle il s'agissait de constituer deux corpus de référence linguistiquement représentatifs du sujet. La mise en forme du corpus italien et français a été menée à l'aide du logiciel d'analyse de textes Sketch Engine. Le système disponible en ligne, qui permet l'étude des corpus dans une perspective multilingue, est doté d'un nombre considérable de fonctionnalités qui le distinguent des autres outils de gestion de corpus disponibles sur le marché (Kilgarrieff *et alii*, 2014). Pour la constitution du corpus de travail, les participants ont individuellement mené leurs recherches sur Internet, en sélectionnant les documents les plus représentatifs et fiables pour le sujet choisi et en les important dans le logiciel. Par la suite, les documents ont été partagés entre les participants et regroupés afin de constituer, pour chaque sous-domaine d'analyse, deux corpus distincts - l'un pour le français et l'autre pour l'italien - et thématiquement comparables.

Chaque corpus contient environ 50-60 documents et a été mis à la disposition de tous les participants via la fonction de partage interne de Sketch Engine. Pour des raisons d'espace gratuit disponible sur les comptes Sketch Engine des étudiants, les deux corpus contiennent un total de 1 000 000 mots équitablement répartis entre la collection de textes français et la collection de textes italiens.

La phase suivante de la démarche adoptée implique l'extraction et la sélection des termes pertinents pour chaque sous-domaine d'étude. Les étudiants ont travaillé en groupes en respectant la subdivision choisie et, pour chaque corpus précédemment compilé, ils ont utilisé la fonctionnalité de Sketch Engine qui prend le nom de *Keywords and term extraction* pour l'extraction automatique de termes simples (composés d'une seule unité lexicale) et des termes complexes (composés de plusieurs unités lexicales). Le résultat de l'extraction a fourni une liste de candidats termes et, parmi ceux-ci, les participants ont procédé à une sélection des termes simples et complexes les plus pertinents pour les sept sous-domaines identifiés. En particulier, s'agissant d'une expérience de traduction active (de l'italien vers le français), les participants ont été invités à se concentrer sur la sélection des termes en langue source. Le résultat de la sélection a produit une liste de 1 530 termes désignant autant de concepts différents pour lesquels compiler le modèle de fiche terminologique bilingue décrit dans la section précédente. Le total des concepts identifiés a été réparti entre les 30 étudiants : chaque participant a été donc chargé de l'analyse de 51 concepts différents.

Les concepts ont été analysés dans leur représentation terminologique à travers la compilation de fiches bilingues directement sur la ressource FAIRterm (mode compilation)¹⁰. Dans cette phase, les participants ont travaillé individuellement : le mode compilation permet à l'utilisateur de visualiser et de modifier uniquement ses propres données. Chaque participant a produit 102 fiches terminologiques : 51 pour l'italien et 51 pour le français. Les 3060 fiches terminologiques sont en cours de révision et de validation finale par l'auteure. Les analyses terminologiques menées par chaque étudiant ont ensuite été mises à la libre disposition de tous les participants. La ressource FAIRterm offre la possibilité d'extrapoler automatiquement une sélection de champs des fiches terminologiques téléchargeable dans un seul fichier Excel via la fonction *Download concordancier*. Le concordancier, autrement nommé « répertoire terminologique » (Gouadec, 1996), est un outil qui joue un rôle prépondérant dans l'assurance de qualité d'une traduction. Les concordances terminologiques qu'il contient sont des équations entre les désignations relevées dans le/les document/s en langue source et les désignations des mêmes objets qui seront utilisés dans le/les document/s en langue cible. Ce répertoire, qui se présente sous forme de fichier Excel, est une version « allégée » des fiches terminologiques précédemment compilées. Il ne contient que la sélection des champs

suivants : 1) le terme en langue source, 2) sa catégorie grammaticale, 3) le domaine de travail, 4) l'équivalent en langue cible, 5) sa catégorie grammaticale, 6) les notes du traducteur, 7) le statut de certitude ou d'incertitude concernant le choix de l'équivalent, 8) le groupe auquel appartient la personne responsable de l'analyse du couple de termes, 9) l'auteur, et 10) la date de compilation¹¹.

L'étape d'extraction automatique du concordancier supporte la phase suivante de la démarche adoptée qui implique l'utilisation d'un outil de traduction assistée par ordinateur (TAO). Pour cette expérience spécifique, les participants ont été invités à traduire à l'aide du système de TAO SDL Trados Studio 2019 (voir la Sitographie). Ce logiciel a été choisi car il permet une gestion avancée des collections de données terminologiques via le système SDL Multiterm. Par rapport à d'autres logiciels de TAO, comme OmegaT ou Matecat qui permettent d'importer des fichiers contenant un nombre limité de catégories de données terminologiques (terme, équivalent et notes), SDL Trados Studio permet d'importer des fichiers structurés plus complexes en termes de quantité d'informations fournies. Dans ce sens, afin de pouvoir bénéficier de l'étude terminologique précédemment menée, les participants ont importé le fichier Excel contenant le concordancier dans le système de gestion terminologique SDL Multiterm. Ensuite, chaque participant a pu procéder à la traduction d'un lot de texte assigné individuellement à l'aide de SDL Trados Studio. Chaque terme identifié dans le texte à traduire est donc automatiquement accompagné de toutes les informations contenues dans le répertoire terminologique au profit de la qualité de la traduction finale¹². En ce sens, la pratique mise en œuvre avec l'élaboration des fiches, leur intégration dans le système TAO et leur utilisation en traduction correspond à un travail complet visant la composante instrumentale demandée aux futurs professionnels des langues telle que décrite par le référentiel de compétence EMT¹³.

2.2. Évaluation de la ressource

Consciente du fait que la validation de l'utilité et de l'ergonomie d'une ressource terminologique est une activité qui nécessite de nombreuses expériences à réaliser sur le long terme, nous avançons quelques observations découlant de cette première expérience décrite.

2.2.1. Retours des utilisateurs

Selon les retours des utilisateurs obtenus suite à la demande de rédaction d'un compte-rendu final pour l'évaluation individuelle de la démarche adoptée, la ressource FAIRterm s'est avérée utile pour le processus de décodage et de

transcodage d'informations spécialisées et conviviale en termes d'interface graphique. Selon la déclaration unanime des étudiants eux-mêmes, cette méthode leur a permis d'élargir leur champ de compétences en traduction spécialisée. En particulier, les étudiants ont apprécié la méthode d'analyse terminologique guidée et structurée en fiches à remplir pour les deux langues de travail. Bien que la phase de saisie des données ait nécessité de nombreux efforts en termes de recherche et de temps consacré, l'évaluation commune a été largement positive car le système permet de stocker et de (re)consulter les informations scrupuleusement obtenues à tout moment sans le risque de perdre le travail effectué précédemment. De plus, la ressource permet d'exporter des données terminologiques au format tabulaire et au format TBX standard (dernière version de 2019). Cet avantage se reflète dans la phase de traduction spécialisée car les données obtenues peuvent être facilement intégrées dans les différents systèmes de TAO, permettant ainsi de rendre disponibles des analyses précises pour chaque terme lors du processus de rédaction du texte en langue cible.

2.2.2. Résultats qualitatifs

D'un point de vue qualitatif, l'analyse terminologique menée avec la ressource a permis de détecter certains phénomènes linguistiques communs qui se reflètent dans cette langue de spécialité pour les deux langues de travail¹⁴. Premièrement, il est possible de remarquer une utilisation fréquente des anglicismes dans la langue italienne : par exemple, *appassimento off-vine* et *appassimento on-vine* ont été traduits par leurs équivalents français respectifs « passerillage hors souche » et « passerillage sur souche ». Il existe également de nombreux cas de variation dénominative des termes désignant le même concept (Freixa, 2006) : par exemple, la variante synonymique populaire du terme français « eau-de-vie » est le terme « gnôle » également sujet à variation orthographique et diatopique (« gnole », « gniole », « gniôle » et « gnaule »). Les champs du modèle de fiche ont permis d'identifier également des cas de synonymie/quasi-synonymie. Par exemple, en italien, les termes *sentore*, *odore* et *profumo* d'un vin sont fréquemment utilisés comme synonymes d'*aroma* bien qu'ils ne véhiculent pas exactement la même signification : les premiers font référence à l'examen olfactif d'un vin, alors que le deuxième indique sa perception du point de vue gustatif. De plus, bien que le but ultime de l'utilisation d'une terminologie spécialisée soit d'éviter toute ambiguïté dans l'interprétation de sa signification, il existe des cas récurrents de polysémie qui ont nécessité une étude approfondie du contexte d'utilisation. Parmi ceux-ci, les termes *vinificatore*/« vinificateur » et *distillatore*/« distillateur » désignent dans les deux cas et pour les deux langues tant 1) l'outil employé pour les deux

procédures, que 2) la personne chargée de leur exécution. Enfin, il est possible d'identifier dans les deux langues une utilisation généralisée de métaphores relatives au monde animal et végétal : parmi celles-ci, on note le « col de cygne » et la « serpentine » en tant que parties de l'alambic et le verre « tulipe » pour la dégustation de la boisson alcoolisée. Enfin, d'un point de vue strictement traductologique, les analyses menées avec la ressource FAIRterm et la constitution de corpus de travail pertinents pour ce domaine de spécialité ont permis d'aider l'utilisateur dans le choix du meilleur équivalent. Il existe de nombreux cas de termes en italien qui peuvent avoir plus d'un équivalent potentiellement correct en français et qui peuvent générer, donc, des difficultés dans le processus de traduction. Pour ne citer qu'un exemple, le terme simple *lievito* en italien présente deux équivalents possibles en français : « levure » et « levain ». Les deux termes se distinguent en ce que le premier, plus générique, est utilisé en référence à la production de différents produits (pain, bière, vin, etc.) tandis que le deuxième a une signification plus spécifique - et une production différente - applicable exclusivement à la préparation de pâtes levées. L'analyse a donc permis de sélectionner le terme « levure » comme équivalent correct de *lievito* dans le contexte de la vinification.

Les études de cas décrites ici ne sont qu'une petite sélection du travail excellent mené par les participants à l'expérience. Actuellement, notre intérêt s'est davantage porté sur une première validation de la ressource développée qui constitue un support stratégique pour la familiarisation avec un domaine d'analyse technique, la maîtrise de la terminologie pertinente et sa fidèle transmission dans le processus de traduction spécialisé.

Conclusions et perspectives

Dans cette étude, nous avons présenté l'exploitation de la ressource FAIRterm en tant que support de la pratique pédagogique et de la professionnalisation en traduction spécialisée. Pour tester sa validité, nous avons décrit une expérience menée afin de former les futurs professionnels de la langue à la traduction de documentation technique dans le domaine de l'œnologie. La ressource est actuellement utilisée pour sa validation dans d'autres cours du Département d'études linguistiques et littéraires, en particulier les cours de *Terminografia orientata alla traduzione* et de *Tecnologie per la traduzione*. Notre perspective est de valider la fonctionnalité de cette ressource dans d'autres domaines de spécialisation et, ensuite, de la mettre à disposition pour un usage à grande échelle.

Pour ce qui est de l'expérience menée, les résultats des analyses réalisées à l'aide des fiches terminologiques FAIRterm seront prochainement intégrés dans

un projet européen. En particulier, notre département collabore depuis deux ans avec l'Unité de Coordination de Terminologie (TermCoord) du Parlement européen dans le cadre du programme « Terminologies sans frontières »¹⁵. L'initiative est organisée autour de nombreux projets consacrés à différents domaines de spécialisation. La collection terminologique sur l'oénologie menée par les étudiants sera donc accueillie par le projet YourTermFOOD (voir la Sitographie) qui vise précisément à collecter les ressources terminologiques multilingues dans le domaine de la gastronomie au sens large.

Bibliographie

- Bureau, Van Dijk. 2007. *Technologies de la Langue en Europe : Marché et Tendances*. Ministère de la Recherche. [En ligne] : [http : www.technolangue.net](http://www.technolangue.net) [consulté le 30 octobre 2020].
- Chessa, F., De Giovanni, C., Zanola, M. T. 2015. *La terminologia dell'agroalimentare*. Milan : FrancoAngeli.
- Freixa, J. 2006. « Causes of denominative variation in terminology: A typology proposal ». *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, n. 12 (1), p. 51-77.
- Gouadec, D. 1996. *Terminologie et phraséologie pour traduire : le concordancier du traducteur*. Paris : Maison du dictionnaire.
- Gouadec, D. 2005. « Modélisation du processus d'exécution des traductions ». *Meta : journal des traducteurs/Meta: Translators' Journal*, n. 50 (2), p. 643-655.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovár, J. Michelfeit, P. Rychl'y, and V. Suchomel. 2014. « The Sketch Engine : ten years on ». *Lexicography*, n.1(1), p. 7-36.
- Lethuillier, J. 2003. « L'enseignement des langues de spécialité comme préparation à la traduction spécialisée ». *Meta : journal des traducteurs/Meta : Translators' Journal*, n. 48 (3), p. 379-392.
- Ludi, F. 2014. « Le vin et sa terminologie ». *Traduire. Revue française de la traduction*, n. 231, p. 52-59.
- Pecman, M., Kübler, N. 2011. « ARTES : an online lexical database for research and teaching in specialized translation and communication ». In *Proceedings of the First International Workshop on Lexical Resources (WoLer)*, p. 87-93.
- Piccardo G. 2020. « CeRTeM - Centro di Ricerca in Terminologia Multilingue dell'Università di Genova ». *Publifarum*, n. 12. [En ligne] : <https://www.publifarum.farum.it/index.php/publifarum/article/view/328> [consulté le 05 février 2021].
- Tsai, C. 2014. « La langue spécialisée du vin : étude comparative de comptes rendus de dégustation ». *Revue française de linguistique appliquée*, n. 19 (1), p. 117-131.
- Vezzani, F., Di Nunzio, G.M. 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*, n. 26 (2), p. 266-298.
- Vezzani, F., Di Nunzio, G. M.2020b. « On the Formal Standardization of Terminology Resources: The Case Study of TriMED ». In *Proceedings of the 12th Language Resources and Evaluation Conference*, p.4903-4910.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. 2016. « The FAIR Guiding Principles for scientific data management and stewardship ». *Scientific data*, n. 3(1), p. 1-9.

Sitographie¹⁶

Application Babbel

<https://www.babbel.com>

Base de données ARTES

<https://artes.app.univ-paris-diderot.fr/artes-symfony/web/app.php>

FAIRterm

<https://shiny.dei.unipd.it/fairterm/compilation.html>

<https://yourterm.org/wp-content/uploads/2021/03/FAIRterm-presentation-update.pdf>

Glossaire œnologique Ce.R.Te.M.

http://www.farum.it/glos_enol/

Normes ISO

<https://bit.ly/3rw3Ba8>

Plateforme *European Open Science Cloud*

<https://eosc-portal.eu>

Projet YourTermFOOD

<https://yourterm.org/yourterm-food/>

SDL - Multiterm

<https://bit.ly/2QcwVlu>

SDL - SDL Trados Studio

<https://bit.ly/2QhxNoR>

Sketch Engine

<https://www.sketchengine.eu>

TermCoord

<https://termcoord.eu>

Notes

1. La plateforme Lexicool rassemble, par exemple, une liste exhaustive des ressources disponibles sur la scène internationale dans le domaine du vin : <https://bit.ly/3pYgFVY>
2. <https://fr.babbel.com/fr/magazine/guide-du-vocabulaire-du-vin/>
3. Pour plus d'informations sur le projet et les publications associées : <https://www.eila.univ-paris-diderot.fr/recherche/artes/index>
4. Pour plus d'informations sur le paradigme théorique sous-jacent à ce projet, nous invitons le lecteur à consulter le lien suivant: <https://purl.org/fairterm>
5. Comme nous l'avons décrit, la ressource a été conçue à l'origine pour la formation d'étudiants en tant qu'apprenants-traducteurs. L'accès à la ressource est cependant ouvert et gratuit pour tout professionnel et/ou chercheur travaillant dans les domaines de la terminologie, de la traduction spécialisée et de la rédaction technique. Le formulaire de demande des identifiants d'accès est disponible au lien suivant : <https://bit.ly/3mwzj63>
6. Pour ce qui est de la langue de navigation de la ressource, au moment présent, l'outil ne présente les informations qu'en anglais. Toutefois, nous sommes en train de travailler dans une perspective multilingue afin de fournir la possibilité pour l'utilisateur de choisir sa propre langue de consultation.
7. Environnement de travail de la ressource FAIRterm : <https://yourterm.org/wp-content/uploads/2021/03/FAIRterm-presentation-update.pdf>
8. Pour plus d'informations sur ces projets, nous invitons le lecteur à consulter la page suivante: <https://shiny.dei.unipd.it/fairterm/consultation.html>
9. <https://shiny.dei.unipd.it/CAMEO/>

10. Les identifiants personnels ont été distribués individuellement à chaque participant au début du cours.
11. La numérotation proposée respecte l'ordre progressif des colonnes du fichier Excel.
12. Dans la section 1, nous avons décrit la conception de la ressource FAIRterm afin de faciliter la réutilisation des données terminologiques conformément à la norme ISO 30042 : 2019. Les fiches terminologiques peuvent en fait être téléchargées selon le format TermBase Exchange (TBX) mis à jour selon la dernière version de la norme susmentionnée. À ce stade, il faut mentionner que le système SDL Multiterm permet également d'importer des fichiers au format TermBase Exchange (TBX) : <https://docs.sdl.com/785445/587005/sdl-multiterm-2019/working-with-sdl-multiterm-convert>. Cependant, la version SDL Multiterm de 2019 utilisée par les étudiants ne prend en charge que la conversion des fichiers TBX organisés selon la version obsolète de la norme de 2008. Pour cette raison, nous avons décidé d'importer des données terminologiques au format tabulaire Excel en attendant une mise à jour par le système SDL Multiterm qui soit compatible avec la dernière version de la norme TBX.
13. https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_fr_web.pdf
14. Pour des raisons d'espace, nous ne décrivons que quelques-uns des cas terminologiques identifiés par les participants. Pour une description plus approfondie de chaque terme analysé et pour le repérage des sources consultées, nous renvoyons le lecteur à la consultation de la ressource suivante qui collecte les données de cette expérience : <https://shiny.dei.unipd.it/CAMEO/>.
15. Le lien suivant fournit l'accès aux détails du projet : <https://yourterm.org>. Les collaborations en cours concernent la participation aux projets YourTermMED (<https://yourterm.org/med-university-projects-2/>) et YourTermFEM (<https://yourterm.org/fem-university-projects/>).
16. La date de la dernière consultation des sites est le 05 février 2021.

Synergies Italie n° 17 / 2021



Industries de la langue
et traitement
automatique
des langues (TAL)





GERFLINT

ISSN 1724-0700

ISSN en ligne 2260-8087

Corpus, *Web* et recherche terminologique : le cas des termes du domaine du bien-être animal

Cosimo De Giovanni

Université de Cagliari, Italie
cdegiovanni@unica.it

<https://orcid.org/0000-0003-1976-3166>

Reçu le 10-02-2021 / Évalué le 15-03-2021 / Accepté le 16-04-2021

Résumé

Dans cet article, nous avons voulu démontrer l'apport des corpus numériques à la recherche terminologique. Nous avons analysé la circulation des deux termes *abattoir mobile* et *unità mobile di macellazione* dans les textes officiels de l'Union européenne (UE), des deux États membres, comme la France et l'Italie, qui réglementent le domaine du bien-être animal (BEA). L'utilisation du corpus numérique a été essentielle à partir du moment où les variantes dénominatives des deux termes pouvaient être extraites et décrites. Ainsi, l'interprétation des contextes d'apparition des candidats termes se révèle fondamentale pour l'analyse du point de vue (PdV) et des choix linguistiques opérés par les locuteurs sur la scène discursive. Les corpus numériques peuvent être utiles pour la construction de nouvelles banques de terminologie et pas pour l'amélioration de la qualité de celles déjà existantes.

Mots-clés : *web-corpus*, *corpus-web*, point de vue, corpus heuristique, bien-être animal

Corpus, *Web* e ricerca terminologica: il caso dei termini nel campo del benessere degli animali

Riassunto

In questo articolo abbiamo voluto dimostrare l'apporto dei corpora digitali alla ricerca terminologica. Abbiamo analizzato la circolazione dei due termini *abattoir mobile* e *unità mobile di macellazione* nei testi ufficiali dell'Unione europea (UE), dei due Stati membri, quali Francia e Italia, che regolamentano il settore del benessere animale (BEA). Il ricorso al corpus digitale è stato essenziale dal momento in cui si sono potute estrarre e descrivere le varianti denominative dei due termini. Allo stesso modo, l'interpretazione dei contesti in cui compaiono i termini candidati si rivela fondamentale per l'analisi del punto di vista (PdV) e delle scelte linguistiche operate dai relatori sulla scena discorsiva. I corpora digitali possono essere utili per la costruzione di nuove banche terminologiche e non solamente per il miglioramento della qualità di quelle già esistenti.

Parole chiave: *web-corpus*, *corpus-web*, punto di vista, corpus euristico, benessere animale

Corpus, web and terminological research: the case of terms in the field of animal welfare

Abstract

In this article we wanted to demonstrate the contribution of numerical corpora to terminological research. We have analyzed the circulation of the two terms *abattoir mobile* and *unità mobili di macellazione* in the official texts of the European Union (EU), of the two Member States, i.e. France and Italy, which regulate the animal welfare sector. The use of the numerical corpus was essential from the moment in which the denominative variants of the two terms could be extracted and described. Thus, the interpretation of the contexts in which candidate terms appear is fundamental for the analysis of the point of view (PdV) and of the linguistic choices made by the speakers on the discursive scene. Digital corpora can be useful not only to improve the quality of existing terminology banks but also to build new ones.

Keywords: web-corpus, corpus-web, point of view, heuristic corpus, animal welfare

Introduction

Ces dernières années, on a assisté à une réduction de l'écart entre le corpus et le web dans les études de linguistique. Les raisons qui ont amené à ce rapprochement sont multiples et diversifiées : changement de cap de certains linguistes, ayant condamné l'utilisation du *Web* dans leurs études, prédominance de l'Internet comme média de communication et d'information (Mille, 2014), développement de nouvelles technologies d'analyse des données brutes dans la Toile (*big data*), absence de grands corpus textuels informatisés pour certaines langues et, enfin, liberté accordée par la linguistique de corpus (LdC) à l'utilisation d'outils et de méthodes d'analyse (Teubert, 2009).

En particulier, dans les années 2000, la rencontre du *Web* et de la LdC a entraîné beaucoup de questions, d'incertitudes, de résistances, voire des débats entre divers groupes de linguistes : à la notion de *Web as Corpus* (Kilgariff, 2001) s'opposait le *Web for building corpus* ou le *Web for Corpus* (Hundt, Nesselhauf, Biewer, 2007). D'un côté, à l'instar du corpus, le *Web* est considéré comme « un grand corps de texte en attente d'être exploité, un énorme tissu de données linguistiques souvent entrelacées avec des fils parallèles¹ » (Resnik, Smith, 2003 : 350).

De l'autre, en revanche, le *Web* est vu comme une ressource linguistique, compte tenu du fait qu'il « ne correspond en rien aux caractéristiques habituelles d'un corpus » (Duclaye *et al.*, 2006 : 53).

Cette opposition - qui a entraîné un changement dans la définition traditionnelle du terme de *corpus* et de ses outils (Gatto, 2014 : 149) - a laissé place, au fil

des ans, à deux nouveaux venus, tant sur le plan terminologique que sur le plan conceptuel : le *corpus-web* (Schaeffer-Lacroix, 2020) et le *web-corpus* (Mayaffre, 2010). Tous les deux peuvent être inscrits sous le terme de *corpus numériques*. Les deux approches se prêtent volontiers aux exigences de l'analyse linguistique des données collectées sans opposer les démarches qualitative et quantitative. De plus, elles sont source d'enrichissement et d'intégration de données déjà collectées à partir de corpus traditionnels (parmi les corpus spécialisés, e.g. *Scientext* du *Lidilem*) ou de produits terminographiques déjà existants (e.g. *IATE*) ou en phase d'élaboration.

1. Le *corpus-web*

Un *corpus-web* est un corpus constitué à partir des données repérées dans le *Web* à l'aide de logiciels, appelés *crawler*, *web crawler* ou *web spider*² (Schäfer, Bildhauer 2013) ayant pour fonction de parcourir périodiquement la toile et d'en collecter le contenu afin de permettre au moteur de recherche de l'analyser et de l'indexer, en identifiant les mots clés et les sujets abordés. Des exemples de *corpus-web* sont le *Web as Corpus Kool Ynitiative (WaCKy)* constitué principalement de trois corpus, respectivement en français, en italien et en allemand, d'environ deux milliards de mots chacun et les corpus *TenTen* repérables dans l'outil en ligne *Sketch Engine*.

L'une des limitations majeures du *corpus-web* est représentée par les droits d'auteur. Pour contourner ce problème, certains corpus contiennent des textes disponibles gratuitement et dépourvus de droits d'auteur ou sous licence *Creative Commons*. Un exemple de cela nous est donné par le corpus italien *Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati (PAISÀ)* avec environ 250 millions de *token*. D'autres corpus sont composés de textes réduits à des échantillons de phrases, c'est le cas du *Leipzig Corpora Collection*, ou en d'autres unités respectant les limites du droit d'auteur. Il existe aussi des corpus qui contiennent des textes qui ne présentent pas ce type de contrainte ou qui restent dans un cadre légal par rapport aux droits d'auteur : e.g. les corpus publiés sur *Sketch Engine*, le *Hungarian Webcorpus* avec 1,48 milliard de mots ou encore le *Leeds Collection of Internet Corpora*.

2. Le *web-corpus*

L'utilisation de moteurs de recherche à des fins linguistiques devient de plus en plus fréquente. Les partisans des deux approches ont toujours reconnu les faiblesses du *Web* dans la recherche linguistique. On peut évoquer, à titre d'exemple, la difficulté de déterminer le degré de représentativité des domaines et des langues,

le caractère incontrôlable des données, la présence de fautes (*dirty corpus*) et, enfin, le caractère anarchique du *Web* (Kilgarriff, Grefenstette, 2003 ; De Giovanni, 2009 ; Maniez, 2011). Ces difficultés sont surmontables à condition d'appliquer des critères linguistiques au cours de la recherche ; dans le cas d'une utilisation à des fins terminologiques, à condition de combiner une analyse par domaines et par usages et, dans les cas extrêmes, d'avoir recours à l'avis d'experts (Lerat, 2003 : 231).

Or, l'introduction du nouveau terme de *web-corpus* sert à démanteler la relation à sens unique existant entre *Web* et *corpus* (*Web-corpus*). De cette manière, il n'est plus question de savoir si le *Web* peut être vu comme un corpus, mais de savoir si le corpus peut exister sans le *Web* (*corpus-Web*) (Mayaffre, 2010).

Au-delà de tous les problèmes de nature épistémologique, on peut affirmer que le *web-corpus* est un corpus constitué des données brutes extraites du *Web* à usages linguistiques. Ces données brutes peuvent être traitées aussi bien *hors-ligne* - à l'aide d'un logiciel externe, gratuit ou payant, (e.g. *AntConc*, *WordSmith*) - qu'*en ligne* - à l'aide d'un moteur de recherche interne au logiciel (e.g. *Webcorp* : Renouf, 2009). Un autre exemple de *web-corpus* est l'outil *Ngram viewer* - un moteur de recherche qui permet de fournir des données fréquentielles sur la distribution chronologique de n-grams (n-mots) - qui repose sur la base des données textuelles de *Google Livres*.

3. Corpus et recherche terminologique

L'abandon de l'approche wüsterienne du terme, considéré comme une étiquette du concept, a été décisif lorsqu'une vision textuelle de la terminologie s'est imposée, ouvrant la voie à de nouveaux modèles théoriques (Raus, 2013). L'examen du fonctionnement du terme dans un contexte linguistique est désormais au cœur de l'activité interprétative du linguiste / terminologue. Aujourd'hui, on reconnaît au terme son statut tant d'élément linguistique que d'élément cognitif (Lerat, 1995 ; Zanola, 2018). De plus, dans la perspective d'une terminologie communicationnelle, le terme possède différentes valeurs en fonction du contexte situationnel en vue de répondre à des besoins de communication spécifiques d'un groupe de professionnels. Des changements affectant le monde (sur le plan social, culturel, scientifique, technique, etc.) peuvent entraîner la variation terminologique d'un ou de plusieurs domaines³.

En nous inspirant des travaux de Raus (2013) et Condamines *et al.* (2004), nous proposons une analyse, d'ordre qualitatif et quantitatif, des variations dénominatives et conceptuelles dans le domaine du bien-être animal (dorénavant BEA).

Dans le cadre des variations dénominales, nous accorderons une large place à l'analyse des variantes morphologiques des unités repérées. Suivra une analyse sémantique des variantes.

Afin d'observer le phénomène de la variation, une analyse sur corpus s'impose. Un premier corpus servira à l'extraction des termes en langue française et italienne, relevant du domaine du BEA, sur une ligne temporelle restreinte (de 1991 à 2009). Ce corpus parallèle est formé de textes émanant de l'Union européenne (dorénavant UE) repérés sur le site *EUR-Lex*. Un deuxième corpus se composera des textes législatifs promulgués en France et en Italie, entre 1991 et 2009. Enfin, l'utilisation de corpus numériques sera indispensable pour vérifier le comportement des termes en contexte.

Pour nos analyses, nous ferons appel à deux autres notions capitales, celles de *candidat terme* et de *point de vue* (dorénavant PdV). On définit le *candidat terme*⁴ comme une occurrence extraite d'un corpus susceptible d'être un terme (Delavigne, 2001 ; L'Homme, 2004). Nous incluons dans cette catégorie non seulement le candidat terme de validation (qui est supposé être terminologique ou non terminologique) mais aussi tout candidat qui entre en concurrence avec les termes officiels ou sont officiellement reconnus par les organismes compétents ou sont traités dans des bases de données terminographiques. La deuxième notion est celle de PdV en langues spécialisées (Boisson 2001) se rapportant à la *perception* d'un objet dans la réalité et dépendant des choix (linguistiques) opérés de la part d'un ou plusieurs locuteurs à des fins spécifiques. Dans le même esprit que Condamines et Rebeyrolle (1997), nous tenterons d'analyser les choix linguistiques adoptés de deux PdV : un premier PdV concernant les choix faits aux niveaux supérieurs, par le terminologue et le législateur ; et, enfin, un second PdV qui concerne les choix linguistiques opérés par un groupe de locuteurs lors de l'actualisation d'une compétence lexicale dans une situation de communication concrète.

4. *Abattoirs mobiles, unità mobili di macellazione* et leurs variantes

4.1. Variabilité dénomminative et conceptuelle dans les textes de l'Union européenne et dans le *Web*

Bien que le BEA reste encore une notion controversée, elle part d'un principe fondateur considérant l'animal comme un être *sensible* (Chessa, De Giovanni, 2019). La présence de sensibilité chez l'animal implique qu'il possède des capacités sensorielles et cognitives et que son bien-être est alors un état mental subjectif, dérivé de son rapport harmonieux avec l'environnement qui l'entoure. Le BEA doit être garanti même au moment de leur mise à mort, surtout dans le cas des animaux

d'élevage destinés à la consommation, en abattoir. Le transport jusqu'aux abattoirs (Le Neindre et al., 2009 : 72) pourrait nuire au BEA du fait de mauvais traitements pendant le chargement et déchargement. Pour cette raison, l'UE a tenté de mieux réglementer les opérations de transport avec une série d'instruments législatifs tout en favorisant, dans le même temps, le déploiement de nouveaux types d'abattoirs afin de réduire la souffrance des animaux pendant la période de pré-abattage.

Le Règlement n°1099 de 2009, sur la protection des animaux au moment de leur mise à mort, prévoit le recours aux *abattoirs mobiles* afin de réduire « la nécessité pour les animaux d'être transportés sur de longues distances [ce qui peut...] contribuer à préserver leur bien-être » et laisse aux États membres le soin d'« établir ou de maintenir des règles nationales pour les abattoirs mobiles ». Le même Règlement distingue les abattoirs mobiles des abattoirs fixes en raison des « techniques de contraintes » et enfin, il estime nécessaire que « les règles techniques [soient] adaptées » (considérant n° 40).

Dans les trois langues de travail du Conseil de l'UE on trouve les termes *abattoir mobile* (fr.), *mobile Schlachthöfe* (all.) et *mobile slaughterhouse* (ang.). Au niveau dénomiatif, le terme italien *unità mobile di macellazione* présente une structure binaire complexe sous forme d'une suite d'expressions enchaînées : {[unità mobile] [di macellazione]}. Au niveau conceptuel, le premier élément *unité* évoque l'objet sous un angle différent : celui d'un groupe d'appareils destiné à effectuer une opération. De même, *macellazione* nous offre un autre PdV : on tient compte de l'ensemble des opérations avec lesquelles les animaux d'élevage, destinés à l'alimentation, sont abattus. Nous pouvons donc voir que le choix du terme italien n'est pas entièrement aléatoire. La présence des deux termes *unità mobili di macellazione* et *unità fissa di macellazione*, nous amène à penser que, sur le plan formel, nous sommes confrontés à une unité plurilexémique semi-ouverte avec des possibilités combinatoires limitées. En effet, une interrogation du moteur de recherche Google (4 février 2021) nous a permis de constater la présence du candidat terme *unità di macellazione* (22 700 résultats)⁵, notamment dans les deux secteurs agroalimentaire et juridique. De plus, nous avons constaté que le terme *unità di macellazione* constitue la base pour trois expressions complexes comme *unità di macellazione autonoma* (gérée directement par une ferme ou une entreprise), *unità di macellazione di animali* et *unità di macellazione annua*. L'interrogation de Google nous a donné aussi une nouvelle dénomination, celle d'*unità di macellazione mobile* (6, 63)⁶. Si dans le terme du Règlement l'élément *mobile* se rattache à la tête de l'expression *unità*, dans le second terme le même élément est placé à la fin : dans ce cas, le déplacement concerne expressément tout le groupe d'appareils destiné à l'abattage. En termes linguistiques, c'est la position des traits sémantiques au sein

de chaque sémème qui motive les deux dénominations. Il est important d'ajouter que l'absence d'un candidat terme comme *unità di macellazione fissa* parmi les données obtenues est due au fait que, à notre avis, l'expression *unità di macellazione* dénomme déjà un abattoir fixe.

Toutefois, les termes *abattoir mobile* et *unità mobile di macellazione* ne sont pas les seuls à circuler dans les textes officiels de l'UE. En fait, par l'interrogation du moteur de recherche du site *EUR-Lex* on découvre qu'il y a au total, pour le français, trois variantes et, pour l'italien, quatre variantes, désignant l'abattoir mobile, distribuées dans les différents types législatifs et réparties sur une période comprise entre 1991 et 2004. En voici quelques exemples (tableau 1) :

Année de publication du texte	fr	It
1991	unités mobiles d'abattage	unità di macellazione mobili
1992	installations mobiles d'abattage	impianti mobili di macellazione
1994	unités mobiles d'abattage	impianti mobili di macellazione
2004	abattoirs mobiles	unità di macellazione mobili / mattatoi mobili / macelli mobili

Tableau 1 : Évolution des variations dénominatives dans les textes de l'Union européenne.

Ayant travaillé dans une diachronie restreinte, où la fenêtre temporelle couvre environ dix ans, le résultat que nous avons obtenu est celui d'une variation restreinte.

4.2 Variabilité dénomminative et conceptuelle dans les textes législatifs nationaux : France et Italie

Les actes juridiques nationaux de chaque État membre de l'UE sont un point de départ utile pour la recherche terminologique, surtout lorsqu'il est possible de comparer des méthodes de dénomination et de conceptualisation de deux langues différentes au sein d'un même domaine. Ceci est également possible grâce à l'instrument d'harmonisation législative pour lequel l'UE demande à ses États membres de modifier leur législation nationale pour atteindre un objectif commun. Une harmonisation des pratiques peut entraîner, sous réserve d'exceptions, une harmonisation des systèmes conceptuels et dénominatifs : tout cela dépend surtout de la politique linguistique de chaque pays (Lavault-Olléon, Grossmann, 2008). L'interrogation des portails *Légifrance* et *Normativa*, les sites officiels pour la diffusion et l'accès aux droits français et italien, nous a permis d'obtenir de

premiers résultats, même si l'interrogation du portail italien a été plutôt difficile. En effet, de ce dernier, nous n'avons pu obtenir qu'une seule donnée utile aux fins de notre recherche : le terme *macello mobile* contenu dans le texte d'une loi de 1994⁷. Or, la langue italienne dispose de deux mots, d'usage commun d'après le *Dizionario internazionale* en ligne (De Mauro), pour désigner le lieu destiné à l'abattage des animaux : ceux de *macello* et de *mattatoio*. L'interrogation de la base de données de *Google Books*, à travers le logiciel *N-gram viewer*, nous a permis de vérifier les occurrences des deux mots entre 1500 et 2019 : le mot *macello* est présent à partir de 1502 et continue son évolution jusqu'à nos jours, tandis que l'apparition du mot *mattatoio* remonte aux débuts du XIX^e siècle⁸. En interrogeant le moteur de recherche *Google*, nous avons pu constater une différence dans la fréquence d'occurrences des termes *mattatoio mobile* (1 680, 1 860) et *macello mobile* (552, 851). Nous n'avons pas d'explication à sur cet écart entre les deux termes. Même *Sketch Engine* (*itTenTen16*) confirme ces derniers résultats. De plus, à travers l'interrogation de la base de données *Google Books* nous apprenons que le terme *macelli mobili* a fait l'objet d'une mention dans deux documents datant des années 40 (*Bollettino parlamentare* et *Rivista delle colonie*) lors de la gestion de la production et de l'abattage de viande par les Italiens en Érythrée.

Au sein des documents officiels français, on a pu constater que face à la seule attestation de l'expression *installation mobile d'abattage* la dénomination courante est *abattoir mobile* (6 fois). Dans le *Rapport d'enquête sur les conditions d'abattage des animaux de boucherie dans les abattoirs français* de 2016, présenté à l'Assemblée nationale, sont mentionnés deux types d'abattoir mobile : les *caissons d'abattage* (un prolongement des abattoirs traditionnels) destinés uniquement à l'abattage et à la saignée à la ferme ; et les (*camions*) *abattoirs mobiles* totalement autonomes construits sur les modèles de camions (articulés) suédois (de l'entreprise Hälsingestintan). En France, le recours aux abattoirs mobiles a non seulement l'objectif de préserver le BEA, mais aussi celui de réduire les coûts de transport trop élevés pour les fermes dispersées dans les campagnes françaises.

Ce qui est intéressant, en revanche, c'est la présence dans les deux langues des expressions *macello itinerante* (144, 4) et *abattoir itinérant* (2 300, 278), identifiées pour la première fois lors de l'interrogation de *Sketch Engine* (une seule attestation par langue) et employées comme des variantes de *macello mobile* et d'*abattoir mobile*. Dans les deux langues la dénomination est motivée de la même façon : de par son étymologie, le concept d'*itinérant* est lié à celui de voyage, par conséquent les termes concernés renvoient à l'activité de se déplacer d'un lieu à l'autre, tout en accomplissant des fonctions et des opérations liées à l'abattage.

Conclusion

Les résultats que nous avons obtenus ne sont évidemment pas définitifs, mais ils constituent une première approximation que de nouvelles analyses permettront d'intégrer. Cependant, ces premiers résultats nous amènent à faire quelques considérations finales. Le BEA, étant un domaine interdisciplinaire situé à la frontière entre le droit, la médecine, la psychologie animale, l'éthique et la technique, regorge de terminologies. Malgré la tentative d'harmoniser le système législatif et le système terminologique de tous les États membres de l'UE, le BEA se caractérise par la présence de variantes terminologiques qui circulent dans des contextes différents. Par exemple, dans notre cas, il a été possible de constater que la circulation des termes varie beaucoup, en français et en italien. Ce qui s'explique par l'exploitation de textes différents dans les deux langues au sein du corpus numérique.

Au-delà des questions épistémologiques (Mayaffre, 2010), le corpus numérique constitue une ressource d'une valeur incontestable. Parmi les deux types de corpus numérique, c'est certainement le *web-corpus* qui pose le plus de problèmes, surtout en termes de volatilité des données collectées et de rapidité avec laquelle elles peuvent varier. Cependant, à travers des « méthodes générales d'investigation » (Lerat, 2003 : 231), le *web-corpus* peut très bien se prêter à des analyses en termes quantitatifs et qualitatifs.

En général, le corpus numérique est un support utile non seulement du fait de sa fonction de validation, mais aussi en raison de sa fonction heuristique dans la mesure où il se propose de donner « à un ensemble de faits une structure qui dégage la vue du linguiste dans une situation où celui-ci poursuit une question mais n'a pas de solution » (Scheer, 2004 : 153-154). Cela pourrait contribuer à faire du corpus numérique une ressource à partir de laquelle il est possible de construire directement des banques terminologiques.

Une dernière considération à faire concerne l'interprétation linguistique des données. Elle constitue le résultat auquel le linguiste est parvenu suite à ses analyses. C'est un point de vue (ou un des points de vue) du linguiste / terminologue sur les PdV des autres locuteurs « présents » dans le discours (Kleiber, 2009). Un point de vue discutable, qui ne fait parfois pas l'unanimité, mais qui, dans le cadre des langues de spécialité, pourrait également être valable si une évaluation de l'expert pour comprendre les variations n'est pas requise.

Bibliographie

- Chessa, F., De Giovanni, C. 2019. « Propositions d'interventions terminologiques dans le domaine du bien-être animal (français-italien) », *Rhesis - International Journal of Linguistics, Philology and Literature*, n° 10.1, p. 34-55.
- Condamines, A., Rebeyrolle, J., Soubeille, A. 2004. « Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus », In : *Proceedings of the 11th EURALEX International Congress*, p. 547-557.

- Condamines, A., Rebeyrolle, J. 1997. « Point de vue en langue spécialisée », *Meta : le journal de traducteurs*, n° 42, p.174-184.
- De Giovanni, C. 2009. *Le corpus entre généralités et spécificités*. Bari / Paris : Schena / Alain Baudry et Cie.
- Delavigne, V. 2001. « Repérage de termes dans un corpus de vulgarisation : aspects méthodologiques ». In : *Conférence TIA*, Nancy, p. 33-43.
- De Mauro, T. *Dizionario internazionale*. [En ligne] : <https://dizionario.internazionale.it> [consulté le 04 février 2021].
- Duclaye, F., Collin, O., Petrier, E. 2006. « Fouille du Web pour la collecte de données linguistiques : avantages et inconvénients d'un corpus hors-normes », In : *Actes de l'atelier Fouille du Web des 6èmes Journées Francophones Extraction et Gestion des Connaissances*, p. 53-64.
- Gatto, M. 2014. *Web as Corpus. Theory and Practice*. London : Bloomsbury.
- Hundt, M., Nesselhauf, N., Biewer, C. 2007. *Corpus Linguistics and the Web*. Amsterdam : Rodopi.
- Kleiber, G. 2009. « D'un contexte à l'autre : aspects et dimensions du contexte ». *L'Information Grammaticale*, n° 123, p.17-32.
- Kilgariff, A., Grefenstette, G. 2003. « Introduction to the Special Issue on the Web as Corpus », *Computational Linguistics*, vol. 9, n° 3, p. 334-347.
- Kilgarriff, A. 2001. « Web as corpus ». In : *Proceedings of the Corpus Linguistics Conference (CL 2001)*. Lancaster : University of Lancaster, p. 342-344.
- Lavault-Olléon, E., Grossmann, F. 2008. « Langue du droit et harmonisation terminologique multilingue : l'exemple de LexALP ». *Lidil*, n° 38, p.11-32.
- Le Neindre et al. 2009. *Douleurs animales : les identifier, les comprendre, les limiter chez les animaux d'élevage*. Expertise scientifique collective, Synthèse du rapport, INRA.
- Lerat, P. 1995. *Les langues spécialisées*. Paris : PUF.
- Lerat, P. 2003. « Web et terminologie philologique ». *Linguistica Antverpiensia*, n° 3, p. 231-244.
- L'Homme, M.-C. 2004. *La terminologie : principes et techniques*. Montréal : PUM.
- Maniez, F. 2011. « Quelques tendances récentes de la linguistique de corpus ». In : Maniez, F. (éd.). *Unité et diversité de la linguistique*. Lyon : Université de Lyon, p. 67-85.
- Mayaffre, D. 2010. « Corpus et web-corpus. Réflexion sur la corporalité numérique ». *Cahiers de praxématique*, n° 54-55, p. 233-248.
- Mille, A. 2014. « D'Internet au web », In : Vitali-Rosati, M., Sinatra, Michael E. (éds). *Pratiques de l'édition numérique*. Montréal, Les Presses de l'Université de Montréal, p. 31-48.
- Raus, R. 2013. *La terminologie multilingue. La traduction des termes de l'égalité H/F dans le discours international*. Bruxelles : De Boeck.
- Renouf, A. 2009. « Corpus Linguistics beyond Google : the WebCorp Linguist's Search Engine ». *Digital Studies/le Champ Numérique*, n° 1.
- Resnik, P., Smith, N. A. 2003. « The Web as a Parallel Corpus ». *Computational Linguistics*, n° 3(29), p. 349-380. [En ligne] : <https://www.aclweb.org/anthology/J03-3002.pdf> [consulté le 04 février 2021].
- Schaeffer-Lacroix, E. 2020. « Les corpus web à travers le prisme de l'ALMT ». *Corpus*, n° 20. [En ligne] : <http://journals.openedition.org/corpus/4579> [consulté le 04 février 2021].
- Schäfer R., Bildhauer F. 2013. « Web Corpus Construction ». *Synthesis Lectures on Human Language Technologies*, n° 6(4), p. 1-145.
- Scheer, T. 2004. « Le corpus heuristique : un outil qui montre mais ne démontre pas ». *Corpus*, n° 3. [En ligne] : <http://journals.openedition.org/corpus/210> [consulté le 04 février 2021].
- Simondon, G. 1958. *Du mode d'existence des objets techniques*. Paris : Aubier.

Teubert, W. 2009. « La linguistique de corpus : une alternative ». *Semen. Revue de sémio-linguistique des textes et discours*, n°27. [En ligne] : <https://journals.openedition.org/semen/8923> [consulté le 04 février 2021].

Zanola, M.T. 2018. *Che cos'è la terminologia?*. Rome: Carocci.

Sitographie⁹

Creative Commons

<https://creativecommons.org/licenses/?lang=fr>

Eur-lex

<https://eur-lex.europa.eu/homepage.html?locale=fr>

Hungarian Webcorpus

<http://mokk.bme.hu/en/resources/webcorpus>

Leeds Collection of Internet Corpora

<http://corpus.leeds.ac.uk/internet.html>

Leipzig Corpora Collection

https://corpora.uni-leipzig.de/en?corpusId=eng-za_web_2014

Ngram viewer

<https://books.google.com/ngrams>

Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati - PAISÀ

<https://www.corpusitaliano.it>

Sketch Engine

<https://www.sketchengine.eu>

WaCKy - Web as Corpus Kool Ynitiative

<https://wacky.sslmit.unibo.it/doku.php?id=corpora>

Notes

1. « *As a great big body of text waiting to be mined, a huge fabric of linguistic data often interwoven with parallel threads* » (notre traduction).

2. Le J.O. du 01/01/2013 recommande l'utilisation du terme *collecteur*.

3. La reconstruction du « mode d'existence » des objets techniques et scientifiques et l'identification de la « carrière objectale » aideraient aux processus de conceptualisation et de formation des termes (Simondon, 1958).

4. Le terme « candidat » doit être ici pris dans son sens originel de 'vêtu de blanc' [du lat. *candidatus*]. En effet, à l'époque de la Rome antique, les citoyens, aspirant à une fonction publique, portaient une toge blanche pour indiquer qu'ils étaient au-dessus de tout soupçon.

5. Le recours à l'opérateur *wildcard*, nous a permis de repérer une seule expression correspondant à la séquence *unità.... di macellazione* : *unità produttiva di macellazione*.

6. Entre parenthèses le premier chiffre se réfère au singulier et le deuxième au pluriel.

7. <https://www.gazzettaufficiale.it/eli/gu/1994/12/16/293/so/162/sg/pdf>

8. Selon le De Mauro, le mot *macello* a été introduit en Italie au XIV^e siècle du lat. *macĕllu(m)* dans le sens de 'marché de viande, poisson et légumes', alors que celui de *mattatoio* est passé en italien, au XIX^e siècle, par le verbe *mattare* dans le sens de 'tuer, mettre à mort'. On pourrait penser que la différence entre les deux termes était d'abord une affaire étymologique et plus tard une affaire de représentation mentale.

9. La date de dernière consultation des sites est le 4 février 2021.



GERFLINT

ISSN 1724-0700

ISSN en ligne 2260-8087

Les corpus et la traduction spécialisée.
Proposition d'un parcours didactique
centré sur la terminologie pour les étudiants
en traduction (FR - IT)

Klara Dankova

Université Catholique du Sacré-Cœur – Brescia/Milan, Italie

klara.dankova@unicatt.it

Reçu le 17-02-2021 / Évalué le 22-03-2021 / Accepté le 16-04-2021

Résumé

Mettant en évidence l'utilité des corpus électroniques pour la traduction, cet article examine la question de la construction et de l'emploi des corpus pour la traduction spécialisée, en proposant un parcours didactique centré sur la recherche terminologique et adressé à des étudiants italophones en traduction. Le parcours est illustré en exploitant un corpus (FR - IT) du domaine de l'urbanisme et de l'aménagement du territoire.

Mots-clés : traduction spécialisée, terminologie, corpus, concordance, urbanisme

**Corpora e traduzione specialistica. Proposta di un percorso didattico
sulla terminologia rivolto a studenti di traduzione (FR - IT)**

Riassunto

Mettendo in evidenza l'utilità dei corpora elettronici per la traduzione, il presente articolo esamina la questione della costruzione e dell'uso dei corpora per la traduzione specialistica, proponendo un percorso didattico sulla terminologia rivolto a studenti italofofoni di traduzione. Il percorso è illustrato utilizzando un corpus (FR - IT) del dominio della pianificazione urbanistica e territoriale.

Parole chiave: traduzione specialistica, terminologia, corpus, concordanza, urbanismo

**Corpora and specialized translation. Proposal for a teaching sequence
focused on terminology for translation students (FR - IT)**

Abstract

By highlighting the usefulness of electronic corpora for translation, this article examines the issues related to the construction and use of corpora for specialized translation. It also proposes a teaching sequence focused on terminological research for Italian-speaking translation students. The sequence is illustrated using a French-Italian corpus in the field of urban and spatial planning.

Keywords: specialized translation, terminology, corpus, concordance, urbanism

Introduction

Réservoirs de textes témoignant du fonctionnement de la langue en discours, les corpus représentent des ressources très utiles pour la traduction et, tout particulièrement, pour la traduction spécialisée. La spécificité de la traduction des textes techniques réside, entre autres, dans la nécessité d'identifier correctement les équivalents des termes utilisés. À part les recherches dans les dictionnaires techniques bilingues et dans les banques de données terminologiques, la terminologie spécialisée peut être découverte aussi en interrogeant des corpus spécialisés du domaine concerné. Toutefois, à l'exception de certains projets spécifiques tels que ceux de l'Union européenne fournissant des corpus relatifs au droit de l'UE¹ et ceux de *ScienQuest*, un outil permettant d'exploiter des corpus de textes scientifiques de plusieurs domaines en français et en anglais, les traducteurs ne disposent pas, à l'heure actuelle, de grands corpus de textes techniques en accès libre. Dans ce contexte, une solution possible au problème de la recherche terminologique s'avère être la construction d'un corpus spécialisé.

Le présent article vise à examiner la question de la construction et de l'emploi des corpus pour la traduction spécialisée, en proposant un parcours didactique centré sur l'analyse des termes et la recherche des équivalents, destiné à des étudiants italophones en traduction FR - IT, ayant un niveau de français avancé (au moins B2 ou C1 du CECRL). Le parcours se base sur la construction des corpus FR - IT concernant l'urbanisme et l'aménagement du territoire à l'aide de l'outil *Sketch Engine*. De grande actualité à cause de ses liens avec le développement durable, ce domaine a été choisi car les sujets traités peuvent être envisagés non seulement au niveau local, mais aussi au niveau international, ce qui entraîne des besoins en traduction.

1. L'emploi des corpus pour la traduction

Même si la question de l'utilisation des corpus électroniques, c'est-à-dire des « banques de données linguistiques exploitables de façon automatique » (Loock, 2016 : 1), pour la traduction est bien loin d'être nouvelle (voir Olohan, 2004 ; Frerot, 2010 : 3), les études récentes montrent que, du moins dans l'espace francophone, les corpus ne représentent pas encore des outils d'aide à la traduction utilisés fréquemment et surtout de façon consciente par les traducteurs professionnels (Loock, 2016 : 1-2). Selon R. Loock (2016 : 2-3), ce sont, en premier lieu, des aspects techniques et des interfaces non conviviales des corpus existants qui découragent les traducteurs de les exploiter et, par conséquent, d'en construire de nouveaux. Dans la formation des futurs traducteurs, il faut mettre en évidence

les avantages que les corpus peuvent apporter à la traduction et encourager les étudiants à s'en servir sans oublier de signaler aussi leurs limites, liées principalement à la structure des textes formant les corpus.

En ce qui concerne l'utilité de l'emploi des corpus pour la traduction, celle-ci s'explique avant tout par le caractère même du processus traductif : le traducteur ne considère pas des mots isolés, mais des séquences de textes plus larges en gardant à l'esprit que la signification d'un mot est toujours déterminée par le contexte dans lequel il se trouve (Teubert, 2010 : 31-32). Le rôle décisif du contexte est encore plus marquant dans le cas des textes de spécialité rédigés par des experts qui utilisent des termes précis renvoyant à des concepts différents en fonction du domaine. Du point de vue terminologique, le plus grand défi pour les traducteurs techniques est constitué par l'emploi des termes désignant des concepts innovants. En effet, le progrès scientifique et technique entraîne sans cesse l'apparition de néologismes dont le recensement - tant dans les ouvrages lexicographiques, que dans les banques de données terminologiques - est loin d'être immédiat. Dans ce contexte, l'utilisation d'un corpus électronique spécialisé contenant des textes techniques récents du domaine constitue une ressource privilégiée pour une recherche terminologique fructueuse et fiable (Colson, 2010 : 116).

Avant de nous concentrer sur la construction des corpus pour la traduction spécialisée, il faut distinguer les corpus comparables et les corpus parallèles. Tandis qu'un corpus comparable est constitué de textes originaux en deux langues concernant le même domaine de spécialité, un corpus parallèle comporte des textes en deux langues, à savoir ceux en langue source et leurs traductions en langue cible (Frerot, 2010 : 3). Pour ce qui est de l'emploi des corpus parallèles pour la traduction, il est important de souligner plusieurs problèmes d'ordre méthodologique et pratique, tels qu'un nombre insuffisant de textes de spécialité traduits pour un domaine donné et, surtout, l'influence de la langue source sur les choix linguistiques dans les textes en langue cible. À la différence des textes originaux, dans le cas des textes traduits, nous n'avons pas la garantie qu'ils représentent, à tous égards, l'usage ordinaire de la langue. Pour faire des généralisations basées sur la comparaison, il est indispensable de combiner l'utilisation des corpus parallèles et comparables (Johansson, 2010 : 55).

2. La construction d'un corpus pour la traduction spécialisée

En premier lieu, il est nécessaire de déterminer le plus précisément possible l'objectif poursuivi, ou bien pour la traduction de quels types de textes le corpus sera utilisé. En fonction de cet objectif, il faut considérer différents critères de choix des textes, dont principalement (Cabré, 2007-2008 : 38-40) :

- le thème ou le sujet : les corpus peuvent concerner une (*corpus monodisciplinaire*) ou plusieurs disciplines (*corpus pluridisciplinaire*) ;
- le niveau de spécialisation : les corpus peuvent comporter des textes d'un niveau de spécialité (par exemple, les manuels pour les élèves du secondaire) ou de plusieurs niveaux de spécialité (par exemple, les manuels pour les élèves du secondaire et ceux pour les étudiants universitaires);
- les sources : les textes du corpus peuvent provenir d'un seul type de source (par exemple, les textes écrits) ou de plusieurs types de sources (par exemple, les textes écrits et audiovisuels);
- le genre textuel : les corpus peuvent contenir des textes d'un seul genre (par exemple, les articles scientifiques) ou de plusieurs genres textuels (par exemple, les articles scientifiques et les critiques);
- la stratégie discursive des textes : les corpus peuvent comporter des textes présentant une seule stratégie discursive (par exemple, les textes argumentatifs) ou plusieurs stratégies discursives (par exemple, les textes argumentatifs et narratifs) ;
- la langue : les corpus peuvent être monolingues ou plurilingues.

Dans le cas présent, notre objectif est de constituer des corpus pour la traduction spécialisée avec une attention particulière à la terminologie. Par conséquent, le choix des textes doit respecter aussi les critères de sélection des sources de documentation pertinentes du point de vue terminologique, à savoir (Zanola, 2018 : 78) :

- documents originaux contenant la terminologie réellement en usage dans le domaine ;
- documents récents et représentatifs du domaine considéré ;
- documents de bonne qualité rédactionnelle écrits par des experts reconnus en la matière.

En particulier, il convient de se concentrer sur la recherche des sources de documentation « primaires » - telles que les publications scientifiques, les manuels techniques, les documents institutionnels et législatifs et la documentation technique officielle - car celles-ci contiennent la terminologie de référence utilisée dans le domaine (Zanola, 2014 : 2-3).

3. Les corpus pour la traduction dans le domaine de l'urbanisme et de l'aménagement du territoire

Nous proposons un parcours didactique concernant la construction et l'exploitation de corpus spécialisés pour la traduction constitué de trois étapes principales : la construction des corpus, l'exploitation des corpus comparables et l'exploitation

du corpus parallèle. L'exploitation des corpus sera guidée par les besoins concrets, liés à la traduction d'un texte vers l'italien. Pour l'illustration de la méthode, nous avons choisi un extrait (3 premières pages) de l'article de F. Tomas *Les temporalités du patrimoine et de l'aménagement urbain*.

Ce parcours nous l'avons expérimenté au cours d'un séminaire de traduction FR - IT à l'Université Catholique du Sacré-Cœur à Brescia. Précisons que pour des raisons pratiques, nous avons préféré effectuer les travaux préparatoires pour la construction des corpus par nous-mêmes. Néanmoins, une fois les corpus construits, il est toujours possible d'impliquer les étudiants en leur demandant de sélectionner et de nettoyer une série de textes pour l'élargissement et la mise à jour des corpus déjà existants.

3.1. La construction des corpus

Dans la première étape de caractère introductif (90 minutes), l'enseignant donne des informations de base sur les corpus en linguistique et explique la différence entre les corpus comparables et parallèles, en mettant en relief la nécessité d'effectuer des recherches en combinant les résultats provenant des deux types de corpus. Ensuite, il présente les critères qu'il faut prendre en considération en constituant un corpus pour la traduction et la recherche terminologique.

L'objectif de cette phase est de faire connaître aux étudiants les critères de sélection de textes aussi bien que le procédé de construction des corpus. L'enseignant décrira la création de trois mini-corpus, à savoir deux corpus monolingues pour l'extraction et l'analyse des termes, l'un contenant des textes en français et l'autre des textes en italien (corpus comparables), et un corpus parallèle composé de textes identiques en français et en italien. Étant donné qu'il s'agit des corpus dont la taille est assez limitée, l'enseignant devrait aussi mettre en évidence les limites de leur exploitation, dont notamment la présence d'un nombre restreint de termes du domaine. Si l'enseignant travaille avec un groupe d'étudiants assez important (au minimum 15 étudiants), il est possible de les impliquer davantage, en les faisant participer activement à la sélection et au nettoyage des textes (voir plus loin) pour la construction des corpus.

Notre objectif est de construire des corpus de textes contenant la terminologie de référence la plus actuelle possible du domaine de l'urbanisme et de l'aménagement du territoire. C'est pourquoi nous n'avons retenu que des documents appartenant au domaine (corpus monodisciplinaires) qui ont été publiés dans la période de 2015 à 2020. Le caractère officiel et contraignant de cette terminologie nous a amené à

nous limiter aux textes provenant de la France (pour le corpus monolingue français) et de l'Italie (pour le corpus monolingue italien). En particulier, ces corpus sont constitués de documents recueillis à partir des sites des institutions et des associations professionnelles du domaine opérant dans les deux pays : le corpus français (40 documents) contient des documents disponibles dans la *Base documentaire de la politique de la ville* et une série de publications de l'*Atelier parisien d'urbanisme*, tandis que le corpus de textes en italien (40 documents) réunit une sélection des publications disponibles sur les sites de l'*Associazione Nazionale degli Urbanisti e dei Pianificatori Territoriali e Ambientali* et de l'*Istituto Nazionale di Urbanistica*. Le corpus parallèle est construit en utilisant les documents de l'Union européenne relatifs à l'urbanisme et à l'aménagement du territoire en français (30 documents) et en italien (30 documents), téléchargeables à partir du site de l'*Office des publications de l'Union européenne*. Du point de vue terminologique, il s'agit de sources de documentation « primaires » incluant plusieurs typologies textuelles de niveaux de spécialisation différents (tableau 1) :

Typologies de documents	Corpus FR	Corpus IT	Corpus parallèle	
			Partie FR	Partie IT
article scientifique	1	30		
communication à un colloque (synthèse)		1		
compte rendu d'une journée d'échanges ou de formation	8			
document institutionnel - étude	11	4	3	3
- rapport annuel	9	1		
- informatif	7	2	9	9
- interne			2	2
document juridique	2	2	16	16
mémoire de master	2			
- nombre total de documents	40	40	30	30

Tableau 1 : La structure des corpus selon la typologie textuelle (corpus FR, corpus IT, corpus parallèle)

Au total, les corpus contiennent 6 typologies de documents : article scientifique, communication à un colloque, compte rendu d'une journée d'échanges ou de formation, document institutionnel, document juridique et mémoire de master. Pour une meilleure précision, au sein de la catégorie « document institutionnel »,

nous avons distingué 4 sous-catégories : document institutionnel sous forme d'étude, document institutionnel faisant partie d'un rapport annuel, document institutionnel informatif destiné au grand public (par exemple, documents décrivant l'état d'avancement d'un projet, dossiers de presse, etc.) et document institutionnel « interne » incluant les rapports d'une institution élaborés pour autre institution. Le tableau montre que la composition des corpus du point de vue de la typologie textuelle est assez déséquilibrée, ce qui nous empêche, pour l'instant, de mener des recherches approfondies sur l'emploi des termes en fonction du type de texte.

Les trois corpus (FR, IT, parallèle) ont été construits en utilisant le gestionnaire de corpus *Sketch Engine*, un logiciel mis au point par *Lexical Computing Limited* pour les recherches en lexicographie et en linguistique de corpus. Ces dernières années, cet outil a été développé davantage pour mieux répondre aux besoins des traducteurs et des terminologues, particulièrement en ajoutant de nouveaux corpus parallèles et en perfectionnant la concordance parallèle (Kilgarriff, 2013 : 6). Nous l'avons choisi car il permet une construction relativement facile des corpus, y compris des corpus parallèles, et une interrogation conviviale des données, notamment grâce aux fonctions *Word Sketch* et *Concordance* (voir plus loin).

Pour la construction des corpus monolingues, il est nécessaire de mettre les documents pdf au format texte, tout en éliminant les éléments qui pourraient altérer les résultats des recherches ou gêner la navigation dans le corpus, tels que les titres, les numéros de page, les notes en bas de page et d'autres notes, les graphiques, les tableaux, les légendes, les références bibliographiques et la sitographie. Une fois les fichiers .txt nettoyés, on peut procéder à la création du corpus dans *Sketch Engine* en utilisant la fonction correspondante (*Créer un corpus*), qui permet de compiler le corpus en l'annotant de façon automatique. Ajoutons qu'il convient de nommer les documents .txt de façon systématique en mentionnant certaines métadonnées, indispensables pour l'élargissement des corpus avec la création des sous-corpus et pour une évaluation plus pointue des résultats. Dans le cas présent, nous avons nommé les fichiers en indiquant leur numéro, la langue et le type du texte (par exemple, 01_IT_article scientifique, 17_FR_document juridique), car nous envisageons d'exploiter ces corpus, une fois devenus plus équilibrés, pour l'analyse de l'emploi des termes en fonction de la typologie textuelle. Précisons que les corpus ainsi créés sont disponibles seulement pour leur auteur dans son espace personnel. Pour pouvoir les exploiter en classe, il faut les partager avec d'autres utilisateurs.

Quant à la construction du corpus parallèle, les documents .txt nettoyés en français et en italien doivent être aussi alignés à l'aide d'un programme d'alignement. Nous avons utilisé le logiciel *Wordfast Online Aligner* disponible en ligne

gratuitement et sans téléchargement. Le résultat de l'alignement est un document *Excel* contenant des segments de texte en français et en italien alignés dans deux colonnes séparées. Ce fichier est ensuite chargé dans *Sketch Engine*, mais avant de le faire, il est nécessaire de vérifier l'exactitude de l'alignement automatique de façon manuelle. Enfin, à partir du document *Excel révisé*, *Sketch Engine* créera deux corpus - l'un en français et l'autre en italien - qui peuvent être utilisés de façon autonome aussi bien qu'en tant que corpus parallèles.

Dans la section dédiée, il est possible de repérer, pour chaque corpus, des données détaillées concernant sa composition, telles que le nombre de documents chargés, le nombre de phrases et le nombre de mots. La taille de nos corpus peut être déterminée de manière précise en indiquant le nombre de mots, à savoir (tableau 2) :

	corpus FR	corpus IT	corpus parallèle	
			partie FR	partie IT
nombre de mots	308.832	343.681	179.914	164.407

Tableau 2 : La taille des corpus (corpus FR, corpus IT, corpus parallèle)

3.2. L'exploitation des corpus comparables en français et en italien

Dans cette étape (120 minutes), les étudiants apprendront à exploiter les corpus monolingues pour la traduction, en particulier en ce qui concerne la recherche terminologique et l'analyse de l'emploi des termes. Dans la première partie (60 minutes), ils effectueront l'extraction terminologique en utilisant la fonction *Mots-clés* de *Sketch Engine* pour se familiariser avec les concepts clés du secteur. Sur la base du score *keyness* (*keyness score*), une valeur calculée en confrontant la fréquence des unités linguistiques dans nos corpus avec celle que ces unités linguistiques ont dans le corpus de référence², le logiciel identifie une liste de candidats termes. Celle-ci doit être nettoyée en éliminant le bruit, c'est-à-dire tous les éléments qui ne sont pas des termes du domaine, tels que les syntagmes libres et les termes appartenant à d'autres domaines. Cette tâche sera effectuée de manière collaborative par les étudiants divisés en quatre groupes pour extraire, respectivement, (1) les termes simples en français, (2) les termes simples en italien, (3) les termes complexes en français et (4) les termes complexes en italien. Les résultats de l'extraction seront ensuite partagés avec le reste de la classe. À titre illustratif, dans le tableau 3, nous indiquons les cinq termes complexes ayant le score *keyness* le plus élevé dans les deux corpus :

corpus FR	score keyness	fréquence	corpus IT	score keyness	fréquence
rénovation urbaine	557,4	512	rigenerazione urbana	140,1	106
renouvellement urbain	186,1	147	servizi ecosistemici	125	60
quartier prioritaire	121,3	79	spazio pubblico	106,3	130
tissu urbain	118,2	70	sistema portuale	90,6	45
logistique urbaine	80,9	35	verde urbano	68,1	41

Tableau 3 : Les 5 termes complexes avec le score *keyness* le plus élevé (corpus FR, corpus IT)

La deuxième partie (60 minutes) a pour objectif, à travers des exercices d'application, de faire acquérir aux étudiants une méthode d'analyse des termes en contexte, basée sur l'emploi de deux fonctions de *Sketch Engine* : *Word Sketch*, permettant d'identifier les combinaisons de mots présentes dans le corpus (par exemple, les collocations verbo-nominales V+N, les groupes nominaux N+ Adj, les groupes prépositionnels Prép. + N, etc.), et *Concordance*, fournissant toutes les occurrences du terme avec son contexte à gauche et à droite. Cette phase est très importante car elle permet de découvrir dans quel contexte et comment les termes analysés sont utilisés dans les textes originaux. Après la lecture du texte choisi pour la traduction (voir section 3), les étudiants seront invités à identifier les termes simples du domaine qui apparaissent souvent dans le texte. En fonction du nombre de termes relevés, les étudiants seront divisés en petits groupes pour identifier leurs équivalents en italien et pour analyser leur emploi en contexte, en particulier en ce qui concerne (1) les verbes formant avec ces termes des collocations V + N (COD) et (2) les termes complexes N + Adj. constitués à partir de ces termes simples. Premièrement, ils analyseront l'équivalent italien dans le corpus IT au moyen de la fonction *Word Sketch* et ils sélectionneront de la liste des résultats suggérés les collocations V + N (COD) et les termes complexes N + Adj. Une fois les analyses achevées, chaque groupe présentera les résultats obtenus en les illustrant avec des exemples concrets tirés du corpus en utilisant la fonction *Concordance*. À titre d'exemple, nous indiquons ci-après la liste de 5 verbes formant des collocations V + N (COD) avec *patrimonio*, l'équivalent italien de « patrimoine », un terme simple récurrent trouvé dans le texte, et la liste de 5 termes complexes N + Adj., formés à partir du même terme.

collocations V + N (COD)	fréquence	termes complexes N+ Adj.	fréquence
valorizzare il patrimonio	3	patrimonio edilizio	34
tramandare il patrimonio	1	patrimonio immobiliare	24
custodire il patrimonio	1	patrimonio abitativo	15
tutelare il patrimonio	1	patrimonio culturale	13
recuperare il patrimonio	1	patrimonio naturale	13

Tableau 4 : Les collocations V + N (COD) et les termes complexes N + Adj. formés avec « *patrimonio* » (corpus IT)

3.3. L'exploitation du corpus parallèle

Dans la dernière étape (60 minutes), les étudiants apprendront à utiliser la fonction *Concordance parallèle* permettant d'obtenir toutes les occurrences d'un terme dans les textes dans une langue, accompagnées des segments de textes correspondants dans l'autre langue. Tout d'abord, l'enseignant souligne la nécessité d'un regard critique sur les solutions de traduction proposées, dérivant du caractère traduit des textes et de leur origine (des documents de l'UE). La prudence s'impose notamment au niveau terminologique, car les termes utilisés dans les documents de l'UE ne sont pas nécessairement pertinents pour la traduction des textes rédigés dans un autre contexte.

Premièrement, on demandera aux étudiants, travaillant de nouveau en petits groupes, seront sollicités à utiliser la fonction *Concordance parallèle* pour trouver des équivalents en italien d'un terme intéressant du domaine repéré dans le texte analysé en vue de la traduction. Ensuite, pour observer l'usage de ces termes par les professionnels en italien, ils seront invités à interroger le corpus italien monolingue et, enfin, ils vérifieront les résultats obtenus dans un corpus de la langue générale plus large, tel que *Italian Web 2016* (4,9 billions de mots), disponible dans *Sketch Engine*. Cette dernière étape se conclut par le partage des résultats d'analyse de différents groupes.

Prenons l'exemple de *aménagement urbain*. Pour ce terme, les textes du corpus parallèle en italien proposent 6 solutions de traduction (selon la fréquence) : *sviluppo urbano* (9), *pianificazione urbana* (3), *pianificazione urbanistica* (2), *riassetto urbano* (2), *pianificazione* (1) et *assetto urbano* (1). Leur analyse dans le corpus italien monolingue montre que, à l'exception de *riassetto urbano* et *assetto urbano*, ces termes sont effectivement utilisés dans les textes italiens. La fréquence de *sviluppo urbano* est relativement élevée (18), ce qui confirme le résultat obtenu dans le corpus parallèle. Quant à la différence entre les termes

pianificazione urbana et *pianificazione urbanistica*, dans le corpus italien, la fréquence de *pianificazione urbanistica* (32) est nettement plus élevée que celle de *pianificazione urbana* (13). Les fréquences de *sviluppo urbano* (5 115), *pianificazione urbana* (1 946) et *pianificazione urbanistica* (6 051) dans le corpus *Italian Web 2016* confirment les tendances d'usage observées dans le corpus spécialisé monolingue. En ce qui concerne *riassetto urbano* et *assetto urbano*, ces termes sont bien présents dans *Italian Web 2016*, mais leurs fréquences (respectivement 82 et 907) indiquent que, par rapport aux autres équivalents proposés, l'usage de ces termes est relativement rare.

Conclusion

En utilisant l'exemple des corpus en français et en italien du domaine de l'urbanisme et de l'aménagement du territoire, nous avons élaboré un parcours didactique proposant une méthode de construction et d'emploi des corpus pour la traduction des textes de spécialité, complété par une série d'exercices, proposés à partir des besoins de traduction d'un texte concret. Actuellement, la taille modeste des corpus ne permet que des recherches limitées, mais il s'agit bien d'une base textuelle solide qui peut être enrichie dans l'avenir. Savoir construire et utiliser les corpus pour la traduction spécialisée peut se révéler très utile dans le futur professionnel des étudiants, notamment s'ils participent à des projets de traduction de longue durée ou s'ils souhaitent se spécialiser dans la traduction de textes relevant d'un domaine déterminé.

Bibliographie

- Cabré, M. T. 2007-2008. « Constituer un corpus de textes de spécialité ». *Cahiers du Centre Interlangue d'Études en Lexicologie*, p. 37-56.
- Colson, J.-P. 2010. « La traduction spécialisée basée sur les corpus : une expérience dans le domaine informatique ». *Synergie Tunisie*, n° 2, p. 115-123. [En ligne] : <https://gerflint.fr/Base/Tunisie2/colson.pdf> [consulté le 15 février 2021].
- Frerot, C. 2010. « Outils d'aide à la traduction : pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs ». *Les Cahiers du GEPE*, n° 2, p. 1-9. [En ligne] : <http://www.cahiersdugepe.fr/index.php?id=1164> [consulté le 15 février 2021].
- Johansson, S. 2010. Multilingual corpora: possibilities and limitations. In: *InterCorp: Exploring a Multilingual Corpus*. Prague: NLN, p. 53-68.
- Kilgarriff, A. 2013. Terminology finding, parallel corpora and bilingual word sketches in the *Sketch Engine*. In: *Proceedings ASLIB 35th Translating and the Computer Conference*. Londres. [En ligne] : https://www.sketchengine.eu/wp-content/uploads/2015/05/Terminology_finding_2013.pdf [consulté le 15 février 2021].
- Loock, R. 2016. « L'utilisation des corpus électroniques chez le traducteur professionnel : quand ? comment ? pour quoi faire ? ». *Ilcea*, n° 27, p. 1-23. [En ligne] : URL : <https://journals.openedition.org/ilcea/3835?lang=ru> [consulté le 15 février 2021].

- Olohan, M. 2004. *Introducing Corpora in Translation Studies*. Londres : Routledge.
- Teubert, W. 2010. Translation equivalence and interpretive corpus linguistics. In: *InterCorp: Exploring a Multilingual Corpus*. Praha : NLN, p. 31-52.
- Tomas, F. 2004. « Les temporalités du patrimoine et de l'aménagement urbain ». *Géocarrefour*, n° 79/3, p. 197-212. [En ligne] : <https://journals.openedition.org/geocarrefour/722?lang=fr#quotation> [consulté le 15 février 2021].
- Zanola, M. T. 2014. « Attività terminologica e fonti di documentazione ieri e oggi: problemi e metodi ». *MediAzioni*, n° 16, p. 1-16.
- Zanola, M. T. 2018. *Che cos'è la terminologia*. Rome: Carocci.

Sitographie³

Associazione Nazionale degli Urbanisti e dei Pianificatori Territoriali e Ambientali
<https://www.urbanisti.it>

Atelier parisien d'urbanisme
<https://www.apur.org/fr>

Base documentaire de la politique de la ville
<http://i.ville.gouv.fr/index.php>

Istituto Nazionale di Urbanistica
<https://www.inu.it>

Office des publications de l'Union européenne
<https://op.europa.eu/fr/home>

ScienQuest
<https://corpora.aiakide.net/scientext20>

Sketch Engine
<https://www.sketchengine.eu>

Wordfast Online Aligner
<https://www.wordfast.net>

Notes

1. <https://ec.europa.eu/jrc/en/language-technologies>
2. <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf>
3. La date de dernière consultation des sites est le 15 février 2021.



ISSN 1724-0700

ISSN en ligne 2260-8087

Corpus Pattern Analysis et Classes d'objets : différences théoriques et retombées pratiques de deux approches de description du lexique

Adriana Orlandi

Université de Modène et de Reggio d'Émilie, Italie
adriana.orlandi@unimore.it

<https://orcid.org/0000-0002-6638-265X>

Marco Fasciolo

Sorbonne Université, France
marco.fasciolo@sorbonne-universite.fr

Reçu le 10-02-2021 / Évalué le 22-03-2021 / Accepté le 07-04-2021

Résumé

Dans cette contribution, nous souhaitons comparer deux approches (théories, méthodes, techniques) lexicographiques conçues dans le cadre du TAL, à savoir la Corpus Pattern Analysis et la Théorie des Classes d'objets. En particulier, nous analyserons la différence théorique entre les notions de *pattern* et de *schéma prédicatif*, et sur le plan applicatif nous évaluerons la capacité des deux approches à discriminer les différentes acceptions d'un mot polysémique. Pour ce faire, nous prendrons en examen les *patterns* du verbe italien *seguire* extraits d'une base de données se basant sur la CPA, la T-Pas, et nous les comparerons avec les schémas prédicatifs du verbe *suivre* définis selon la Théorie des Classes d'objets.

Mots-clés : pattern, schéma prédicatif, Corpus Pattern Analysis, Théorie des classes d'objets, T-Pas

Corpus Pattern Analysis e Classes d'objets: differenze teoriche e ricadute pratiche di due approcci alla descrizione del lessico

Riassunto

In questo contributo ci proponiamo di confrontare due approcci (teorie, metodi, tecniche) lessicografici concepiti nell'ambito del TAL, ovvero la *Corpus Pattern Analysis* e la *Théorie des Classes d'Objets*. In particolare, analizzeremo la differenza teorica tra le nozioni di *pattern* e di *schéma prédicatif*, mentre, sul piano applicativo, valuteremo la capacità dei due approcci di discriminare le diverse accezioni di una parola polisemica. Per fare ciò, prenderemo in esame i *pattern* del verbo *seguire* estratti da un database realizzato secondo i principi della CPA, la risorsa T-Pas, e li confronteremo con gli schemi predicativi del verbo francese *suivre* definiti secondo la *Théorie des Classes d'Objets*.

Parole chiave: *pattern*, schema predicativo, *Corpus Pattern Analysis*, Teoria delle « classi di oggetti », T-Pas

Corpus Pattern Analysis and *Classes d'objets*: theoretical differences and practical implications of two approaches to lexicon description

Abstract

The aim of this paper is to compare two lexicographic approaches (theories, methods, techniques) conceived within the NLP framework, namely the *Corpus Pattern Analysis* and the *Théorie des Classes d'Objets*. In particular, we will analyse the theoretical difference between the notions of *pattern* and *schéma prédicatif*, and we will evaluate the capacity of the two approaches to discriminate the different meanings of a polysemic word. In order to do so, we will examine the patterns of the Italian verb *seguire* extracted from a database created according to the principles of CPA, the T-Pas resource, and compare them with the predicative patterns of the French verb *suivre* defined according to the *Théorie des Classes d'Objets*.

Keywords: pattern, *schéma prédicatif*, Corpus Pattern Analysis, *Théorie des classes d'objets*, T-Pas

1. Introduction : la *Théorie des Classes d'Objets* et la *Corpus Pattern Analysis*

La *Théorie des Classes d'Objets* (dorénavant *TCO*) est une approche lexicographique, française, conçue au sein du laboratoire *LLI*, ensuite *LDI*, par M. Mathieu-Colas et G. Gross, comme un développement de l'idée de lexique-grammaire de M. Gross. Une première synthèse de cette approche est offerte par Le Pesant et Mathieu-Colas (1998), mais l'idée est précédente : cf. par exemple, Gross (1992, 1994, 1995). Pour une présentation récente, nous mentionnerons Gross (2012), Fasciolo et Gross (2020). Dans cette contribution, nous nous proposons de comparer la *TCO* avec la *Corpus Pattern Analysis* (dorénavant *CPA*) à la base de la ressource T-Pas (Jezek *et al.*, 2014). Pour une présentation de la *CPA*, nous renvoyons à Hanks (2004, 2013), Chinkova et Hanks (2010).

L'idée centrale de la *TCO* est que la signification d'un mot doit être étudiée dans le cadre de la *phrase*. Or, il y a deux acceptions pertinentes de *phrase*. La première acception - mise en avant par la *TCO* - conçoit la phrase comme un pivot prédicatif (typiquement un verbe) saturé par ses arguments. La seconde acception - qui reste en arrière-plan dans la *TCO*, mais qui demeure cruciale - conçoit la phrase comme une architecture formelle GN □ GV au sens de l'analyse en constituants immédiats. Dans le cadre de la *TCO*, la meilleure façon pour étudier la signification d'un mot consiste à déployer une structure prédicative relevant de la première acception de phrase à l'intérieur de la structure formelle relevant de la seconde acception de phrase¹. Ce faisant, on obtient des *schémas prédicatifs* comme les suivants :

<humain> *suivre* <voie de communication> *jusqu'à* <lieu>

<événement> *suivre* <événement>

Les étiquettes <...> identifient des classes d'arguments conceptuellement homogènes, occupant les positions distributionnelles majeures du noyau de la phrase. Un *emploi prédicatif*, dans le cadre de la TCO, est défini comme le paquet de propriétés impliquées par chaque schéma prédicatif. Pour un schéma prédicatif donné, par exemple, un prédicat entre dans certaines relations lexicales (et pas d'autres), affecte certains rôles aux arguments (et pas d'autres), implique une certaine *Aktionsart*, peut être lié à certaines traductions, etc. Chacune de ces propriétés est définie exclusivement par rapport à un schéma prédicatif donné. Encore une fois, pour les détails, nous renvoyons à Gross (2012), Fasciolo (2019b) et à Fasciolo et Gross (2020).

La TCO distingue explicitement deux niveaux de classes d'arguments : les *hyper-classes* et les *classes d'objets stricto sensu*. Les *hyper-classes* sont des catégories *ontologiques*, universelles, qui délimitent des champs de cohérence conceptuelle *a contrario* : à savoir, plus la marge d'incohérence est grande, plus l'hyper-classe est positivement définie et ontologiquement centrale. Voici quelques exemples. Un verbe comme *peser* a une marge d'incohérence très petite car il s'applique à toute <entité matérielle> par opposition aux entités non-matérielles : *J'ai pesé la table, mon chien et mon collègue* vs. **J'ai pesé ma joie*. Un verbe comme *tuer* a une marge d'incohérence majeure car il oppose les <vivants> aux non-vivants : *J'ai tué mon collègue et son chien* vs. **J'ai tué la table*. Un verbe comme *scandaliser*, finalement, a une marge d'incohérence encore plus grande car il distingue les <humains> des non-humains : *J'ai scandalisé mon collègue* vs. **J'ai scandalisé mon chien, *J'ai scandalisé ma table*. Les *classes d'objets* au sens strict sont des classes *sémantiques* que le lexique circonscrit à l'intérieur des hyper-classes grâce à des faisceaux de *prédicats appropriés* : par exemple, *tailler* pour les <arbres>, *tondre* pour les <herbes>, *siroter* pour les <boissons>, etc. Ces classes d'objets sont extrêmement fines et sont les plus pertinentes pour la description lexicographique. Pour un approfondissement de la distinction entre hyper-classes et classes d'objets, on pourra se reporter à Prandi (1998), Gross (2008), Fasciolo (2018, 2019a, 2019b) et Fasciolo et Gross (2020).

Passons maintenant à la CPA :

En ce qui concerne les verbes, chaque pattern dans la CPA se base sur la structure des rôles phrastiques de l'anglais tels qu'ils sont décrits dans les grammaires. (...) Chaque rôle phrastique est « peuplé » par un ensemble paradigmatique de collocations (...). Les collocations pertinentes d'un verbe dans un rôle phrastique

sont habituellement des noms partageant quelques traits fondamentaux de leur signifié, à savoir un **type sémantique**. Les types sémantiques sont stockés dans une **ontologie superficielle** [*shallow ontology* dans le texte, ndt] organisée de manière hiérarchique. (...) Par exemple, un des patterns du verbe *translate*, (...) est le suivant² :

PATTERN: [[Human]] translate ([[Document]]) (from [[Language 1]]) (into [[Language 2]]) (Chinkova, Hanks, 2010: 3)

Le constat est vite fait : dans les deux cas, les notions de *valence* (ou structure argumentale) et de paradigme d'arguments jouent un rôle central. Les notions de *classes d'objets* et de *types sémantiques* d'une part, et les notions de *schéma prédicatif* et de *pattern* de l'autre semblent en effet très proches (cf. également Hanks, 2013 : 11 et Feltracco, 2016 : 70-71).

Il paraît donc légitime de se demander s'il existe des différences non marginales entre la *TCO* et la *CPA*. À notre avis, la réponse est positive. Ces différences, cependant, ne concernent pas tellement les résultats pratiques des deux approches, mais plutôt leurs présupposés théoriques et leurs implications méthodologiques. Dans la suite, nous contrasterons la *TCO* et la *CPA* sur deux points spécifiques :

- la différence entre les notions de *schéma prédicatif* et de *pattern* (§2)
- l'analyse d'une portion de la polysémie du verbe prédicatif *suivre/seguire* (§3).

2. *Pattern vs. Schéma prédicatif*

2.1. Qu'est qu'un *pattern* ?

La notion de *pattern* de la *CPA* hérite de celle de Hunston et Francis (2000) : une notion *corpus-driven*. Chinkova et Hanks (2010 : 1-2) insistent sur la déviation à partir de *patterns* normaux. Ces derniers sont induits par le corpus en se basant sur un présupposé partagé au sein des approches cognitivistes : à savoir, plus un phénomène est récurrent, plus il s'imprime dans l'esprit des parlants et plus il acquiert le statut d'objet linguistique saillant. Ce présupposé - explicité par exemple par Perek (2015 : 28-29) - est mobilisé par Hanks (2013 : 5) lorsqu'il distingue entre *saillance cognitive* et *saillance sociale* d'un phénomène. La première relève du caractère inoubliable d'une collocation, issu de son écart par rapport à une alternative normale. La seconde relève, justement, de la banalité des collocations normales : c'est-à-dire, tellement fréquentes qu'elles sont inaperçues par les individus. Un *pattern* (normal) au sens de la *CPA* est donc une collocation socialement saillante, enrichie par des types sémantiques (Hanks, 2013 : 92) : par exemple, *hasarder une hypothèse, une explication, une réponse*, etc³. Ce cadre

théorique a des implications non banales, nous semble-t-il, sur la conception de la valence.

Depuis que Tesnière (1959) a mis en avant cette notion, la séparation entre les arguments (ou actants) et les non-arguments a toujours été *la* question cruciale. La notion de valence est épistémologiquement solide si, et seulement si, pour une acception spécifique d'un verbe prédicatif, on peut établir *le* nombre de ses arguments. Dans la mesure où l'existence d'une limite tranchée entre arguments et non-arguments est remise en cause, parler de *valence d'un verbe* (par exemple) perd de pertinence.

Or, sur la base de ces prémisses, considérons un verbe prédicatif que les manuels classeraient comme tétra-valent : *vendre*, par exemple. Supposons, ensuite, que sa manifestation la plus fréquente en corpus - *saillance sociale* - soit avec seulement deux arguments. La question que nous posons est la suivante : à la lumière de la définition de *pattern* mentionnée ci-dessus, que faut-il en conclure à propos de la structure argumentale de *vendre* ?

La réponse est claire. Il n'y a pas quelque chose comme *le nombre* d'arguments de *vendre* : ce verbe n'est pas tétra-valent en soi, mais il est prototypiquement bivalent et il admet marginalement l'insertion d'arguments supplémentaires. Cela est ce que nous *montre* le corpus. Cette conclusion est parfaitement cohérente avec les prémisses et est tirée, par exemple, par Mereu (2017, 2020) et Mereu et Piunno (2019). Le problème, cependant, est que, par-là, la notion de valence se dissout car il n'y a plus des arguments et des non-arguments, mais des compléments accompagnant le verbe avec une fréquence majeure ou mineure. Or, selon nous, une telle conséquence est inévitable pour toute approche partageant la notion de *pattern* de la CPA. C'est précisément sur cet aspect que la TCO diverge de la CPA.

2.2. Qu'est-ce qu'un schéma prédicatif ?

Bien que la TCO soit centrée sur le pivot prédicatif - la source de la valence - la notion de *schéma prédicatif* hérite de la notion de *phrase* dans son acception distributionnelle et formelle (Harris, 1951). Cela se reflète sur la conception du *contexte*. Selon Gross (2010 : 191) :

Le contexte n'est jamais une donnée immédiate quand il s'agit de textes, qui ne sont jamais des suites de phrases simples (...) : des prédicats accompagnés de la suite la plus longue de leurs arguments. Et pourtant la phrase simple [= le schéma prédicatif] est l'unité la plus importante des textes.

La notion de *contexte* ci-dessus mentionnée n'est pas pragmatique : elle ne renvoie pas à une occurrence discursive ou textuelle, mais à la structure abstraite de la phrase - le *schéma prédicatif* - qui explicite *tous* les arguments du verbe. Cette structure - elle - n'est pas un fait du corpus, mais une donnée eidétique fonctionnant comme condition préalable pour interpréter les faits de corpus. En principe, rien n'empêche que les faits du corpus reproduisent parfaitement la structure du schéma prédicatif ; cependant, dans la pratique, cela n'arrive jamais. Typiquement, les faits de corpus se limitent à manifester des fragments d'un schéma prédicatif. C'est la raison pour laquelle Gross (2010 : *cit.*) affirme que les phrases simples sont les unités de base des textes, mais qu'elles ne sont pas, pour autant, *dans* les textes.

Nous retrouvons ici l'une des distinctions nobles de la linguistique : *langue* vs. *parole*. La notion de *schéma prédicatif* de la *TCO* relève du niveau de la *langue*, alors que la notion de *pattern* de la *CPA* relève de la *parole*. Prandi (2019) saisit clairement cette différence en concevant la phrase comme modèle de l'énoncé : en ses termes, la notion de *schéma prédicatif* relève du niveau de la *phrase-modèle*, alors que la notion de *pattern* relève du niveau l'énoncé. Ce bouleversement de cadre théorique a un impact sur la conception de la valence contraire à celui sous 2.1.

Considérons la question : *Quelle est la valence du verbe X ?* Dans le cadre de la *TCO*, pour répondre à cette question, il faut procéder en deux étapes. Premièrement, il faut explorer le corpus à la recherche d'attestations exhibant la suite de candidats arguments la plus longue : jusqu'ici, la *TCO* et la *CPA* procèdent en parallèle. Deuxièmement, une fois que cette suite a été identifiée, il faut évaluer le statut d'argument de chaque complément, *a priori* - indépendamment du corpus - avec une batterie de tests basés sur la cohérence conceptuelle. Pour une présentation de ces tests et une discussion de leur fonctionnement, nous renvoyons à Prandi (2004, 2007, 2013). Le point important, maintenant, est que la *TCO* et la *CPA* divergent par rapport à cette seconde étape.

Certes, dans la littérature issue de l'approche de la *CPA* (Jezek, 2018), on reconnaît que certains arguments peuvent ne pas être exprimés pour des raisons pragmatiques ou communicatives. Cependant, ce constat est souvent fait dans le cadre plus général d'une approche se proposant d'induire la valence *a posteriori* : à partir du corpus et de ses *patterns*. Or, c'est précisément cette démarche que la *TCO* n'adopte pas.

2.3. Le poids de la fréquence

Concevoir la structure argumentale à travers le prisme de la notion de *schéma prédicatif* - plutôt qu'à travers celui de la notion de *pattern* - signifie envisager

la valence comme *a priori* par rapport au corpus. Cette perspective implique que la fréquence d'un complément n'est pas pertinente pour déterminer son statut d'argument.

Supposons qu'en corpus le verbe *vendre* (*quelque chose à quelqu'un*) apparaisse fréquemment sans le destinataire, mais avec seulement le vendeur et le bien. Supposons qu'en corpus le verbe *parler* (*avec quelqu'un de quelque chose*) apparaisse fréquemment sans mentionner ni l'interlocuteur, ni le propos, et qu'il soit en revanche accompagné par des adverbiaux comme *à tort et à travers* ou *à voix basse*. Si nous raisonnons en termes de *pattern*, force en est d'en conclure que la valence prototypique de ces verbes est « SUJET_{VENDEUR} + vendre + COD_{BIEN} » et « SUJET_{LOCUTEUR} + parler + ADVERBE_{MANIÈRE} ». D'une part, le destinataire de la vente cesse d'être un argument prototypique de *vendre* ; de l'autre, la manière devient un argument prototypique de *parler*, à la place de l'interlocuteur et du propos. Si nous raisonnons en termes de *schéma prédicatif*, cette conclusion n'a aucun sens. D'un point de vue communicatif, dans *vendre un appartement* et *parler à voix basse*, l'expression du bien vendu et du volume de la voix peut être plus importante que l'expression du destinataire de la vente et de l'interlocuteur de l'acte de parole ; cependant, cela n'implique ni que le destinataire cesse d'être un argument de *vendre*, ni que le volume devienne un argument de *parler*. De même, au niveau de la fréquence, nous voyons les personnes plus souvent habillées que nues ; pourtant, nous n'en concluons ni que les habits sont des parties prototypiques du corps de la personne, ni que la distinction entre corps et habits est une question de *continuum*.

La notion de *schéma prédicatif* ne vise pas à minimiser les faits de fréquence qu'on vient d'évoquer ; plus simplement, elle nous invite à séparer deux critères logiquement indépendants : un critère de *pertinence communicative* d'un complément d'une part et un critère de *pertinence idéationnelle* de ce complément (son statut d'argument) de l'autre. Le premier légifère à propos de ce qu'il est pertinent d'omettre ou d'exprimer lorsqu'on énonce un procès pour communiquer un message ou pour l'insérer dans un texte : ce critère doit donc être étudié *a posteriori*, à partir du corpus. Le second légifère à propos des participants logiquement essentiels, ou pas, pour concevoir un procès *avant* son emploi pour communiquer un message ou son insertion dans un texte : ce critère doit donc être étudié *a priori*. Les modifications issues de la pertinence communicative tiennent pour acquise la structure argumentale établie par la pertinence idéationnelle.

3. Discrimination des acceptions : le cas de *seguire-suivre*

Au-delà des différences qui caractérisent les deux approches sur le plan théorique et méthodologique, la CPA et la TCO partagent une même visée applicative lexicographique : la réalisation de dictionnaires électroniques pour le traitement automatique du langage⁴. À ce niveau, la détermination des *patterns* et des *schémas prédicatifs* constitue le moyen privilégié pour discriminer les différentes acceptions d'un mot polysémique. Le travail sur les corpus a en effet montré l'impact de la polysémie sur la variété des structures argumentales possibles. Considérons par exemple les schémas prédicatifs du verbe *suivre* isolés suivant la TCO, et les *patterns* du verbe *seguire* dans la ressource T-Pas, isolés moyennant la CPA, et comparons-les⁵. Pour des raisons d'économie, nous ne considérerons que les acceptions de *seguire-suivre* en tant qu'« activité » (*s. qqn de près, s. un chemin, s. une règle, s. une explication, s. un patient, etc.*), laissant de côté les acceptions de *seguire-suivre* en tant que « relation » (*le mardi suit lundi, un chapitre suit un autre, etc.*). Aussi, nous nous bornerons à comparer les acceptions qui sont réellement partagées par les deux langues, écartant celles qui n'existent que dans l'une des deux (*ex. s. un produit, au sens de « continuer à le produire », ne se dit qu'en français*).

La première remarque que l'on peut formuler est que les schémas prédicatifs sont plus nombreux que les T-Pas (*patterns*). On compte en effet 8 T-Pas pour *seguire* et 16 schémas prédicatifs pour *suivre*, le double. Les T-Pas de *seguire* isolent les acceptions principales du verbe (la numérotation est la nôtre) :

- 1: [[Animate1]] segue [[Animate2]]
- 2: [[Animate]] segue [[Route]]
- 3: [[Human]] segue [[Concept | System]]
- 4: [[Human]] segue [[Concept]]
- 5: [[Human]] segue [[Activity]]
- 6: [[Human]] segue [[Event]]
- 7: [[Human1]] segue [[Human2]]
- 8: [[Human1 | Human Group1]] segue [[Human2 | Human Group2]]

Les schémas prédicatifs de *suivre*, en plus d'isoler les acceptions principales du verbe, définissent également quelques-unes de ce qu'un dictionnaire de langue classerait habituellement comme des sous-acceptions. Par exemple, la T-Pas 2 ci-dessus ([[Animate]] segue [[Route]]) identifie *grosso modo* la même acception que le schéma prédicatif suivant (nous ne reproduisons par commodité que le schéma suivi d'un synonyme, du domaine d'emploi et d'un exemple) :

suivre/N0:hum(métonymie intégrée sur voiture)/N1:loc<voies>/N2:jusqu'à loc
 Sy : parcourir
 D : véhicule
 Ex : *On suit le boulevard jusqu'à la place*

Il s'agit de l'acception « parcourir un chemin ». On remarquera au passage que l'acception est tout de même un peu plus spécifique dans la *TCO* (l'agent est un être humain, idéalement un chauffeur) et que le schéma prédicatif est plus complet (il comprend également N2, la destination). Nous y reviendrons.

Ce qui nous intéresse ici de souligner est que la *TCO* fournit également deux autres emplois voisinant le sens de « parcourir un chemin », mais qui s'en différencient par un schéma prédicatif spécifique :

suivre/N0:hum(métonymie intégrée sur voiture)/N1:loc<point de repère/signal>/N2:jusqu'à loc
 Sy : adopter orientation
 D : véhicule
 Ex : *On suit le nord / l'indication / la boussole jusqu'à la place*

suivre/N0:inc<moyens de transport/voies>/N1:loc<voies/côtes>
 Sy : longer
 D : véhicule
 Ex : *Le train suit la route. La frontière suit le fleuve*

Les (sous-)accepions mentionnées ci-dessus concernent respectivement l'action faite par un être humain d'adopter une orientation, et l'action faite par un moyen de transport ou un élément du paysage de longer une voie ou un autre élément du paysage. D'un point de vue conceptuel, toutes ces accepions représentent autant de déclinaisons de l'acception principale « parcourir un chemin » ; elles ne sont pas présentes dans la base de données T-Pas.

La deuxième remarque que l'on peut avancer concerne le degré de précision des *patterns* vis-à-vis des schémas prédicatifs. En comparant les *patterns* de *suivre* avec les schémas prédicatifs de *suivre*, on constate que les schémas prédicatifs sont souvent plus précis et plus spécifiques que les T-Pas.

Tout d'abord, les schémas prédicatifs sont plus complets pour les raisons que nous avons illustrées *supra* (§ 2.2), c'est-à-dire qu'ils ne sont pas fondés sur la fréquence mais sur la valence en dehors du contexte d'utilisation. C'est le cas, comme nous l'avons vu, pour la T-Pas 2 ([[Animate]] segue [[Route]]), mais aussi pour les T-Pas 1 ([[Animate1]] segue [[Animate2]]) et 7 ([[Human1]] segue [[Human2]]). Prenons la T-Pas 7. La définition de ce *pattern* est confiée à l'implicature suivante :

[[Human1]] sta dietro con la mente o nel ragionamento a [[Human2]]

(littér., « va derrière avec l'esprit ou dans le raisonnement »)

Le schéma prédicatif correspondant dans la *TCO* est :

suivre/N0:hum/N1:hum/N2:dans ina<raisonnement>

On voit bien que l'argument N2 (dans ina<raisonnement>, où <raisonnement> est une classe d'objets) est absent dans la T-Pas 7 ; il n'est présent qu'au niveau de l'implicature.

Ensuite, les classes d'objets qu'utilise la *TCO* pour définir les emplois de *suivre* ont la caractéristique de s'appliquer à des classes plus restreintes de référents par rapport aux types sémantiques utilisés par la *CPO*. L'extension des arguments étant plus limitée, le sens est souvent plus précis. Tel est le cas pour les T-Pas 3, 5, et 8.

Prenons en guise d'exemple la T-Pas 5 :

T-Pas 5 : [[Human]] segue [[Activity]]

Implicature : [[Human]] assiste o partecipa allo svolgersi di [[Activity]]

(littér., « assiste ou participe au déroulement de »)

Le schéma prédicatif correspondant dans la *TCO* est plus spécifique, car ses arguments ont une extension plus limitée :

suivre/N0:hum/N1:ina<spectacle télévisé>

Sy : regarder

D : général

Ex : *On suit un match, une série*

suivre/N0:hum/N1:ina<cours>

Sy : participer à, assister à

D : enseignement

Ex : *On suit un séminaire, un master, un stage*

suivre/N0:hum/N1:ina<traitement>

Sy : se soumettre à, subir un

D : médecine

Ex : *On suit une cure, un régime*

La T-Pas 5 isole un signifié plus large par rapport aux trois schémas prédicatifs ci-dessus, car, comme le dit l'implicature, le *pattern* [[Human]] segue [[Activity]] s'applique tant aux situations où l'individu assiste au déroulement d'une activité (ex. *Paul suit un match à la télé*), qu'aux situations où l'individu participe activement au déroulement d'une activité (ex. *Paul suit un cours d'anglais*).

En revanche, les schémas prédicatifs isolés par la *TCO* distinguent plusieurs cas de figure, comme le montrent bien la classe d'objets N1, le synonyme et l'exemple. Il s'agit d'acceptations plus restreintes.

Pour finir, nous aimerions souligner le rôle clé du domaine d'emploi dans le cas de la *TCO*. Le domaine d'emploi est prévu aussi dans la description des T-Pas, mais au moment de l'extraction des *patterns* à partir de la base de données, les domaines n'étaient pas fournis, ce qui nous empêche de faire une comparaison avec T-Pas sur ce point. Dans la *TCO*, la présence du domaine d'emploi est fondamentale. Prenons la T-Pas 8 :

T-Pas 8 : [[Human1 | Human Group1]] segue [[Human2 | Human Group2]]
 Implicature: [[Human1 | Human Group1]] mostra attenzione in favore di, assiste, aiuta [[Human2 | Human Group2]]
 (littér., « est attentif vis-à-vis de, assiste, aide »)

Or, considérons la perspective de la *TCO* et imaginons d'appliquer des domaines tels que la médecine et l'enseignement aux types sémantiques [[Human1 | Human Group1]] et [[Human2 | Human Group2]], appelés *hyper-classes* dans la *TCO*. Les domaines permettent de restreindre l'extension des hyper-classes, donnant lieu aux schémas d'arguments suivants :

suivre/N0:hum<soignant>/N1:<patient>

Sy : soigner, surveiller

D : médecine

Ex : *Le médecin suit le patient*

suivre/N0:hum<professeur>/N1:<étudiant>

Sy : surveiller

D : enseignement

Ex : *Le professeur suit ses élèves*

Par rapport à la T-Pas 8, les schémas prédicatifs issus de la *TCO*, tout en sélectionnant la même acception du verbe *suivre*, sont beaucoup plus spécifiques. Remarquons que les étiquettes des classes d'objets se confondent avec des rôles : les humains dans le domaine médical, en effet, sont typiquement des soignants et des patients. Un autre exemple montrant l'importance des domaines d'emploi pour la *TCO* est le suivant (pour cet exemple, nous faisons référence au sens de *suivre* en tant que « relation » et pas en tant qu'« activité ») :

suivre/N0:<événement>/N1:<événement>

Sy : venir après

D : général

Ex : *Un beau coucher de soleil a suivi la pluie*

suivre/N0:<événement>/N1:<événement>

Sy : être concomitant à, aller en parallèle avec

D : économie

Ex : *L'offre n'a pas suivi la demande*

Les domaines diffèrent par rapport aux hyper-classes pertinentes. Dans le cadre du langage général, un événement est par exemple un phénomène atmosphérique, mais dans le cadre de l'économie un événement est une variation de valeur comme la hausse de la demande, des prix etc. La *TCO* porte un intérêt tout particulier vers la description des langages de spécialité (d'où l'importance des domaines d'emploi), des langages qui, au sens de G. Gross, sont vus comme autant d'îlots formant l'archipel du langage tout court.

Conclusion

Notre comparaison entre la *TCO* et la *CPA* peut être résumée comme suit. Le *pattern* de la *CPA* est un outil qui est censé décrire le corpus *et*, en même temps, il est une donnée induite du corpus. Le *schéma prédicatif* de la *TCO*, lui aussi, est un outil qui est censé décrire le corpus, *mais* il n'est pas une donnée du corpus. Nous sommes confrontés, par-là, à une différence épistémologique majeure entre les deux approches, qui concerne leur rapport vis-à-vis du corpus.

Si le *pattern* est un outil pour décrire le corpus *et* une donnée du corpus, alors, dans le cadre de la *CPA*, le corpus contient en lui-même les conditions de possibilité de sa propre connaissance. Si le *schéma prédicatif* est un outil pour décrire le corpus, *mais* il n'est pas une donnée du corpus, alors, dans le cadre de la *TCO*, le corpus ne contient pas en lui-même les conditions de possibilité de sa propre connaissance. Ces dernières se situent au niveau de structures eidétiques - les *schémas prédictifs* - qui transcendent le corpus.

De ce point de vue, ce n'est pas surprenant que, dans le cadre de la *CPA*, on parle de la signification en termes d'événements ayant lieu en corpus (*i.e.* des énoncés) :

Qu'est-ce que, alors, le signifié d'un mot ? Il est raisonnable de croire que, dans l'utilisation quotidienne de la langue, les signifiés sont des événements et non des entités (...). Est-ce que les signifiés existent en dehors des contextes

*transactionnels dans lesquels ils sont utilisés ? Je dirais que non*⁶. (Hanks, 2013 : 73) [c'est nous qui soulignons].

Le schéma prédicatif de la *TCO*, en revanche, n'est pas un événement qui se produit en corpus, mais précisément une *entité* - un objet d'ordre eidétique stable et partagé - fonctionnant comme condition préalable pour donner du sens aux événements du corpus.

À bien y regarder, la *CPA* s'avère ainsi affectée par une curieuse contradiction. D'une part, elle se propose comme une théorie du lexique : donc des signifiés stablement partagés par les parlants. De l'autre, elle s'appuie sur l'acception de *signification* (*meaning*) mise en avant par Grice (cf. Hanks, 2013 : 88), qui relève du niveau conversationnel, pragmatique, et donc contingent. Cela, cependant, n'est pas totalement surprenant. Dans le cadre de la *CPA*, la seule façon pour penser le partage des signifiés est dans les termes des *patterns* d'énonciations les plus fréquents. En revanche, dans le cadre de la *TCO*, le partage des signifiés n'est pas ancré dans la régularité de *patterns énonciatifs*, mais dans le partage des *schémas prédicatifs*, qui sont des structures *a priori* présupposées par les différents *patterns* énonciatifs.

Sur le plan applicatif, la *CPA* pourrait tirer profit des travaux menés au sein du laboratoire LDI : la description des schémas prédicatifs étant plus fine dans la *TCO* que dans la *CPA*, cela augmenterait le pouvoir de désambiguïsation des *patterns*. Il serait également intéressant de comparer l'ontologie de la *CPA* avec les classes d'objets et les hyper-classes de la *TCO*. Puisque la description des signifiés verbaux s'appuie sur la description des éléments occupant la place d'argument, il est aisé de comprendre que le succès d'un outil électronique visant au traitement automatique du langage ne peut que dépendre d'une description fiable et soignée des classes d'objets ainsi que des types sémantiques.

Pour finir, il a été récemment démontré (Giacomini, Di Muccio-Failla, 2019) que les résultats obtenus par l'application de la méthode *CPA* pourraient être sensiblement améliorés en adoptant une approche « mixte » qui ne se base pas seulement sur l'examen des concordances, mais aussi sur l'extraction de collocations, ainsi que sur l'exploitation des dictionnaires déjà existants. Il est donc souhaitable que les approches et les méthodes existantes en matière d'analyse de corpus sortent des frontières nationales et entament un processus de connaissance réciproque.

Bibliographie

Buvet, P.A., Grezka, A. 2009. « Les dictionnaires électroniques du modèle des classes d'objets ». *Langages*, 176/4, p. 63-79.

- Chinkova, S., Hanks, P. 2010. « Validation of Corpus Pattern Analysis - Assigning pattern numbers to random verb samples », Octobre.
- Fasciolo, M. 2018. « Les noms du fond, ou la fonction des noms dits 'sommitaux' ». *Congrès Mondial de Linguistique Française*, en ligne.
- Fasciolo, M. 2019a. « Faut-il distinguer le lexique de l'ontologie ? » *Langue Française*, 204/4, p. 21-35.
- Fasciolo, M. 2019b. « The syntax of the lexicon : classes of objets and uses of predicates ». *PhrasaLex*.
- Fasciolo, M., Gross, G. 2020. *La sintassi del lessico*. Torino: UTET.
- Feltracco, A. 2016. « T-PAS: Building a Corpus-based Reference Tool for Italian ». *RiCOGNIZIONI*, 3/6, p. 69-78.
- Giacomini, L., Di Muccio-Failla, P. 2019. « Investigating Semi-Automatic Procedures in Pattern-Based Lexicography ». In : *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 conference*, p. 490-505.
- Gross, G. 1992. « Forme d'un dictionnaire électronique ». In : Clas, A. et Safar, H. (dir.) *L'environnement traductionnel*. Sillery : PUQ.
- Gross, G. 1994. « Classes d'objets et description des verbes ». *Langages*, 28/115, p. 15-30.
- Gross, G. 1995. « Une sémantique nouvelle pour la traduction automatique : les classes d'objets ». *La tribune des industries de la Langue et de l'information électronique*, 17-18-19, Paris.
- Gross, G. 2008. « Les classes d'objets ». *Lalies*, p. 111-165.
- Gross, G. 2010 « Sur la notion de contexte ». *Meta*, 1, 55, p. 187-198.
- Gross, G. 2012. *Manuel d'analyse linguistique*. Villeneuve d'Ascq : PUS.
- Hanks, P. 2004. « Corpus Pattern Analysis ». In : *Euralex Proceedings*. Vol. I, Lorient : Université de Bretagne-Sud, p. 87-98.
- Hanks, P. 2013. *Lexical Analysis*. Cambridge Mass : MIT Press.
- Hanks, P. et Pustejovsky, J. 2005. « A Pattern Dictionary for Natural Language Processing ». *Revue française de linguistique appliquée*, 10/2, p. 63-82.
- Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago-Londres : UCP.
- Hunston, S., Francis, G. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam - Philadelphie : John Benjamins.
- Jezek, E. 2018. « Partecipanti impliciti nella struttura argomentale dei verbi ». In : Dallabrida, S. et Cordin, P. (dir.) *La grammatica delle valenze. Spunti teorici, strumenti e applicazioni*. Firenze : Franco Casati Editore, p. 55-72.
- Jezek, E. Magnini, B., Feltracco, A., Bianchini, A. et Popescu O. 2014. T-Pas : « A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing ». In : Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik.
- Le Pesant, D., Mathieu-Colas, M. 1998. « Introduction aux classes d'objets ». *Langages*, 32/131, p. 6-33.
- Mereu, L. 2017. « La struttura argomentale in un approccio « usage-based » ». *Studi e saggi linguistici*, 55/2, p. 36-65.
- Mereu, L. 2020. *Semantica della frase*. Rome : Carocci.
- Mereu, L., Piuino, V. 2019. « The Argument Structure of Verbs of Hitting and Breaking in Italian ». *Lingue e linguaggio*, 18/1, p. 143-176.
- Perek, F. 2015. *Argument Structure in Usage-Based Construction Grammar*. Amsterdam-Philadelphie : John Benjamins.
- Prandi, M. 1998. « Contraintes conceptuelles sur la distribution : réflexions sur la notion de classe d'objets ». *Langages*, 32, 131, p. 34-44.
- Prandi, M. 2004. *The building blocks of meaning*. Amsterdam-Philadelphie : John Benjamins.

- Prandi, M. 2007. « Les fondements méthodologiques d'une grammaire descriptive de l'italien ». *Langages*, 3/167, p. 70-84.
- Prandi, M. 2013. *L'analisi del periodo*. Roma : Carocci.
- Prandi, M. 2019. Phrase et énoncé. De l'ordre symbolique à l'ordre indexical. In : Neveu, F. (dir.) *Proposition, phrase, énoncé*. Londres : ISTE éditions, p.131-154.
- Tesnière, L. 1959 [1966] *Éléments de syntaxe structurale* [2 éd.]. Paris : Klincksieck.

Notes

1. Cela n'implique pas que la structure formelle de la phrase soit complètement isomorphe par rapport à la structure conceptuelle prédicats-arguments, mais que la première (qui relève d'une « syntaxe *au-delà* du lexique ») permet d'exhiber de la façon la plus claire la seconde (qui relève d'une « syntaxe *du* lexique »). Sur ce point, cf. Fasciolo et Gross (2020).
2. « Each verb pattern in CPA is based on the structure of English clause roles described in systemic grammar (...) Each clause role in a pattern is 'populated' by a paradigm set of collocations (...). The relevant collocations of a verb in a clause role are usually nouns that share some basic aspect of their meaning, which can be expressed as a **semantic type**. Semantic types are stored in a hierarchically structured **shallow ontology**. (...) For instance, one of the patterns of the verb *translate* (...) is as follows: (...) ».
3. Dans la terminologie de Hanks (2013 : 10-11), *hasarder une hypothèse* est un *pattern* « normal », alors que *hasarder une explication, une réponse* sont des « exploitations » de cette norme.
4. La CPA a inspiré trois projets lexicographiques concernant respectivement l'anglais, l'espagnol et l'italien : le PDEV (*Pattern Dictionary of English Verbs*), coordonné par Hanks (cf. Hanks et Pustejovsky, 2005), le projet *Verbario*, coordonné par le Tecling, Université pontificale catholique de Valparaíso (cf. <http://www.tecling.com/index.php?l=verbario>), le T-Pas, coordonné par Jezek (cf. Jezek *et al.*, 2014). La TCO est à la base d'un dictionnaire électronique du français (cf. Gross, 1992 et plus récemment Buvet et Grezka, 2009). Tous ces projets visent au traitement automatique du langage.
5. Il aurait été préférable de comparer les deux approches par des exemples relevant d'une seule langue, mais les applications de la CPA et de la TCO les plus consolidées portent sur des langues différentes : l'anglais, l'espagnol et l'italien pour la CPA, le français pour la TCO (cf. note 4). Les deux approches ne sont donc pas comparables à partir d'une seule et même langue. C'est pourquoi nous avons fait le choix de comparer le verbe *seguire*, dont les *patterns* étaient disponibles à la plateforme tpas.fbk.eu, avec le verbe *suivre*, dont les schémas prédicatifs nous ont été fournis par G. Gross, que nous tenons à remercier.
6. « What, then, is a word meaning ? It is reasonable to assume that in everyday use of language, meanings are events, not entities (...). Do meanings exist outside the transactional contexts in which they are used? I would argue that they do not ».

Synergies Italie n° 17 / 2021



Varia





L'enseignement/apprentissage des unités phraséologiques en terminologie : un parcours didactique en français L2 sur le commerce international

Silvia Calvi

Université de Vérone, Italie

silvia.calvi@univr.it

Reçu le 08-02-2021 / Évalué le 08-03-2021 / Accepté le 10-03-2021

Résumé

Nous nous proposons d'approfondir la question de l'enseignement/apprentissage des unités phraséologiques de type *locution* et *collocation*, en travaillant sur la terminologie du commerce international. Nous allons illustrer une proposition didactique adressée aux étudiants universitaires de français L2 ayant au moins un niveau C1 du CERCL. Pour la préparation du matériel didactique nous avons exploité le corpus DIACOM-fr créé au Département de Langues et Littératures étrangères de l'Université de Vérone. Notre proposition didactique vise à développer des compétences métaterminologiques, méthodologiques et terminologiques.

Mots-clés : français L2, terminologie, corpus spécialisé, unités phraséologiques, commerce international

L'insegnamento/apprendimento delle unità fraseologiche in terminologia: un percorso didattico in francese L2 sul commercio internazionale

Riassunto

Il presente articolo intende approfondire la questione dell'insegnamento/apprendimento delle unità fraseologiche di tipo *locuzione* e *collocazione*, lavorando sulla terminologia del commercio internazionale. Illustreremo una proposta didattica rivolta a studenti universitari di francese L2, con un livello minimo C1 del QCER. Per la preparazione del materiale didattico abbiamo utilizzato il corpus DIACOM-fr, creato presso il Dipartimento di Lingue e Letterature straniere dell'Università di Verona. La nostra proposta didattica si pone come obiettivo lo sviluppo di competenze metaterminologiche, metodologiche e terminologiche.

Parole chiave: francese L2, terminologia, corpus specialistico, unità fraseologiche, commercio internazionale

Teaching/learning phraseological units in terminology: a didactic proposal in French L2 on international trade

Abstract

We propose to investigate the teaching/learning of phraseological units, more specifically of *idioms* and *collocations*, in the terminology of international trade. We will illustrate a didactic proposal addressed to university students of French L2 with at least a C1 level from the CEFR. For the preparation of the didactic material we have used the corpus DIACOM-fr, created at the Department of Foreign Languages and Literatures of the University of Verona. Our didactic proposal aims at developing metaterminological, methodological and terminological competences.

Keywords: French L2, terminology, specialized corpus, phraseological units, international trade

Introduction

L'enseignement/apprentissage du lexique et des connaissances métalinguistiques en L2 revêt une importance indéniable en langue générale et en langue de spécialité. Plusieurs études s'intéressent à la question de l'enseignement/apprentissage des unités phraséologiques, dont la connaissance permet aux apprenants de développer leurs compétences linguistiques au niveau de la production et de la compréhension écrite et/ou orale (Cavalla, Labre, 2009 ; Frassi, 2018 ; Bonadonna, Zollo, 2019).

Dans cet article, nous nous occupons de la question de l'enseignement/apprentissage des unités phraséologiques en langue de spécialité (français L2), avec une attention particulière au domaine du commerce international. Nous proposons un parcours didactique s'adressant aux étudiants universitaires de français L2 ayant au moins un niveau linguistique C1 du CERCL. Cette étude s'inscrit dans un cadre de réflexions théoriques et pratiques menées au sein du Département de Langues et Littératures étrangères de l'Université de Vérone qui a obtenu en janvier 2018 un financement de la part du Ministère italien de l'Éducation, de l'Université et de la Recherche pour le projet *Le Digital Humanities applicate alle lingue e letterature straniere (Les Humanités Numériques appliquées aux langues et littératures étrangères)*. Plus particulièrement, l'équipe de recherche DIACOM-fr dont je fais partie est en train de réaliser un corpus de langue française concernant le domaine du commerce international.

Notre travail se compose de trois parties : la présentation de notre approche théorique, l'illustration du corpus DIACOM-fr et la proposition d'un parcours

didactique que nous avons développé dans le but ultime de démontrer qu'une séquence didactique, qui s'appuie sur l'emploi d'un corpus en langue de spécialité, permet de développer plusieurs compétences, notamment des compétences métaterminologiques, méthodologiques et terminologiques.

1. Prémises théoriques

Dans cette étude, nous adoptons la définition et le classement des *unités phraséologiques* proposés par la Lexicologie Explicative et Combinatoire (dorénavant LEC), branche lexicale de la Théorie Sens-Texte (Mel'čuk, Clas, Polguère, 1995 ; Mel'čuk, 2013). Les unités phraséologiques, ou *phrasèmes*, sont des énoncés multilexémiques non libres sur l'axe paradigmatique et/ou syntagmatique. La classification des unités phraséologiques dépend des contraintes sur ces deux axes qui désignent respectivement les contraintes de sélection et de combinaison. Selon les contraintes de sélection, les unités phraséologiques se distinguent en phrasèmes lexicaux et phrasèmes sémantico-lexicaux. Les premiers représentent des phrasèmes contraints par rapport au sens que le locuteur veut exprimer, par exemple *gravement malade* ou *les carottes sont cuites*, des expressions que le locuteur choisit lorsqu'il désire exprimer respectivement les sens de 'très malade' et de 'il n'y a plus rien à faire'. Par contre les phrasèmes sémantico-lexicaux sont des phrasèmes contraints non seulement par rapport au sens que le locuteur veut exprimer, mais aussi par rapport au contexte communicatif dans lequel il se trouve, par exemple à qui de droit, une expression à employer en fonction du sens 'à la personne que cela peut concerner' et du contexte communicatif spécifique, c'est-à-dire dans des lettres ou des courriels. En ce qui concerne les contraintes de combinaison, les unités phraséologiques se distinguent en phrasèmes compositionnels (ex. *gravement malade*) et phrasèmes non compositionnels (ex. *les carottes sont cuites*) en considérant une expression multilexémique comme compositionnelle si elle inclut tous les sens de ses composantes qui par conséquent gardent leur autonomie. Le croisement de ces deux dimensions permet d'identifier trois types différents d'unités phraséologiques, comme représenté dans le tableau 1 :

Unité phraséologique	Non compositionnelle	Compositionnelle
Lexicale	<i>Locution</i>	<i>Collocation</i>
Sémantico-lexicale	Cas impossible (il n'y a pas dans la langue d'exemples de phrasèmes sémantico-lexicaux non compositionnels)	<i>Cliché</i>

Tableau 1 : La LEC et les types d'unités phraséologiques

Aux fins de cette étude, nous n'approfondirons que les locutions et les collocations, deux entités phraséologiques qui diffèrent sur la base de leurs propriétés sémantico-syntaxiques. Les locutions ne sont pas libres sur l'axe syntagmatique, alors que les collocations se caractérisent par un certain degré de liberté ; par exemple, l'insertion d'éléments linguistiques à l'intérieur d'une unité phraséologique est impossible dans le cas de la locution (ex. **manger une belle grenouille*), mais tout à fait possible dans le cas de la collocation (ex. *poser une bonne question*). Cela dépend du fait que dans le cas de la locution les composantes perdent leur autonomie et représentent, ainsi, une seule lexie (ex. *manger la grenouille*), alors que dans le cas de la collocation les éléments - la base et le *collocatif* - sont distincts (ex. dans la collocation *poser une question*, la base *question* et le collocatif *poser* gardent leur autonomie). De plus, la LEC distingue trois types de locutions et deux types de collocations. Les locutions sont classées selon leur degré d'opacité et de transparence: a) les locutions opaques, c'est-à-dire les locutions dont le sens n'est pas compositionnel, sont appelées *locutions fortes* (ex. *manger la grenouille* qui signifie 's'approprier frauduleusement les fonds déposés' n'inclut pas les sens de 'manger' et de 'grenouille'), b) les *locutions faibles* sont des locutions transparentes qui incluent les sens de toutes les composantes aussi bien qu'un autre sens à connaître pour comprendre le sens global de la locution de référence (ex. *donner le sein* qui signifie 'action d'une femme qui nourrit son bébé en lui donnant le lait contenu dans son sein', les sens de 'donner' et 'sein' sont donc tout à fait transparents, mais il faut aussi considérer le sens de 'nourrir un bébé' pour comprendre le sens global de la locution), c) les locutions dont un élément est opaque et l'autre transparent sont dites *semi-locutions* (ex. *loup de mer* qui signifie 'marin très expérimenté', le sens de *loup* est opaque, alors que celui de *mer* est transparent). Quant aux collocations, la LEC les formalise en s'appuyant sur la notion de *fonction lexicale* (dorénavant FL). Sans approfondir ici cette notion, il suffit de rappeler que la FL est une fonction au sens mathématique du terme, $f(x)=y$, où x et y sont deux lexies reliées par un lien syntagmatique ou paradigmatisé par la fonction f . Dans le cas des collocations, la fonction désigne le lien sémantico-syntaxique qui relie y , le collocatif, à x , sa base. La LEC distingue deux types de collocations : les collocations standard et les collocations non standard. Les premières sont décrites par des FL standard, voire des liens sémantico-syntaxiques récurrents dans la plupart des langues du monde - à titre d'exemple nous pouvons mentionner des sémantismes comme le sens de bon, de beaucoup, de vrai, la présence de verbes de réalisation etc. (ex. *rire à se rouler par terre* qui signifie 'rire beaucoup', collocation formalisée par la fonction Magn indiquant le sens d'intensité, $Magn(\text{rire})= \text{à se rouler par terre}$). Les collocations non standard sont décrites par des FL non standard,

c'est-à-dire des fonctions non formalisées ; ces fonctions représentent des liens qui s'appliquent à un nombre très limité de bases et de collocatifs, bien souvent à une seule base et à un seul collocatif (ex. *année bissextile*, collocation formalisée par une fonction non standard indiquant un sens très spécifique, c'est-à-dire « (année) ayant 366 jours »).

Notre équipe de recherche (Frassi *et al.*, 2020) vient de démontrer que dans les langues de spécialité, les termes complexes correspondent aux unités phraséologiques que nous venons d'introduire. De plus, nous précisons que nous avons décidé de considérer les collocations non standard comme des locutions faibles : en effet, ces collocations partagent plus de propriétés avec les locutions faibles qu'avec les collocations standard, par exemple l'impossibilité d'associer plusieurs collocatifs à une base ou l'impossibilité d'ajouter des matériaux linguistiques dans ce type de collocations (Mel'čuk, Clas, Polguère, 1995 : 150-151, Frassi *et al.*, 2020 : 331-332).

L'enseignement des langues, et notamment du lexique, est aussi au cœur des recherches menées au sein de la LEC qui soutient que l'enseignement/apprentissage du lexique passe nécessairement par l'acquisition des notions métalexicales. Ces réflexions se focalisent principalement sur la langue générale, mais des études plus récentes démontrent également leur productivité en langue de spécialité. Plus particulièrement nous nous appuyons sur l'étude de Bonadonna (2020) qui, en analysant les notions métalexicales étudiées par la LEC, a proposé une séquence didactique permettant de travailler sur les mêmes notions en langue de spécialité, notamment sur la terminologie du commerce international. Le présent article vise à approfondir l'étude de Bonadonna : en effet, parmi les notions métaterminologiques de base mentionnées, notre proposition didactique porte presque exclusivement sur l'enseignement/apprentissage des propriétés morphologiques et syntagmatiques du terme (Bonadonna, 2020 : 25-26). Nous croyons fermement que l'emploi d'un corpus spécialisé à des fines didactiques donne la possibilité de réaliser du matériel riche en données, permettant de montrer les termes non plus de manière isolée mais dans des contextes communicatifs réels. C'est justement à partir de cette conviction que nous avons exploité un corpus créé ad hoc pour la préparation du matériel didactique.

2. Le projet DIACOM

Dans ce paragraphe nous allons présenter le corpus que nous avons employé : il s'agit du corpus DIACOM-fr. Ce corpus sera exploité pour l'extraction de termes simples et complexes à insérer dans une base de données terminologiques de type *réseau lexical* (Polguère, 2014). Cette base de données terminologiques sera

adressée à plusieurs destinataires, entre autres des enseignants et des apprenants de français commercial désirant en étudier la terminologie de base. Le corpus DIACOM-fr se compose de 583 textes (dix millions de *word tokens*) sélectionnés selon trois critères :

1. Un critère chronologique, critère d'une importance significative vu qu'il nous permet de définir notre corpus comme un corpus diachronique visant à étudier l'évolution des termes : nous avons considéré des textes datant de trois périodes-clés du commerce international 1) 1850-1914 (la deuxième révolution industrielle), 2) 1945-1970 (le boom économique), 3) 1985-2020 (le développement du commerce électronique et du marketing).
2. Un critère thématique : nous avons sélectionné des textes appartenant à des sous-domaines différents, la macro-économie (politique commerciale, pays d'étude et questions sociales), le secteur (produits et services), les types d'entreprise/activités d'entreprise (management, marketing, logistique, commerce électronique et droit).
3. Un critère textuel : nous avons recueilli des textes de typologies textuelles différentes, c'est-à-dire des textes institutionnels, académiques, des documents d'entreprise et des articles de la presse spécialisée.

À partir de ce corpus, grâce aux programmes *Stanza* (Qi, 2020) et *TermoStat* (Drouin, 2003), nous avons effectué une extraction terminologique des termes simples et complexes en gardant la subdivision par tranches temporelles : nous n'avons retenu que les termes de base du domaine du commerce international, en nous appuyant sur le critère de la fréquence. Nous planifions de réaliser des fiches terminologiques pour les termes complexes aussi bien que pour les termes simples. Nous signalons que nous traiterons les locutions et les collocations de manière différente, en suivant le traitement que la LEC leur a accordé dans le *Dictionnaire Explicatif et Combinatoire* : les unités phraséologiques de type *locution* constituant un tout pré-confectionné, elles auront leur propre fiche terminologique, alors que les unités phraséologiques de type *collocation* seront présentées dans la fiche de référence de la base de la collocation, plus particulièrement lorsque nous donnerons les propriétés de combinatoire du terme. Chaque fiche présentera plusieurs renseignements, entre autres le niveau d'enseignement : nous allons distinguer une échelle A, B, C qui va du niveau A (niveau élémentaire) au niveau C (niveau avancé). Plus particulièrement, en ce qui concerne l'enseignement/apprentissage des unités phraséologiques nous croyons que l'attribution du niveau d'enseignement dépendra de plusieurs facteurs : entre autres, le score de spécificité, c'est-à-dire la valeur indiquant la spécificité et par conséquent l'importance du terme dans le

domaine de référence, et le niveau d'opacité/transparence de l'unité phraséologique. L'exploitation du corpus DIACOM-fr et l'analyse quantitative et qualitative des résultats de l'extraction automatique nous a beaucoup aidés dans la création du matériel didactique que nous allons illustrer dans le prochain paragraphe.

3. Proposition didactique

Notre proposition didactique se compose de quatre parties qui prévoient la participation active des étudiants. Plus dans le détail, nous avons planifié cette séquence dans le but de la proposer aux étudiants du cours de *Langue Française 3 - Terminologie et Traduction* de l'Université de Vérone. En effet, cette proposition didactique s'aligne aux objectifs principaux du cours, c'est-à-dire la connaissance de notions clés de la terminologie et l'acquisition de la méthodologie de travail à suivre lors de la réalisation des produits terminographiques. Notre séquence didactique s'appuie sur l'approche par tâches à laquelle plusieurs études ont eu recours dans l'enseignement des langues étrangères et dans la formation des traducteurs (Li, 2013 ; Alipour, 2018). Dans cette approche, l'apprenant, accomplissant plusieurs tâches, joue un rôle actif, alors que l'enseignant se pose comme guide. Chaque activité se compose de trois moments principaux : 1) la pré-tâche, le moment préparatoire où l'enseignant donne les informations théoriques nécessaires à exécuter l'activité établie ; 2) la tâche, le moment d'exécution de l'activité prévue ; 3) la post-tâche, le moment conclusif où les apprenants et l'enseignant partagent leurs réflexions. L'exploitation du corpus DIACOM-fr permet de surmonter l'une des limites principales de cette approche, la pénurie de ressources disponibles (Kübler, 2014) pour la création d'activités pratiques à proposer. Notre séquence didactique - au total 7 heures - se compose de quatre étapes qui permettront aux étudiants d'acquérir des compétences métaterminologiques, méthodologiques et terminologiques.

La première étape de 60 minutes se pose deux objectifs principaux : l'approfondissement des compétences encyclopédiques préalables sur la terminologie du commerce international et l'acquisition des compétences métaterminologiques. Tout d'abord, les étudiants vont apprendre la distinction entre terme appartenant ou moins au domaine de référence, en approfondissant leurs compétences encyclopédiques, et, ensuite, en ce qui concerne leurs compétences métaterminologiques, ils vont comprendre la distinction entre terme simple et terme complexe. Après le moment de la pré-tâche où l'enseignant définit au niveau théorique le terme et ses propriétés morphologiques, l'activité prévoit la participation active des étudiants. L'enseignant montre quelques résultats d'une extraction automatique de candidats termes simples et complexes ; il demande aux étudiants d'en retenir les seuls termes appartenant au domaine du commerce international.

Ensuite, pour ces termes, les étudiants doivent distinguer, d'une manière plutôt intuitive, les termes simples des termes complexes en s'appuyant sur des critères graphiques. Ce n'est que dans le moment de la post-tâche que l'enseignant donne le classement correct des données montrées. La deuxième étape suit le modèle de la première : 120 minutes dont l'objectif principal porte sur l'acquisition d'autres compétences métaterminologiques, plus particulièrement les étudiants apprendront la différence entre unité phraséologique et syntagme libre et, en ce qui concerne les unités phraséologiques, entre locution et collocation. Cette activité se compose de la pré-tâche, le moment d'explication théorique, de l'activité de classement des termes complexes résultant de la première étape et du moment conclusif de réflexion et de partage. En ce qui concerne l'activité de classement, l'enseignant demande de distinguer tout d'abord les unités phraséologiques des syntagmes libres et ensuite les locutions des collocations en s'appuyant sur leur nature sémantico-syntaxique : en effet, les étudiants doivent observer si l'unité phraséologique constitue un tout pré-confectionné (*locution*) ou si les éléments de l'unité gardent leur autonomie (*collocation*). Pour chaque unité phraséologique les étudiants doivent donner une définition, en outre, dans le cas de la locution, ils doivent indiquer le degré d'opacité (élevé, moyen, faible) et, dans le cas de la collocation, ils doivent indiquer la base et le sémantisme de l'unité phraséologique (par exemple, le sens de *bon*, de *beaucoup*, le verbe de réalisation etc.). Dans ces deux activités, l'enseignant s'appuie sur le corpus DIACOM-fr pour repérer des données réelles à inclure dans les matériaux didactiques. L'enseignant peut choisir les termes, objets des fiches terminologiques, en s'appuyant sur leur niveau d'enseignement : à partir des termes les plus élémentaires il peut rendre plus complexe l'activité en ajoutant des termes de niveau avancé. Le tableau 2 montre des exemples de données résultant d'une extraction automatique faite à travers *Stanza* et *TermoStat* du sous-corpus DIACOM-fr 1985-2020. Nous avons recueilli des exemples représentatifs, et nous les avons classés, en simulant le travail préalable de préparation du matériel didactique de la part de l'enseignant. Les termes simples, les locutions et les collocations sont ordonnés selon leur niveau d'enseignement : des termes les plus élémentaires aux termes du niveau le plus avancé.

Candidat hors domaine	Terme simple	Syntagme Libre	Locution	Collocation
1) <i>Effet</i> 2) <i>Base de donnée</i> 3) <i>Variable</i>	1) <i>Commerce</i> 2) <i>Consommateur</i> 3) <i>Marque</i>	1) <i>Milliard de dollars</i> 2) <i>Nombre d'entreprise</i> 3) <i>Processus de mise</i>	1) <i>Commerce électronique</i> 2) <i>Bien de consommation</i> 3) <i>Intention d'achat</i>	1) <i>Hausse des prix</i> 2) <i>Conclure un accord</i> 3) <i>Favoriser la coopération capital</i>

Candidat hors domaine	Terme simple	Syntagme Libre	Locution	Collocation
4) <i>Donnée</i> 5) Émission de carbone 6) <i>Modèle</i> 7) <i>Moteur de recherche</i> 8) <i>Dioxyde de carbone</i> 9) <i>Industrie pharmaceutique</i> 10) <i>Administration locale</i>	4) <i>Investissement</i> 5) <i>Exportation</i> 6) <i>Importation</i> 7) <i>Douane</i> 8) <i>Rendement</i> 9) <i>Hospitalité</i> 10) <i>Arbitrage</i>	4) <i>Type de commerce</i> 5) <i>Type de spécialisation</i> 6) <i>Nombre de clients</i> 7) <i>Forme de publicité</i> 8) <i>Terme de production</i> 9) <i>Notion de confiance</i> 10) <i>Contexte commercial</i>	4) <i>Transport intérieur</i> 5) <i>Droit antidumping</i> 6) <i>Commerce de gros</i> 7) <i>Commerce de détail</i> 8) <i>Compte de résultat</i> 9) <i>Équation de gravité</i> 10) <i>Placement de portefeuille</i>	4) <i>Part de marché</i> 5) <i>Résultat positif</i> 6) <i>Résultat solide</i> 7) <i>Stimuler la demande</i> 8) <i>Augmenter la productivité</i> 9) <i>La productivité augmentée</i> 10) <i>Attirer le capital</i>

Tableau 2 : Exemples de données résultant d'une extraction automatique du corpus DIACOM-fr

La troisième étape de notre proposition didactique (120 minutes) vise à développer des compétences méthodologiques : les étudiants auront la possibilité d'explorer le travail des terminologues, notamment le travail de constitution et d'exploitation d'un corpus spécialisé. Dans le moment de la pré-tâche l'enseignant présente la notion de *corpus spécialisé* : concrètement, il peut avoir recours au DIACOM-fr, disponible sur la plateforme de l'Université de Vérone, pour en expliquer les critères de constitution. De plus, il peut illustrer la démarche d'extraction terminologique automatique à travers la présentation des fonctions principales du logiciel *TermoStat*. Après cette explication théorique, l'activité prévoit un travail de groupe : à partir du corpus DIACOM-fr les groupes, composés de trois/quatre étudiants, constituent un sous-corpus de dix textes et ils effectuent l'extraction automatique des termes complexes à travers *TermoStat*. Ensuite, après avoir éliminé le bruit dans les résultats, ils s'occupent de leur classement en distinguant termes simples, syntagmes libres, locutions et collocations. Ce n'est qu'au moment de la post-tâche que l'enseignant peut vérifier l'acquisition des compétences méthodologiques : en effet, chaque groupe ne doit pas seulement illustrer les résultats obtenus, mais il doit aussi expliquer les critères de constitution du sous-corpus et la démarche suivie pour arriver au classement conclusif. Pendant la dernière étape (120 minutes) les étudiants approfondissent leurs compétences terminologiques. En effet, au moment de la pré-tâche, l'enseignant présente plusieurs notions théoriques : il montre un exemple de fiche terminologique dont il explique les informations principales, notamment l'équivalent en langue italienne,

la partie du discours, la définition, les contextes d'emploi, les principales relations paradigmatiques et syntagmatiques. Dans le tableau 3, nous illustrons des exemples de fiches terminologiques.

Champ de la fiche	Exemple de terme simple (base de plusieurs collocations)	Exemple de terme complexe de type locution
Terme vedette	Client	Commerce électronique
Équivalent en italien	Cliente	Commercio elettronico
Partie du Discours	Nom Masculin	Locution nominale masculine (transparente)
Définition	Personne et/ou entreprise qui achète des biens et/ou des services	Commerce de biens et/ou services en utilisant internet ou d'autres réseaux informatiques
Contexte d'emploi (du Corpus DIACOM-fr)	« L'obtention d'un client est la plupart du temps le fruit d'un appel d'offre remporté face à des agences concurrentes, au terme d'une compétition consistant à présenter l'offre la plus adaptée au besoin exprimé par le client potentiel »	« Le commerce électronique n'est plus un simple support promotionnel de la distribution des biens de consommation, il s'apparente de plus en plus à un marché parallèle à la distribution classique »
Relations paradigmatiques	Opérateur économique (nom générique), consommateur, acheteur (synonymes), vendeur (nom conversif), marché (nom de lieu), clientèle (nom du groupe)	Commerce (nom générique), cybercommerce (synonyme), commerçant, vendeur (noms d'agent), produit, marchandise, service, bien (noms de patient)
Relations syntagmatiques_ Collocations Verbe+ Nom	Fidéliser les clients, attirer les clients (verbes de préparation)	Promouvoir/ faciliter/ favoriser le commerce électronique, (verbes de préparation)
Relations syntagmatiques_ Collocations Nom+ Adjectif	Un client fidèle (sens de bon), Un mauvais client (sens de mauvais)	X

Tableau 3 : Exemples de fiches terminologiques pour les termes *client* et *commerce électronique*

L'activité proposée engage les groupes dans la préparation de vingt fiches terminologiques : dix fiches pour des termes complexes de type *locution* (ex. *commerce électronique*) et dix fiches pour les bases des collocations obtenues pendant la

troisième étape de cette séquence didactique (ex. *client* base de plusieurs collocations comme *fidéliser les clients*, *attirer les clients*, *un client fidèle*, *un mauvais client*). Les fiches terminologiques sont présentées et vérifiées au moment de la post-tâche. Enfin chaque groupe partage ses fiches terminologiques avec les autres : cela permet à tous les étudiants de collecter un nombre significatif de fiches qui leur permettront d'enrichir leurs compétences terminologiques en ce qui concerne la terminologie du commerce international. Bref, dans ces deux dernières étapes, le corpus DIACOM-fr est employé aussi par les étudiants qui ont un accès direct aux textes préalablement traités et sélectionnés par notre équipe de recherche, permettant aux étudiants de travailler sur des textes déjà prêts au traitement automatique.

Conclusion

Cette proposition didactique naît de plusieurs réflexions théoriques, notamment de la conviction que l'enseignement/apprentissage du lexique passe nécessairement par l'acquisition de notions métalexicales. À partir de cela nous avons essayé d'adapter aux langues de spécialité une méthodologie qui, d'habitude, se prête à l'enseignement/apprentissage du lexique de la langue générale. Dans l'objectif de réduire la pénurie des ressources disponibles pour l'étude des langues de spécialité, nous avons créé un corpus *ad hoc* concernant la terminologie du commerce international, le corpus DIACOM-fr. Plus particulièrement, nous espérons avoir démontré l'utilité d'un corpus spécialisé à des fins didactiques et la richesse des données qui peuvent être repérées et exploitées. Certes pour l'avenir, nous visons à tester cette proposition didactique : cela sera possible lors de la mise en ligne du corpus sur une plateforme accessible au public. Ce ne sera que la mise en pratique de notre proposition didactique qui nous permettra d'observer si cette séquence atteint les objectifs conçus à un niveau essentiellement théorique. Enfin, il est évident que la méthodologie que nous avons adoptée pour l'élaboration de notre proposition didactique peut s'appliquer à d'autres domaines de la langue de spécialité, en enrichissant le matériel disponible pour l'enseignement de la discipline terminologique.

Bibliographie

- Alipour, M. 2018. « Création d'activités lexicales basées sur corpus pour la formation des traducteurs ». *ILCEA*, n° 32, p. 1-18.
- Bonadonna, M.F., Zollo, S. 2019. « A didactic sequence for phrasemes in L2 French ». *Computational and corpus-based Phraseology, Actes Europhras 2019* (Malaga, 25-27 septembre 2019), p. 53-60.

- Bonadonna, M.F. 2020. « L'enseignement des notions métaterminologiques en français L2 : un parcours fondé sur les corpus spécialisés ». *Cahiers de lexicologie*, n° 117, p. 11-33.
- Cavalla, C., Labre, V. 2009. « L'enseignement en FLE de la phraséologie du lexique des affects ». In : Tutin, A., Nokanova, I. (éds.), *Le lexique des émotions et sa combinatoire lexicale et syntaxique*. Grenoble : Ellung, p. 297-316.
- Drouin, P. 2003. « Term extraction using non-technical corpora as a point of leverage ». *Terminology*, n° 9/1, p. 99-115.
- Frassi, P. 2018. « L'enseignement/apprentissage de la collocation entre contraintes grammaticales et contenu sémantique ». *Études de Linguistique Appliquée*, n° 189, p. 63-84.
- Frassi, P. et al. 2020. « Fouille de textes et repérage d'unités phraséologiques ». *Terminologie & Ontologie : théories et applications*. Chambéry : Presses Universitaires Savoie Mont Blanc, p. 321-338.
- Kübler, N. 2014. « Mettre en œuvre la linguistique de corpus à l'université ». *Recherches en didactique des langues et des cultures*, n° 11-1.
- Li, D. 2013. « Teaching Business Translation: a Task-Based approach ». *The Interpreter and Translator Trainer*, n° 7/1, p. 1-26.
- Mel'čuk, I., Clas, A., Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Paris/ Louvain-La-Neuve : Duculot.
- Mel'čuk, I. 2013. « Tout ce que nous voulions savoir sur les phrasèmes mais... ». *Cahiers de lexicologie*, n° 102/1, p. 129-149.
- Polguère, A. 2014. « From writing dictionaries to weaving lexical networks ». *International Journal of Lexicography*, n° 27/4, p. 396-418.
- Qi, P. et al. 2020. « Stanza: A Python natural language processing toolkit for many human languages ». *Association for Computational Linguistics (ACL) System Demonstrations*.

Synergies Italie n° 17 / 2021



Comptes rendus
de lecture



Annalisa Buonocore

Université de la Vallée d'Aoste, Italie

a.buonocore@univda.it



Sabina Gola, Michel Pierrard, Evie Tops et Dan Van Raemdonck (dir.) (2019). *Enseigner et apprendre les langues au XXI^e siècle*, Bruxelles, Peter Lang, 155 p.

Les professionnels qui travaillent actuellement dans le domaine de l'éducation sont de plus en plus confrontés à de nouveaux enjeux et défis. Les nouvelles technologies, l'enseignement bi-plurilingue, l'interculturel, de nouvelles méthodes d'apprentissages représentent autant d'opportunités stimulantes pour les enseignants et les apprenants.

Toutefois, si d'un côté il est devenu impossible de négliger ces nouveautés que l'on va de plus en plus intégrer à l'enseignement traditionnel, de l'autre côté, un travail de systématisation, d'approfondissement et d'analyse, qui puisse offrir des pistes de réflexion et garantir un emploi plus conscient et réfléchi de ces mêmes méthodes et outils innovants, est de plus en plus urgent.

Le volume *Enseigner et apprendre les langues au XXI^e siècle* est un recueil d'études pionnières où de nouvelles méthodes sont objet d'analyse scientifique. Ce volume a le mérite d'apporter des éclaircissements sur des approches didactiques très hétérogènes et sur certains aspects controversés qui suscitent un certain scepticisme sur leur validité ou efficacité. Ceux-ci constituent une aide aux concepteurs et aux enseignants qui veulent innover et diversifier leur offre d'éducation.

Dans le premier chapitre, Laurent Gajo (Université de Genève) illustre les changements apparus ces dernières années dans la didactique des langues étrangères et les conditions d'émergence d'une didactique du plurilinguisme en Europe. Cette évolution a engendré de nombreuses approches et des pratiques différentes en classe de langue. L'auteur propose une schématisation de ces approches en remarquant que dans le processus de didactisation on se positionne sur deux pôles : « langue en mention » vs « langue en usage » et « monolinguisme » vs « plurilinguisme ». Il encourage le caractère dynamique de ces axes, sur lesquels on peut se déplacer pour créer de nouvelles stratégies didactiques. Enfin, à travers des exemples tirés des dernières recherches dans la didactique du plurilinguisme, il montre les potentialités d'un travail bi-plurilingue ancré dans l'usage.

Le deuxième chapitre concerne une importante innovation pédagogique, les Moocs (*Massive Open Online Course*). Nicolas Roland de l'Université libre de Bruxelles (ULB) s'en occupe en étudiant le Mooc « *Spice-up your English* », un cours de français langue étrangère qui a été conçu pour les étudiants et le personnel de l'ULB et qui se base essentiellement sur des vidéos. Plus particulièrement, l'étude se concentre sur les apprenants et leur appropriation de ressources audiovisuelles. Cette contribution a le mérite de souligner un aspect essentiel lié à cette méthode innovatrice, les concepteurs de Moocs ne créent pas de vidéos pour donner simplement des informations mais ils sont capables désormais de créer des liens humains, « un fil rouge humain ». L'étude montre aussi qu'une même vidéo pédagogique peut donner lieu à un processus d'appropriation des ressources très différent selon des caractéristiques propres à chaque apprenant. Au moment de la création de ces vidéos les concepteurs doivent en tenir compte.

Le troisième chapitre est consacré à une étude sur l'efficacité d'une plateforme d'auto-apprentissage, « Le Projet Voltaire ». Cette plateforme du Centre Académique des Langues de l'Université néerlandophone de Bruxelles est utilisée dans un cours de français sur objectifs spécifiques (FOS) et a comme objectif d'améliorer l'orthographe des apprenants francophones mais aussi des apprenants pour lesquels le français est langue seconde (FL2). La recherche vise à vérifier si « Le Projet Voltaire » améliore l'expression écrite et les structures grammaticales des apprenants et quel profil d'apprenant fait-il progresser le plus. On s'est basé sur une comparaison entre deux groupes d'apprenants. L'un a utilisé la plateforme, l'autre a bénéficié de deux heures par semaine d'interaction orale avec des francophones natifs. Même si la recherche est faite pendant une période très courte et les données ne sont pas toujours statistiquement remarquables, Julie Luyckx et Hélène Stengers ont pu constater un impact positif de la plateforme sur la progression des apprenants.

Le quatrième chapitre s'occupe d'un aspect souvent peu considéré dans les écoles, celui d'aider les élèves à travailler en autonomie. Lenka Zouhar Ludvíková nous explique que pour favoriser le développement de cette compétence, à l'Université Masarik (Brno), les professeurs du Centre linguistique de l'Université ont organisé *L'Anglais Autonomously Course*. Il s'agit d'un cours qui permet aux étudiants d'apprendre l'anglais et qui, en même temps, favorise l'autonomie dans le processus d'apprentissage en le rendant conscient et riche de stimuli. La possibilité donnée aux apprenants de faire des choix et de prendre des décisions sur leurs cours à la place des institutions est un grand atout. Ils peuvent décider des objectifs, des contenus, des méthodes, des temps et de l'évaluation. Cette possibilité est due à des sessions que les étudiants ont avec les enseignants

qui ont le rôle de guide. Ils essaient d'aider les apprenants à trouver leur chemin. De plus ces sessions, étant en anglais et visant à améliorer le processus d'apprentissage, représentent un exemple bien réussi d'apprentissage proactif de la langue.

Dans le cinquième chapitre, Tim Lewis (The Open University, UK) nous offre des réflexions sur l'apprentissage des langues en « *e-tandem* » (apprentissage fait à travers des échanges en ligne, synchrones ou asynchrones, et qui utilisent des outils tels que des courriels, les chats, les web conférences). L'augmentation d'instruments tels que les chats ou *Skype* ont permis de favoriser l'utilisation du discours comme moyen de communication. Le fait d'avoir des conversations, libres et authentiques, aide les apprenants à améliorer leurs compétences dans le discours beaucoup plus que durant des échanges dans les classes de langue où les modèles sur lesquels se basent les interactions sont beaucoup plus limités que lors des conversations libres. L'auteur souligne l'aspect interculturel des échanges en *e-tandem*, qui emploient toujours deux langues. En plus, la nécessité de la part de l'apprenant de comprendre dans les échanges les intentions des interlocuteurs et de faire comprendre ses intentions communicatives permet aux apprenants de s'engager dans les aspects de langue qui concernent la formation du signifié, le lexique mais aussi la pragmatique.

Enfin, dans le sixième chapitre, Caroline Staquet a pour objet de son analyse scientifique les présupposés scientifiques qui sous-tendent l'enseignement d'une matière par l'intégration d'une langue étrangère. Pour son étude elle a sélectionné 5 mots-clés : « Intégration », « *Cosmopolitan identity* », « convergence », « best practice », « société de la connaissance » et en a fait une analyse interdiscursive. La mise en parallèle des mots analysés avec les contextes où ces mots apparaissent, a montré le caractère idéologique et promotionnel des nombreuses publications scientifiques de référence sur l'EMILE.

Manon Hayette
Université de Mons, Belgique



Christophe Rey (2020). *Dictionnaire et Société*, Paris, Honoré Champion, 256 p.

Dans cet ouvrage, Christophe Rey, professeur en sciences du langage à l'Université de Cergy-Pontoise et membre sénior de l'Institut Universitaire de France, nous propose un point de vue novateur sur le dictionnaire en mettant en exergue l'intérêt que peut représenter ce type d'outils pour les sociolinguistes en tant que corpus d'investigation. Ce plaidoyer se justifie, selon l'auteur, par le fait que les dictionnaires sont trop facilement assimilés à l'unique fonction de gardiens de la norme, et pas assez exploités en sociolinguistique. L'auteur considère exclusivement la tradition lexicographique française et indique que son exposé n'est pas nécessairement transposable à d'autres traditions lexicographiques.

À la suite d'une courte introduction précisant le sujet de l'ouvrage ainsi que sa structure et le domaine de recherche de l'auteur, trois chapitres s'offrent au lecteur, auxquels s'ajoute une conclusion générale.

Le premier chapitre (p. 15-96) porte sur la lexicographie dans ses dimensions culturelle et sociale et présente une typologie des différents dictionnaires (dictionnaires extensifs/restrictifs, dictionnaires de mots/de choses, dictionnaires de langue/de la langue et dictionnaires d'usage/des usages), dont la diversité s'avère fréquemment négligée. L'auteur considère le dictionnaire comme un outil de grammatisation, capable de faire exister la langue tout en décrivant (de manière inévitablement incomplète) une culture et une communauté à une époque donnée. Dans cette perspective, il est le révélateur d'une certaine évolution de la langue et contient des traces d'idéologie, de militantisme et de différentes formes de censure.

Il est cependant important de souligner que, bien que Christophe Rey voie le dictionnaire comme un corpus de recherche passionnant pour les sociolinguistes, il rappelle que cet outil ne peut pas constituer leur seule source d'investigation. En effet, le dictionnaire reste « une photographie subjective de la langue à un instant T » (p. 48) et son élaboration est donc aussi le résultat des choix lexicographiques

effectués par son auteur (dans lesquels entre nécessairement une part de subjectivité), dont la vision nécessairement partielle peut refléter des contraintes commerciales et des considérations sociétales. Enfin, toujours dans le premier chapitre, l'auteur fournit un aperçu captivant de ses propres recherches portant sur le traitement des régionalismes dans le *Petit Robert* 2009 et dans la neuvième édition du *Dictionnaire de l'Académie française* de 1992, permettant ainsi au lecteur d'observer dans quelle mesure des marquages dépréciatifs tels que « vieux » ou « familier » peuvent nuire au maintien d'usages régionaux dans un contexte où la langue française s'avère toujours plus conquérante.

Une étude sociohistorique et sociolinguistique des dictionnaires est présentée dans le deuxième chapitre (p. 97-183). Après avoir brièvement rappelé l'existence des sommes, gloses et autres glossaires du Moyen Âge, l'auteur s'attelle à analyser le degré d'ouverture des dictionnaires vers l'étranger. En effet, cette ouverture est à la fois une composante des dictionnaires bilingues et multilingues des ^{xiv}^e et ^{xv}^e siècles, des monolingues anciens (le « triptyque fondateur » du ^{xvii}^e siècle, cf. *infra*) qui voulaient faire rayonner la France à l'étranger (grâce à des commentaires sur l'usage, à une attention particulière à la prononciation [sauf dans le cas du *Dictionnaire de l'Académie*] ou à des explications sur le caractère polysémique de certains mots), mais aussi des monolingues d'aujourd'hui qui incluent des indications sur le genre des mots, non essentielles pour les locuteurs natifs. L'auteur s'attarde, dans la section suivante, sur ce qu'il appelle la « Trinité de la lexicographie », à savoir les monolingues de Richelet (1680), de Furetière (1690) et de l'Académie (1694), qui font l'objet d'un récit détaillé tenant compte de tout le contexte politique et historique. Cet exposé est couplé à une présentation des encyclopédies, elle-même suivie d'un historique allant des accumulateurs de mots de la première moitié du ^{xix}^e siècle aux grands dictionnaires de la première moitié du ^{xx}^e siècle (*Le Grand Robert, Trésor de la langue française*), en passant par le *Littré* ([1863-1872] avec l'apparition d'une dimension historique caractérisée par le marquage de l'étymologie) et par les dictionnaires de types encyclopédiques tels que le *Larousse* (^{xix}^e siècle) et le *Quillet* (^{xx}^e siècle). Enfin, l'auteur mentionne à nouveau la lexicographie en langue régionale, selon lui sous-exploitée, et montre l'intérêt d'une lexicographie différentielle qui aborderait d'autres variétés de français que celle de l'Hexagone, envisageant ainsi la langue française dans toute sa diversité. Il s'agit là d'un bel exemple des confluences qui peuvent naître de la combinaison de la lexicographie et de la sociolinguistique.

Cette conclusion constitue une parfaite transition au chapitre trois (p. 185-219) qui revient sur la thèse défendue par l'auteur selon laquelle le dictionnaire, en tant que témoignage historique de la culture, des langues, de la pédagogie et des progrès

techniques, possède un grand potentiel en tant que corpus d'investigation pour les sociolinguistes. Pour illustrer cette hypothèse, des études antérieures mêlant lexicographie et sociolinguistique ainsi que des pistes pour des recherches à venir sont discutées. L'auteur rappelle notamment les nombreuses analyses déjà réalisées sur les marques (diachroniques, diatopiques et diastratiques) qui accompagnent les différentes entrées des dictionnaires. Christophe Rey, lui-même particulièrement intéressé par la question du traitement des régionalismes par les lexicographes, s'est penché sur le cas du « synonyme régional » adopté dans l'édition 2009 du *Nouveau Petit Robert*. Selon l'auteur, les régionalismes, présentés dans cet ouvrage comme de simples synonymes sans mention de leurs origines géographiques, seraient une nouvelle fois stigmatisés par rapport au français standard. En outre, les marques de genre, notamment la féminisation des noms de métiers, ont également fait couler beaucoup d'encre dès la fin des années 2000.

L'intérêt personnel de l'auteur pour les variations diatopiques explique l'importance des exemples et l'intérêt des discussions traitant de cet aspect. Toutefois, le lecteur curieux d'autres types de variations pourrait rester sur sa faim et il aurait dès lors été peut-être judicieux que l'auteur élargît son propos à ces autres variations, en présentant avec plus de détails les études menées dans ces domaines.

Par ailleurs, s'agissant des recherches devant être conduites par les lexicographes, l'auteur insiste à nouveau sur le fait que beaucoup reste à accomplir dans le domaine de l'étude du traitement lexicographique des régionalismes. Si l'incomplétude est le lot de tout dictionnaire, il serait souhaitable d'être mieux informé des critères utilisés par les lexicographes pour réaliser leurs dictionnaires, notamment à l'aide des préfaces. L'auteur pose ensuite la question suivante : si ces outils s'avèrent incapables d'offrir une image complète et parfaite de la langue, de quelles évolutions linguistiques sont-ils les témoins ? L'histoire riche de plusieurs siècles du dictionnaire en fait un candidat de choix pour des études diachroniques en sociolinguistique. Il serait aussi opportun, selon Christophe Rey, d'observer l'influence de la dictionnaire et de l'application lexicographique des outils informatiques, notamment sur les dictionnaires millésimés, dont la rédaction dépend généralement de facteurs extralinguistiques. À cette liste des études à mener, l'auteur ajoute un domaine qui a déjà été traité par les métaléxicographes : les « contacts » entre les langues dans la nomenclature du dictionnaire. Cette question se rapprochant fortement de celle de « contact des langues », il serait, par conséquent, du ressort du sociolinguiste d'analyser l'organisation des langues dans les dictionnaires, du moins pour cette dimension.

Toutes ces pistes de recherches ne font que confirmer l'intérêt de la lexicographie comme terrain d'investigation pour le sociolinguiste. Néanmoins, selon l'auteur, la sociolinguistique n'est pas la seule discipline qui sortirait grandie de cette collaboration : la sociolinguistique pourrait contribuer à donner au dictionnaire une apparence moins rigide et artificielle.

En conclusion, cet ouvrage, extrêmement documenté et d'une construction claire, s'adresse avant tout aux sociolinguistes, ainsi qu'à tous les amoureux de la lexicographie. La présence d'une bibliographie détaillée ainsi que d'un index des notions et des noms propres répondra aux attentes des chercheurs et à leurs besoins spécifiques.

Bien que, pour des raisons compréhensibles et inhérentes, entre autres, au travail scientifique, l'auteur ait été amené à limiter son propos, il serait souhaitable que d'autres études viennent compléter ce tableau, notamment sur la question des belgicisms, helvétismes et autres québécoisismes. Enfin, de nombreuses pistes d'études présentées dans cet ouvrage (dictionnaire, traitement des régionalismes, préfaces et lexicographes), particulièrement ancrées dans la tradition française, seraient également applicables à d'autres langues et cultures. Autrement dit, l'auteur ouvre et balise avec ce premier ouvrage un immense terrain de recherches pour les sociolinguistes désireux de se pencher sur les dictionnaires ou pour les lexicographes souhaitant s'ouvrir à la sociolinguistique.



Maria Margherita Mattioda

Université de Turin, Italie

marita.mattioda@unito.it

Geneviève Tréguer-Felten (2018). *Langue commune, cultures distinctes. Les illusions du globish*, Laval, Les Presses de l'Université Laval, 214 p.

En 2013 le juriste et philosophe François Ost faisait le constat suivant : « l'expérience enseigne que tout langage unique ne tarde jamais longtemps à se « babéliser » (Ost, 2013). Face à une société mondialisée et dominée par un modèle communicationnel présenté de toute part comme inéluctable, la dialectique entre la langue universelle et la pluralité post-babélienne est toujours à l'œuvre pour remettre en question ces axiomes qui tendent à s'imposer selon les contingences. C'est exactement dans les plis de la langue commune prêchée par les grandes entreprises opérant à l'international - ce « globish » évoqué dans le sous-titre -, que nous fait plonger Geneviève Tréguer-Felten dans ce livre solide et stimulant afin de faire jaillir toutes les limites de l'adoption d'une langue commune (l'anglais) pour la communication en situation d'interculturalité. Par le biais de l'analyse de discours comparative appliquée à la communication professionnelle des entreprises, l'auteure s'attache à montrer les « illusions d'une évidence » à travers l'« épaisseur » des discours : si d'une part la langue globale est un fait assis dans la réalité des activités humaines mondialisées, d'autre part le manque de réflexion concernant l'emploi généralisé de l'ELF (anglais *lingua franca*) sous un angle plus proprement linguistique et culturel empêche d'en cerner toutes ses limites. L'originalité de l'approche adoptée dans cet ouvrage consiste à considérer cette *lingua franca* à la lueur de la culture inscrite par les locuteurs dans tout discours et par les univers de sens qu'ils mobilisent, d'où les « accidents de parcours », la diversité des interprétations, la « mal-entente ». À ce propos, les *Annexes* clôturant le livre sont éclairantes sur les fondements théorico-méthodologiques qui encadrent cette recherche visant à rendre perceptible l'inscription des locuteurs dans le discours grâce à une approche interprétative de la culture associée à l'analyse de discours contrastive appliquée à un corpus de textes très spécifiques (brochures, courriels, autoprésentations, codes et principes...) issus d'entreprises françaises, étatsuniennes et chinoises. De par un *modus operandi* outillé (l'analyse lexicométrique) et orienté par les transformations des échanges

numériques, l'auteure s'intéresse à la manière dont ces entreprises construisent et affichent leur *ethos* dans la langue de la communication internationale (ELF) et examine les tensions discursives qui en ressortent. Or, ces choix communicationnels dans une langue soi-disant commune, mais dépourvue de fonds culturel partagé demandent aux destinataires de faire appel à leur propre « univers de sens » d'où la question de la représentation de l'autre en tant qu'étranger, comme le suggère la langue adoptée, ou en tant qu'*alias* culturels comme l'indique leur « culture discursive ».

Dans les huit chapitres composant cet ouvrage, le lecteur - chercheur, professionnel ou étudiant en formation - pourra trouver matière à repenser les représentations des langues et des pratiques communicationnelles institutionnalisées dans les organisations (entre autres) à partir de l'osmose entre langue, culture, discours. Celle-ci révèle l'impossible transparence de la *lingua franca*, affirme la valeur de la traduction en tant que pratique indispensable pour transposer des « univers de sens » et fait émerger l'exigence de développer une certaine « attention », voire une « conscience » plurilingue pour mieux répondre aux défis de la communication interculturelle.

Dans le premier chapitre, sont illustrées les dérives communicationnelles engendrées par la non-coïncidences des univers de sens. En effet, le contexte culturel donne la couleur à tout discours, d'où le titre éloquent *Tu dis vert, j'entends bleu !*, mais l'absence d'un terrain d'entente partagé en ELF alimente les incompréhensions et les malentendus. Comme le montre l'analyse des échanges dans les courriels rassemblés au sein de plusieurs firmes multinationales, produits par des locuteurs de divers statuts, l'efficacité de la communication se heurte non seulement à des problèmes relevant de la maîtrise de la *lingua franca*, mais encore à la distance culturelle entre les usagers, ce qui constitue l'obstacle principal pour passer de la « commun'action » à la « commun'entente ». Un retour sur les fondements mêmes de la communication met en évidence le fait que la culture constitue le fond sur lequel les discours, les situations, les relations sociales prennent sens. Ainsi, la communication par l'intermédiaire de l'anglais international n'offre pas de cadre commun aux locuteurs qui font appel à leur propre culture pour construire le sens, d'où la notion de « mal-entente », proposée par l'auteure, pour désigner des incompréhensions latentes (Chapitre 2. *D'où vient cette « mal-entente » ?*).

Dans le chapitre suivant (*Un cas franco-chinois*), Geneviève Tréguer-Felten s'attache à l'analyse contrastive des autoprésentations de grandes entreprises chinoises et françaises en ELF étalées sur le temps puisqu'un premier corpus de brochures a été recueilli à la fin des années 1990, intégré en 2008 par des documents numériques téléchargés sur Internet à partir de la rubrique *About us*.

Après avoir tracé le contexte macro-économique et présenté les différentes structures textuelles et énonciatives qui organisent le discours *corporate*, caractérisé d'une part par l'effacement énonciatif compensé par la prise en charge visuelle du discours (sous corpus chinois) et d'autre part par l'inscription syntaxique de l'énonciateur e/ou du destinataire (sous corpus français), dans les chapitre 4 (*L'ethos des entreprises chinoises*) et 5 (*L'ethos des entreprises françaises*) la réflexion se poursuit dans le but de mieux cerner l'image de soi que ces firmes projettent dans et par leurs discours. Un ethos d'entreprise spécifique à chaque pays émerge dans le corpus examiné. Les entreprises chinoises se caractérisent par une inscription dans l'histoire et une plus forte intégration dans la société qui en fait leur réputation ; ainsi leur ethos social se projette-t-il dans la trame discursive en déployant ses valeurs et sa « bienveillance » à l'égard de son auditoire qui se veut universel. Le binôme énonciateurs-destinataires vise à instaurer des relations harmonieuses et une véritable communion entre eux. En revanche, les entreprises françaises se focalisent sur le moment présent et laissent transparaître des qualités éthiques quelque peu différentes dans un discours davantage hiérarchisé esquissant deux profils principaux : le mentor et le leader. Le premier « fait preuve de bienveillance en guidant le client destinataire du discours », le second soutient ce rôle en affichant sa position et son expertise, ce qui détermine une relation asymétrique entre les énonciateurs et les destinataires. Le discours déployé s'appuie sur un argumentaire logique et sur l'évidence des données chiffrées visant à produire une image rationnelle et solide censée correspondre à celle attendue par l'auditoire.

La comparaison entre les images « idéales » inscrites dans le corpus par ces entreprises et les contextes spécifiques qui leur permettent de construire le sens occupe presque entièrement le chapitre 6 (*Dissemblances et univers de sens*). La communication institutionnelle que ces entreprises chinoises et françaises s'adressent montre qu'en dépit du choix de l'ELF à l'intention d'un public international, toutes leurs productions discursives portent la marque du « *collective ethos* » sous-jacent, ce qui explique le décalage entre l'auditoire étranger désigné par l'ELF et l'auditoire inscrit en creux dans leurs discours. De plus, l'analyse détaillée de la démarche rhétorico-argumentative adoptée permet de saisir la circulation des schèmes de pensée propres à chaque culture à travers les ressources rhétoriques employées et de révéler le transfert de la culture discursive des énonciateurs vers l'ELF. Cette démarche est illustrée concrètement dans le chapitre 7 (*Un exemple d'application pratique*) où l'exploration d'un corpus complémentaire permet de tester les résultats obtenus en élargissant la variété des langues (ELF, anglo-américain et français), des genres discursifs (codes éthiques) et des secteurs d'activité. La focalisation de l'analyse sur la relation-client, que l'on tend à présenter comme

un concept managérial « universel », corrobore le fait que les dissemblances dans les textes comparables remontent à l'inscription de cette notion dans des univers de sens propres aux cultures concernées. Il en résulte le besoin de revenir sur la place de la traduction dans les entreprises et sur le rôle à lui attribuer en s'interrogeant sur les avantages et les inconvénients de la considérer un simple « habillage d'un message » en anglais international ou un véritable atout stratégique pointant l'équivalence linguistique, discursive et culturelle dans une autre langue (Chapitre 8. *L'ELF une langue bien particulière*).

Dans sa conclusion, Geneviève Tréguer-Felten rebondit sur l'idée à la base de son ouvrage : l'anglais lingua franca ne possède aucune « culture discursive » distincte, mais elle est traversée par les « univers de sens » de ses locuteurs indépendamment de leurs niveaux linguistiques. C'est pourquoi, elle propose à bon escient de jeter un nouveau regard sur la communication internationale des entreprises en passant par le développement d'une conscience de la relativité linguistique - et par conséquent de la traduction - en tant que moyen anthropologique et culturel du *vivre ensemble* et de la communication interculturelle.

Bibliographie

Ost, F. 2013. « Babel revisité. Défense et illustration du multilinguisme », *Synergies Italie*, n°9., p. 25-34. [En ligne]: <https://gerflint.fr/Base/Italie9/ost.pdf> [consulté le 15 mai 2021].



Giovanni Tallarico

Université de Vérone, Italie

giovanni.tallarico@univr.it



Michela Tonti (2020). *Le nom de marque dans le discours au quotidien. Prisme lexiculturel et linguistique*, Paris, L'Harmattan, 206 p.

L'onomastique commerciale a fait l'objet de nombreuses études dans le domaine francophone, de l'ouvrage fondateur de Galisson et André (1998) jusqu'à l'étude d'Altmanova (2013). Plus récemment, nous pouvons citer l'article de Faure (2018) et l'important volume (faisant plus de 700 pages) dirigé par Altmanova et Le Tallec (2019).

Publié dans la collection *laboratorio@francesisti.it* financée par la S.U.S.L.L.F., le travail de Michela Tonti se situe explicitement, dès le titre, dans le sillage des études à visée lexiculturelle de Galisson, mais en les prolongeant par une approche basée sur un corpus, ce qui constitue une innovation remarquable. Presque 2 000 noms de marque (dorénavant NdM) sont passés au crible de l'analyse de Tonti, dans le but d'en mettre en valeur les implicites culturels et leur rôle pragmatique et interactionnel.

Le premier chapitre pose les jalons théoriques de l'étude et tente de définir le concept de NdM, au carrefour de disciplines tels le marketing, la sémiotique, le droit et, bien entendu, la linguistique. À cet égard, l'objectif de cette étude est d'identifier les caractéristiques des noms propres en général et des NdM, d'un point de vue morphosyntaxique, sémantique et pragmatique. Après un utile passage en revue des études sur l'onomastique commerciale, qui ont privilégié jusqu'à présent les analyses formelles et descriptives, Tonti plaide pour une linguistique *in vivo*, qui ne se borne pas à enquêter sur les aspects étymologiques du NdM mais qui vise à étudier le fonctionnement discursif global des NdM.

Le deuxième chapitre présente la méthodologie adoptée (*corpus-based*) et la constitution de la base de données des NdM, qui englobe des sources telles les catalogues en ligne et les comparateurs de prix, d'une part, et la base de données de l'I.N.P.I. de l'autre, cette dernière ayant fourni la grille d'analyse pour catégoriser les NdM. Le corpus, en langue française, exploite la ressource *Araneum*

Francogallicum Maius (1,2 milliard de mots environ), qui regroupe des sources journalistiques, des sites institutionnels et d'entreprise, des forums de discussion et des blogs personnels.

La véritable exploration du corpus commence avec le chapitre trois : il est question tout d'abord du comportement discursif des NdM, entre polysémie et homonymie. En effet, il s'avère que presque un tiers (636 sur 1 987) des NdM de l'échantillon analysé correspondent à des lexies de la langue commune, en priorité des noms propres ou communs. La notion galissonienne de « palimpseste verbo-culturel » est mobilisée afin de rendre compte du fonctionnement en discours d'une expression comme *y a bon banania*, qui reprend le célèbre slogan de la marque *Banania*. L'examen du sigle *BHV* (NdM d'une chaîne de bricolage, *Bazar de l'Hôtel de Ville*) et de son homonyme, correspondant aux initiales de l'arrondissement de Bruxelles-Hal-Vilvorde, montre que l'affectation référentielle peut être multiple, entraînant un risque d'opacité pour les lecteurs non avertis. En revanche, les occurrences discursives du NdM *L'Argus* ont tendance à véhiculer en priorité le sens de « cote », « estimation ». Ensuite, la valeur lexicoculturelle des NdM est mise en valeur par des lexies renvoyant à des toponymes, à des mots du terroir, de la tradition, ainsi que de la culture « cultivée » (littérature, mythologie, art, etc.), dont Michela Tonti explore l'implicite culturel.

Le quatrième chapitre se propose d'étudier l'impact quantitatif des NdM dans le discours : pour ce faire, l'auteure a recours à la notion de *notoriété* telle qu'elle est envisagée par le marketing et qui permet de connaître le point de vue des consommateurs/locuteurs sur une marque. Le recours prioritaire au genre du blog permet d'avoir accès à des échantillons authentiques de discours, avec leur charge évaluative plus ou moins explicite. Tonti se concentre tout d'abord sur les NdM les plus fréquents (avec plus de 1 000 occurrences dans le corpus) et en étudie les marqueurs adjectivaux d'appréciation (ou, cas plus rare, de dépréciation), sans négliger les NdM moins fréquents, car ils s'avèrent eux aussi bien ancrés dans l'imaginaire collectif des locuteurs, grâce notamment à leur ancienneté. Pour finir, est présentée une catégorisation fine des emplois « (du) genre » + NdM, « (du) type » + NdM, oscillant entre axiologie et exemplarité et témoignant de l'intégration des NdM au stock lexicoculturel de la communauté linguistique.

Le cinquième et dernier chapitre, de loin le plus vaste, aborde le thème de la variation tous azimuts des NdM. En premier lieu, il est question de variation orthographique : à ce niveau, il émerge un certain attachement des scripteurs à l'accentuation graphique, parfois par hypercorrection. Au niveau morphologique, la marque du pluriel est l'indice d'une catégorisation : par exemple, *Barbies* peut devenir l'hyperonyme de toute une typologie de poupées partageant les mêmes

traits distinctifs. L'étude du corpus montre une certaine saillance de patrons syntaxiques réguliers, tels « c'est du NdM + (tout pur/ tout craché) », témoignant à la fois d'usages métonymiques, d'une distinction non opératoire entre NdM et nom de produit de la part des locuteurs et de l'axiologisation des propriétés référentielles des NdM. La partie suivante est consacrée à la créativité lexicale qui se forme à partir des NdM. Parmi les matrices morphosémantiques, la dérivation (*rolexé, chanelisant, McDonaldisation*, etc.) se taille la part du lion, suivie par la composition. À cet égard, on peut regretter une certaine confusion entre les procédés de la compositation et de la fractocomposition (voir Sablayrolles, 2019) : comme le rappelle Tonti elle-même, la fractocomposition correspond à la « fusion de deux lexies dont la première est tronquée » (p. 144), on a donc du mal à voir pourquoi *blédi(ne)truc* ou *blédi(ne)cantoche* relèveraient de la compositation plutôt que de la fractocomposition. Parmi les matrices syntactico-sémantiques, la lexie *Sephorette* est présentée comme une métonymie du NdM Sephora, là où nous voyons plutôt une simple suffixation à partir de la base *Sephor-*. Bien que plus marginales, l'auteure relève également des matrices morphologiques (troncation, siglaison), phraséologiques (détournement d'expressions), ainsi que la matrice externe, qu'il s'agisse d'emprunts véritables ou de constructions allogènes du type *disneyfan* ou *appleman*. L'attention se porte ensuite sur la dimension sémantique de la variation des NdM en corpus : à ce propos, les noms *ripolin* et *ripolinage* et le verbe *ripoliner* (ainsi que son participe passé) sont analysés dans leur combinatoire et pour ce qui est de leurs domaines d'application. Étant donné que les acceptions figurées de *ripoliner* paraissent stables et lexicalisées (cf. *Petit Robert 2020*), on peut se demander si la notion de *dégénérescence* du NdM (Petit 2000) n'aurait pas gagné à être prise en compte dans ce contexte. Pour finir, Tonti aborde les emplois métaphoriques et métonymiques des NdM dans des expressions telles que *la barbie de service* ou *être très nike*.

Entité aux contours flous, le NdM nécessite un traitement multidisciplinaire que l'auteure a le mérite de revendiquer. Mis à part certaines redondances et un style parfois un peu trop dense, on peut saluer une contribution importante à l'étude des NdM, qui associe de manière heureuse une approche quantitative à une exploration qualitative. Le recours à un vaste corpus contemporain se montre indispensable pour analyser la dynamique sémantique et syntaxique des NdM en discours et pour mettre au clair des implicites culturels parfois insoupçonnés. On peut parier que cet ouvrage contribuera de manière significative à l'essor des études sur l'onomas-tique commerciale et qu'il servira de repère méthodologique : pour étudier les NdM et leurs avatars, il s'agit en effet de donner la parole aux locuteurs et d'observer les réappropriations linguistiques qu'ils en font au cours d'énoncés authentiques.

Bibliographie

- Altmanova, J. 2013. *Du nom déposé au nom commun. Néologie et lexicologie en discours*. Milan : Educatt.
- Altmanova, J., Le Tallec, G. (éds.). 2019. *Lexicalisation de l'onomastique commerciale. Créer, diffuser, intégrer*. Bern : Peter Lang (Gramm-R. Études de linguistique française).
- Faure, P. 2018. « La néologie commerciale. L'exemple des noms de médicaments ». *Cahiers de lexicologie*, n° 113, p. 129-144.
- Galisson, R., André, J.-C. 1988. *Dictionnaires de noms de marques courants : essai de lexiculture ordinaire*. Paris : Didier.
- Petit, G. 2000. « Un hybride sémiotique. Le nom déposé ». *Lingvisticæ investigationes*, n° XXIII/1, p. 161-192.
- Sablayrolles, J.-F. 2019. *Comprendre la néologie. Conceptions, analyses, emplois*. Limoges : Lambert-Lucas.

Synergies Italie n° 17 / 2021



Annexes



Profils des contributeurs



• Coordinatrices scientifiques •

Annick Farina est maître de conférences en langue française et traduction et Présidente du Centre de Langues étrangères de l'Université de Florence. Après avoir travaillé sur les relations entre lexicographie et société au Canada (Dictionnaires de langue française au Canada, Champion, 2001), elle a publié de nombreux articles sur la lexicographie francophone monolingue et sur la lexicographie bilingue français-italien, analysant différents aspects micro- et macrostructuraux. Elle dirige actuellement l'unité de recherche Lessico multilingue dei Beni Culturali (LBC), qui travaille à la création de ressources lexicographiques plurilingues centrées sur le patrimoine italien et orientées vers les besoins des traducteurs du domaine du tourisme.

Valeria Zotti est maître de conférences en langue et traduction françaises à l'Université de Bologne. Ses recherches portent principalement sur la lexicographie, la linguistique de corpus et le TAL. Elle a publié, entre autres, les ouvrages *La variation des français. Dictionnaires, bases de données, corpus* (avec A. Farina, Champion, 2014), *Informatica umanistica: risorse e strumenti per lo studio del lessico dei beni culturali* (avec A. Pano Alaman, Firenze University Press, 2017), et *The Language of Art and Cultural Heritage: a Plurilingual and Digital Perspective* (avec A. Pano Alaman, Cambridge Scholars Publishing, 2020).

• Auteurs des articles •

Francesca Bisiani est maîtresse-assistante et chercheuse associée au Centre de recherche sur les relations entre le risque et le droit (C3RD) à la Faculté de Droit de l'Université Catholique de Lille. Elle est docteure en Linguistique de l'Université de Paris et de l'Université d'Udine et Trieste. Ses recherches portent actuellement sur la terminologie institutionnelle dans les espaces multilingues et sur l'impact des dispositifs d'intelligence artificielle sur les concepts juridiques et sociopolitiques.

Silvia Calvi est doctorante en littératures étrangères, langues et linguistique, avec une spécialisation en langues étrangères et traduction à l'Université de Vérone. Ses intérêts concernent principalement trois domaines de la linguistique : la terminologie, la lexicographie spécialisée et l'enseignement/apprentissage du français L2.

Sa thèse de doctorat concerne l'extraction semi-automatique et la représentation des collocations en terminologie dans une base de données terminologiques. Elle a mené des études terminologiques dans le secteur du commerce et du sport.

Klara Dankova est docteure en sciences du langage et littératures étrangères à l'Université Catholique du Sacré-Cœur à Brescia-Milan. Sa thèse, soutenue en 2020, porte sur la terminologie des fibres textiles en français à partir du XVIII^e siècle (titre de la thèse : *La néologie et la construction morphosémantique dans la communication spécialisée en français : les fibres chimiques entre la synchronie et la diachronie*). Dans ses recherches, elle s'intéresse notamment à la terminologie en français dans une perspective synchronique et diachronique.

Cosimo De Giovanni est enseignant-chercheur en linguistique française à l'Université de Cagliari. Il s'occupe principalement de lexicographie, de terminologie, de linguistique de corpus et de phraséologie. Plus récemment, il s'est consacré à l'étude de l'orthographe et à l'histoire de la langue française. Il est vice-président de l'Association italienne de phraséologie et parémiologie *Phrasis*.

Marco Fasciolo est maître de conférences en linguistique générale et française à l'Université Paris Sorbonne et membre du laboratoire STIH. Il est l'auteur de *Rethinking Presuppositions* (Cambridge Scholars, 2019), (avec G. Gross) *La sintassi del lessico. Manuale di linguistica aperto all'informatica e alla filosofia* (UTET, 2020) et *Grammaire Philosophique du Verbe* (Classiques Garnier, en cours d'édition). Ses recherches portent sur l'articulation entre les structures conceptuelles de l'ontologie naturelle d'une part, et le lexique et la syntaxe de l'autre.

Eleonora Marzi est linguiste à l'Université de Bologne. Elle s'occupe d'interculturalité et d'analyse du discours en s'appuyant sur des technologies informatiques telles que le web sémantique et le traitement du langage naturel. Parmi ses derniers articles, citons : « Le texte liquide entre linéarité et causalité : les caractéristiques de l'hypertexte dans la littérature numérique » (2021), « L'implicito nella linguistica: uno studio sull'immaginario letterario nella pubblicità » (2020), « La néonymie à l'épreuve de la textométrie : le cas du terme *francisation* » (2020).

Adriana Orlandi est enseignante-chercheuse à l'Université de Modène et de Reggio d'Émilie, en Italie, où elle enseigne la traduction. Elle s'intéresse depuis longtemps aux relations entre sémantique et syntaxe, avec une attention toute particulière pour la sémantique de la modification adjectivale. Elle a publié *Le paradoxe de l'adjectif* (De Boeck, 2020). Ses recherches portent également sur la traduction et la traductologie, ainsi que sur la phraséologie.

Federica Vezzani est chercheuse post-doc à l'Université de Padoue. Ses recherches portent sur la terminologie, la terminographie et la traduction spécialisée. Elle a conçu et implémenté la ressource terminologique multilingue et polyvalente TriMED pour le domaine médical, l'outil de compilation de fiches terminologiques FAIRterm et la base de données CAMEO pour le domaine commercial. Elle est membre du comité de direction du projet européen YourTerm.org au sein de l'Unité de Coordination de Terminologie (TermCoord).

• **Auteurs des comptes rendus** •

Annalisa Buonocore a obtenu un doctorat de recherche en sociolinguistique à l'Université de Heidelberg. Elle est enseignante titulaire à l'école secondaire de second degré et enseigne aussi à l'Université de la Vallée d'Aoste. Elle a rédigé *Dialettali e neodialettali in inglese* (Cofine, 2003) et *Varietà dialettali microareali della costiera amalfitana* (Peter Lang, 2009). Ses travaux portent sur la sociolinguistique et la didactique du FLE.

Manon Hayette est titulaire d'un Master en traduction multidisciplinaire (Faculté de Traduction et d'Interprétation — École d'interprètes Internationaux) à l'Université de Mons (Belgique). Ses domaines de recherche sont, entre autres, la traductologie, la lexicologie, la lexicographie, la sociolinguistique, la phraséologie, la littérature comparée et les études chinoises. Elle entamera bientôt la rédaction d'une thèse de doctorat portant sur le traitement lexicographique des *chengyu*, unités phraséologiques quadrisyllabiques du chinois.

Maria Margherita Mattioda est maître de conférences en langue et traduction – langue française auprès du Département de Langues et Littératures étrangères de l'Université de Turin, où elle enseigne la traduction et la communication spécialisée. Ses domaines de recherche sont la communication professionnelle dans ses aspects lexico-culturels et rhétorico-pragmatiques, les discours des entreprises et la traduction spécialisée. Elle a coordonné, avec Mélanie Bourdaa, le n° 13 de *Synergies Italie* sur *Le storytelling dans tous ses états* en 2017.

Giovanni Tallarico est maître de conférences en langue et traduction française à l'Université de Vérone. Ses recherches portent sur la lexicographie bilingue (*La dimension interculturelle du dictionnaire bilingue*, Honoré Champion, 2016), la lexicologie (notamment la néologie et l'étude des emprunts), la langue de la publicité, la terminologie (du tourisme et du sport) et la traductologie. Il a aussi travaillé sur les aspects linguistiques du texte littéraire, en abordant des auteurs tels que J. Tardieu, E. Carrère, R. Sattouf et V. Novarina. De la fiction radiophonique) et *Histoire de Bouvard et Pécuchet, copistes* (libre adaptation du roman de G. Flaubert), il collabore aussi avec des dessinateurs dans des œuvres à quatre mains,

parmi lesquelles *Géométrie dans la poussière* (avec Killoffer, Verticlaes, 2004), *Les Carnets de Gordon McGuffin* (avec Nicolas de Crécy, Futuropolis, 2009) ou encore *Cendres des hommes et des bulletins* (avec S. Aquindo, Le Tripode, 2016).

Liana Tronci est maître de conférences en Linguistique générale et historique à l'Université pour Étrangers de Sienne, où elle enseigne la sociolinguistique et la typologie des langues. Ses recherches portent sur les langues classiques et les langues romanes.

Projet pour le n° 18 / 2022



Plurilinguisme et littératie dans les pays francophones et italophones **Coordonné par Sara Vecchiato (Universite d'Udine, Italie)**

Dans un cadre général de grands changements sociaux, accélérés par la crise pandémique mondiale, les compétences linguistiques s'inscrivent dans des espaces sociaux en tension (Berthoud, Burger, 2014) et augmentent leur importance à bien des égards. En effet, sur un marché où l'impact des technologies de l'information est en train de remodeler la division du travail (Frey, Osborne, 2017 ; Levy, Murnane, 2012), plusieurs professions de niveau « moyen » disparaissent. Face à cet *effet sablier* (Vittori, 2009) partageant de plus en plus la société, le fait de parler plusieurs langues est important pour l'épanouissement personnel et professionnel de l'individu (Grin, 2015). La formation linguistique touche de près l'ensemble des compétences mobilisées en lecture et en écriture, désignées sous le nom de « littératie » (ou « lettrisme », cf. Pruvost, 2019) : alors que l'évolution de la société impose des niveaux de plus en plus élevés de littératie, les phénomènes de l'analphabétisme et de l'illettrisme marginalisent des fractions importantes d'adultes (OCDE, 2020 ; Vertecchi, Gallina, 2007). Puisque le plurilinguisme est configuré comme un « capital culturel » (Bourdieu, Passeron, 1970 ; Coffey, 2018), l'éducation aux langues a une valeur d'émancipation sociale (De Mauro, 2018).

Premièrement, la question du plurilinguisme est d'une grande actualité à un moment historique où l'Europe débat de son avenir et où l'équilibre entre les langues est en train d'être redessiné (Ginsburgh et al., 2017 ; Somssich, 2016). En effet, s'il existe un consensus sur le fait que pouvoir s'exprimer et accéder à des services dans sa propre langue est un droit (Varenes, 2016), la co-édition de textes en plusieurs langues au niveau européen soulève un certain nombre de problématiques (Raus, 2010), si bien que le coût du maintien du plurilinguisme au niveau européen fait débat (Ginsburgh, Moreno-Ternero 2020). D'autre part, il est bien connu qu'au niveau économique, le manque de compétences en langues étrangères parmi les employés entraîne un préjudice qui a été quantifié dans plusieurs études (Commission Européenne 2009 ; Holmes, 2018), tandis que certains chercheurs ont souligné le rôle du plurilinguisme comme stratégie de communication efficace dans les entreprises (Gerolimich, Vecchiato, 2016 ; Lüdi, 2020). La valeur économique des langues étrangères dans la carrière professionnelle d'un individu semble être une donnée acquise, quoique les

études scientifiques sur ce sujet aient une extension variable selon les pays, notamment en ce qui concerne l'Italie (Gazzola, 2017). Toujours est-il que le modèle d'une langue véhiculaire unique a longtemps été remis en question (Hogan-Brun, 2017 ; Mattioda, 2013). Parmi les alternatives possibles, la pratique de l'*intercompréhension*, ou *lingua receptiva*, a reçu l'attention de la Commission européenne (Ten Thije et al., 2017) et pourrait représenter un tournant communicatif, notamment pour les langues romanes (Badelita, 2016). En même temps, les compétences plurilingues et pluriculturelles (Coste et al., 2009 ; Gibault, 2018) risquent d'être présentées de manière enthousiaste comme la panacée à toutes les inégalités sociolinguistiques, alors qu'elles peuvent aussi faire l'objet d'exploitation par le marché, sans que l'individu en tire aucun bénéfice (Duchêne, 2011 ; Duchêne, Heller, 2012).

Deuxièmement, le plurilinguisme comme gage de la connaissance et de la reconnaissance de l'Autre constitue le pari de l'Observatoire Européen du plurilinguisme qui a soumis au Conseil de l'Europe et au Parlement européen une *Charte européenne du plurilinguisme* (Observatoire européen du plurilinguisme, 2005 ; Telmon, 2018). En effet, l'utilisation de répertoires linguistiques plurilingues (Lüdi, 2006) permet à la personne d'entrer en contact avec d'autres cultures et d'explorer son identité d'un autre point de vue (Kinginger, 2008). Toutefois, le plurilinguisme n'est ni conçu ni accepté de la même manière au niveau international et au niveau national : la pression en faveur du monolinguisme justifie alors des revendications empreintes de purisme qui justifient l'abandon des répertoires linguistiques plurilingues au niveau local (Cerquiglioni, 2003 ; sous presse ; Telmon *et al.*, 2012). Cette divergence sur le plan symbolique devient d'une pertinence absolue dans certains contextes tels que les régions frontalières (Buonocore, 2018). Lorsque, d'autre part, l'éducation linguistique intervient positivement dans la formation de l'individu, le plurilinguisme peut donner lieu à une expérience privilégiée de réflexion sur soi, comme c'est le cas dans la proposition pédagogique de « l'autobiographie linguistique » (Favero, Sofia, 2018).

Troisièmement, le plurilinguisme touche directement la transmission du savoir, bien que ce rôle soit resté relativement discret (Berthoud, Gajo, 2020). En effet, le monolinguisme de plus en plus répandu dans la communication scientifique comporte le risque pour les langues qui ne sont pas « hyper-centrales » (au sens de Calvet, 1999), comme le français et l'italien, de mener des batailles d'arrière-garde pour défendre un héritage culturel du passé (Cabiddu, 2017 ; Hamel, 2013). Inversement, la « littératie plurilingue » peut entrer en jeu dans les entreprises, les institutions, et surtout dans l'éducation (Egli Cuenat et al., 2020 ; Legros et al., 2009 ; Lo Bianco, 2000). Les éducateurs se confronteront alors à une valorisation fortement asymétrique des langues selon le pays et au phénomène de l'illettrisme dans la langue maternelle accompagné de la scolarisation de l'enfant dans une langue qu'il maîtrise mal ou pas du tout (Conseil de l'Europe et Assemblée parlementaire, 2006). Voilà pourquoi les études de langues étrangères et le temps consacré à la lecture-écriture dans l'enseignement font l'objet d'une attention particulière dans la planification de l'éducation au sein de l'Union européenne (DEPP, 2020).

Ce numéro de *Synergie Italie* se veut une occasion de faire le point sur le plurilinguisme et la littératie, en les considérant à la fois de façon autonome et à travers leurs points d'intersection. Un accent particulier sera mis sur leur rôle dans les pays francophones et italophones. Voici une liste (non exhaustive) des questions examinées :

- plurilinguisme et identité(s) sociale(s), compétences pluriculturelles, droits linguistiques ;
- littératie plurilingue, compétences d'écriture dans un milieu multilingue ;
- illettrisme en L1, éducation au plurilinguisme, intercompréhension, lingua receptiva ;
- valeur économique des langues, marché des langues ;
- communautés translinguales, communautés plurilingues et écritectes.

Un appel à contributions a été lancé en février 2021.

Contact : synergies.italie@gmail.com

Consignes aux auteurs



- 1** L'auteur aura pris connaissance de la politique éditoriale générale de l'éditeur (le Gerflint) et des normes éditoriales et éthiques figurant sur le site du Gerflint et de la revue. Les propositions d'articles seront envoyées pour évaluation à synergies.italie@gmail.com avec un court CV résumant son cursus et ses axes de recherche en pièces jointes. L'auteur recevra une notification. Les articles complets seront ensuite adressés au Comité de rédaction de la revue selon les consignes énoncées dans ce document. Tout texte ne s'y conformant pas sera retourné. Aucune participation financière ne sera demandée à l'auteur pour la soumission de son article. Il en sera de même pour toutes les expertises des textes (articles, comptes rendus, résumés) qui parviendront à la Rédaction.
- 2** L'article sera inédit et n'aura pas été envoyé à d'autres lieux de publication. Il n'aura pas non plus été proposé simultanément à plusieurs revues du Gerflint. L'auteur signera une « déclaration d'originalité et de cession de droits de reproduction et de représentation ». Un article ne pourra pas avoir plus de deux auteurs.
- 3** Proposition et article seront en langue française. Les articles (entrant dans la thématique ou épars) sont acceptés, toujours dans la limite de l'espace éditorial disponible. Ce dernier sera réservé prioritairement aux chercheurs francophones (doctorants ou post-doctorants ayant le français comme langue d'expression scientifique) locuteurs natifs de la zone géolinguistique que couvre la revue. Les articles rédigés dans une autre langue que le français seront acceptés dans la limite de 3 articles non francophones par numéro, sous réserve d'approbation technique et graphique. Dans les titres, le corps de l'article, les notes et la bibliographie, la variété éventuelle des langues utilisées pour exemplification, citations et références est soumise aux mêmes limitations techniques.
- 4** Les articles présélectionnés suivront un processus de double évaluation anonyme par des pairs membres du comité scientifique, du comité de lecture et/ou par des évaluateurs extérieurs. L'auteur recevra la décision du comité.
- 5** Si l'article reçoit un avis favorable de principe, son auteur sera invité à procéder, dans les plus brefs délais, aux corrections éventuelles demandées par les évaluateurs et le comité de rédaction. Les articles, à condition de respecter les correctifs demandés, seront alors soumis à une nouvelle évaluation du Comité de lecture, la décision finale d'acceptation des contributions étant toujours sous réserve de la décision des experts du Conseil scientifique et technique du Gerflint et du Directeur des publications.
- 6** La taille de police unique est 10 pour tout texte proposé (présentation, article, compte rendu) depuis les titres jusqu'aux notes, citations et bibliographie comprises). Le titre de l'article, centré, en gras, n'aura pas de sigle et ne sera pas trop long. Le prénom, le nom de l'auteur (en gras, sans indication ni abréviation de titre ou grade), de son institution, de son pays et son adresse électronique (professionnelle de préférence et à la discrétion de l'auteur) seront également centrés et en petits caractères. L'auteur possédant un identifiant ORCID ID (*identifiant ouvert pour chercheur et contributeur*) inscrira ce code en dessous de son adresse. Le tout sera sans couleur, sans soulignement ni hyperlien.

7 L'auteur fera précéder son article d'un résumé condensé ou synopsis de 6-8 lignes maximum suivi de 3 ou 5 mots-clés en petits caractères, sans majuscules initiales. Ce résumé ne doit, en aucun cas, être reproduit dans l'article.

8 L'ensemble (titre, résumé, mots-clés) en français sera suivi de sa traduction en italien puis en anglais. En cas d'article non francophone, l'ordre des résumés est inchangé. Les mots-clés seront séparés par des virgules et n'auront pas de point final.

9 La police de caractère est Times New Roman, taille 10, interligne 1. Le texte justifié, sur fichier Word, format doc, doit être saisi au kilomètre (retour à la ligne automatique), sans tabulation ni pagination ni couleur. La revue a son propre standard de mise en forme.

10 L'article doit comprendre entre 15 000 et 30 000 signes, soit 6-10 pages Word, éléments visuels, bibliographie, notes et espaces compris. Sauf commande spéciale de l'éditeur, les articles s'éloignant de ces limites ne seront pas acceptés. La longueur des comptes rendus de lecture ne dépassera pas 2 pages en Word. Comptes rendus et entretiens seront en langue française.

11 Tous les paragraphes (sous-titres en gras sans sigle, petits caractères) seront distincts avec un seul espace. La division de l'article en 1, 2 voire 3 niveaux de titre est suffisante.

12 Les mots ou expressions que l'auteur souhaite mettre en relief seront entre guillemets ou en *italiques*. Le soulignement, les caractères gras et les majuscules ne seront en aucun cas utilisés, même pour les noms propres dans les références bibliographiques, sauf la majuscule initiale.

13 Les notes, brèves de préférence, en nombre limité, figureront en fin d'article avec appel de note automatique continu (1,2,...5 et non i,ii...iv). L'auteur veillera à ce que l'espace pris par les notes soit réduit par rapport au corps du texte.

14 Dans le corps du texte, les renvois à la bibliographie se présenteront comme suit: (Dupont, 1999 : 55).

15 Les citations, toujours conformes au respect des droits d'auteurs, seront en italiques, taille 10, séparées du corps du texte par une ligne et sans alinéa. Les citations courtes resteront dans le corps du texte. Les citations dans une langue autre que celle de l'article seront traduites dans le corps de l'article avec version originale en note.

16 La **bibliographie** en fin d'article précèdera les notes (sans alinéa dans les références, ni majuscules pour les noms propres sauf à l'initiale). Elle s'en tiendra principalement aux ouvrages cités dans l'article et s'établira par classement chrono-alphabétique des noms propres. Les bibliographies longues, plus de 15 références, devront être justifiées par la nature de la recherche présentée. Les articles dont la bibliographie ne suivra pas exactement les consignes 14, 17, 18, 19 et 20 seront retournés à l'auteur. Le tout sans couleur ni soulignement ni lien hypertexte.

17 **Pour un ouvrage**

Baume, E. 1985. *La lecture – préalables à sa Pédagogie*. Paris : Association Française pour la lecture.

Fayol, M. et al. 1992. *Psychologie cognitive de la lecture*. Paris: PUF.

Gaonac'h, D., Golder, C. 1995. *Manuel de psychologie pour l'enseignement*. Paris : Hachette.

18 **Pour un ouvrage collectif**

Morais, J. 1996. La lecture et l'apprentissage de la lecture : questions pour la science. In : *Regards sur la lecture et ses apprentissages*. Paris : Observatoire National de la lecture, p.49-60.

19 Pour un article de périodique

Kern, R.G. 1994. « The Role of Mental Translation in Second Language Reading ». *Studies in Second Language Acquisition*, n°16, p. 41-61.

20 Pour les références électroniques (jamais placées dans le corps du texte mais toujours dans la bibliographie), les auteurs veilleront à adopter les normes indiquées par les éditeurs pour citer ouvrages et articles en ligne. Ils supprimeront hyperlien, couleur et soulignement automatique et indiqueront la date de consultation la plus récente [consulté le], après vérification de leur fiabilité et du respect du Copyright.

21 Les textes seront conformes à la typographie française.

22 Graphiques, schémas, figures, photos éventuels seront envoyés à part au aux formats Word et PDF ou JPEG, en noir et blanc uniquement, avec obligation de références selon le *copyright* sans être copiés/collés mais scannés à plus de 300 pixels. Les articles contenant un nombre élevé de figures et de tableaux et/ou de mauvaise qualité scientifique et technique ne seront pas acceptés. L'éditeur se réserve le droit de refuser les tableaux (toujours coûteux) en redondance avec les données écrites qui suffisent bien souvent à la claire compréhension du sujet traité.

23 Les captures d'écrans sur l'internet et extraits de films ou d'images publicitaires seront refusés. Toute partie de texte soumise à la propriété intellectuelle doit être réécrite en Word avec indication des références, de la source du texte et d'une éventuelle autorisation.

NB : Toute reproduction éventuelle (toujours en noir et blanc) d'une image, d'une photo, d'une création originale et de toute œuvre d'esprit exige l'autorisation écrite de son créateur ou des ayants droit et la mention de paternité de l'œuvre selon les dispositions en vigueur du Code français de la propriété intellectuelle protégeant les droits d'auteurs. L'auteur présentera les justificatifs d'autorisation et des droits payés par lui au propriétaire de l'œuvre. Si les documents sont établis dans un autre pays que la France, les pièces précitées seront traduites et légalisées par des traducteurs assermentés ou par des services consulaires de l'Ambassade de France. Les éléments protégés seront publiés avec mention obligatoire des sources et de l'autorisation, dans le respect des conditions d'utilisation délivrées par le détenteur des droits d'auteur.

24 Seuls les articles conformes à la politique éditoriale et aux consignes rédactionnelles, seront édités, publiés, mis en ligne sur le site web de l'éditeur et diffusés en libre accès par lui dans leur intégralité. La date de parution dépendra de la coordination générale de l'ouvrage par le rédacteur en chef. L'éditeur d'une revue scientifique respectant les standards des agences internationales procède à l'évaluation de la qualité des projets à plusieurs niveaux. L'éditeur, ses experts ou ses relecteurs (évaluation par les pairs) se réservent le droit d'apprécier si l'œuvre convient, d'une part, à la finalité et aux objectifs de publication, et d'autre part, à la qualité formelle de cette dernière. L'éditeur dispose d'un droit de préférence.

25 Les prépublications de l'article et de ses métadonnées ne sont pas autorisées. Une fois éditée sur gerflint.fr, seule la version « PDF-éditeur » de l'article peut être déposée pour archivage dans les répertoires institutionnels de l'auteur exclusivement, avec mention exacte des références et métadonnées de l'article. L'archivage de numéros complets est interdit. Tout signalement ou référencement doit respecter les normes internationales et le mode de citation de l'article, tels que dûment spécifiés dans la politique de la revue. Par ailleurs, les Sièges, tant en France qu'à l'étranger, n'effectuent aucune opération postale, sauf accord entre le Gerflint et un organisme pour participation financière au tirage.



Synergies Italie, n° 17 /2021
Revue du GERFLINT
Groupe d'Études et de Recherches
pour le Français Langue Internationale

En partenariat avec
la Fondation Maison des Sciences de L'Homme de Paris

Président d'Honneur : Edgar Morin

Fondateur et Président : Jacques Cortès

Conseillers et Vice-Présidents : Ibrahim Al Balawi, Serge Borg et Nelson Vallejo-Gomez

PUBLICATIONS DU GERFLINT

<https://catalogue.bnf.fr/ark:/12148/cb14524060t>

ISNI 0000 0001 1956 5800

IdRef : 077342070

Le Réseau des Revues Synergies du GERFLINT

Synergies Afrique centrale et de l'Ouest	Synergies Monde
Synergies Afrique des Grands Lacs	Synergies Monde Arabe
Synergies Algérie	Synergies Monde Méditerranéen
Synergies Argentine	Synergies Pays Germanophones
Synergies Amérique du Nord	Synergies Pays Riverains de la Baltique
Synergies Brésil	Synergies Pays Riverains du Mékong
Synergies Chili	Synergies Pays Scandinaves
Synergies Chine	Synergies Pologne
Synergies Corée	Synergies Portugal
Synergies Espagne	Synergies Roumanie
Synergies Europe	Synergies Royaume-Uni et Irlande
Synergies France	Synergies Russie
Synergies Inde	Synergies Sud-Est européen
Synergies Iran	Synergies Tunisie
Synergies Italie	Synergies Turquie
Synergies Mexique	Synergies Venezuela

Essais francophones : Collection scientifique du GERFLINT

Direction du Pôle Éditorial International :

Sophie Aubin (Universitat de València, Espagne)

Contact: gerflint.edition@gmail.com

Site officiel : <https://www.gerflint.fr>

Webmestre : Thierry Lebeau-pin (France)

Synergies Italie, n° 17 / 2021

Couverture, conception graphique et mise en page : Emilie Hiesse (*Créactiv'*) - France

© GERFLINT – Sylvains-les-Moulins – France – Copyright n° ZSN67E3

ARK : <https://catalogue.bnf.fr/ark:/12148/cb42702496d>

Bibliothèque Nationale de France - Juillet 2021

GERFLINT

Groupe d'Études et de Recherches pour le Français
Langue internationale

Programme mondial de diffusion scientifique
francophone en réseau

www.gerflint.fr

Ce numéro de *Synergies Italie* est l'occasion de présenter quelques recherches en cours en France et en Italie dans le domaine linguistique qui ont des applications concrètes au sein des industries des langues, ainsi que d'analyser les produits (corpus, bases de données terminologiques, outils de traitement automatique) existants, un quart de siècle après l'essor de ces industries. Nous nous sommes intéressées aux retombées que l'emploi généralisé des technologies du traitement automatique des langues (TAL) a eues dans ces deux pays sur le grand public, sur les professionnels de la langue (traducteurs, terminologues, lexicographes) et dans le monde de la recherche universitaire. Nous nous sommes interrogées en particulier sur la présence de la langue française et de la langue italienne (et de la traduction de l'une à l'autre) à l'intérieur des nouveaux instruments développés.

Annick Farina et Valeria Zotti