



**HAL**  
open science

## Apport d'un corpus de presse spécialisé parallèle français/italien à l'analyse des marqueurs et de la relation de méronymie

Anne Condamines, Marie-Pierre Escoubas-Benveniste, Silvia Federzoni

### ► To cite this version:

Anne Condamines, Marie-Pierre Escoubas-Benveniste, Silvia Federzoni. Apport d'un corpus de presse spécialisé parallèle français/italien à l'analyse des marqueurs et de la relation de méronymie. *Études de linguistique appliquée : revue de didactologie des langues-cultures et de lexiculturologie*, 2022, 208 (4), pp.429-446. 10.3917/ela.208.0049 . halshs-03890533

**HAL Id: halshs-03890533**

**<https://shs.hal.science/halshs-03890533>**

Submitted on 16 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## APPORT D'UN CORPUS DE PRESSE SPÉCIALISÉ PARALLÈLE FRANÇAIS/ITALIEN À L'ANALYSE DES MARQUEURS ET DE LA RELATION DE MÉRONYMIE

*Résumé.* Cet article décrit les apports d'un corpus spécialisé parallèle français italien pour l'étude de la méronymie. Ces apports concernent d'une part la définition de nouveaux marqueurs de relation et, d'autre part, la définition de la relation méronymique elle-même. En effet, le départ entre cette relation et les relations temporelle, causale ou de conséquence apparaît parfois difficile à faire. Par ailleurs, dans certains extraits, la relation ne peut être qu'inférée. L'étude met au jour ces différentes variations, en lien avec les possibilités de la traduction ; elle montre aussi que, parfois, les ambiguïtés sont levées lors du passage d'une langue à l'autre. Interviennent alors la question de la confiance que l'on peut accorder à ces choix, qui ont un impact sur le mode de représentation et, dans certains cas, la nécessité de faire intervenir un expert du domaine pour valider (ou pas) ces choix.

### INTRODUCTION

L'étude présentée dans cet article s'inscrit dans la perspective générale de construire un réseau conceptuel à partir d'un corpus en utilisant une approche semi-automatique. Dans cet objectif, nous évaluons l'apport de l'utilisation d'un corpus parallèle (textes en français traduits en italien). On sait en effet que le processus de traduction autorise une légère inflexion du sens qui en permet un éclairage plus fin pour le lecteur (Lavault-Olléon, 1996). Pour construire un réseau conceptuel, par hypothèse supposé commun au corpus d'origine et au corpus traduit, l'apport de ce « décalage » dans la traduction peut être très utile. C'est en tout cas l'hypothèse que nous faisons. Dans cet article, nous nous focalisons sur la méronymie, avec, comme point de départ une liste de marqueurs recensés dans les deux langues et projetés sur chacun des corpus. Nous observons ce que les équivalents dans l'autre langue apportent, soit à la liste des marqueurs initiaux, soit à la caractérisation de la relation méronymique elle-même.

La section 1 rend compte des études utilisant des corpus en terminologie et, en particulier, des corpus alignés. Elle mentionne aussi comment les études en traductologie sont convoquées dans notre approche. Enfin, les premiers résultats obtenus, pour l'hyperonymie, sont résumés. Les sections 2 et 3 détaillent les résultats pour la méronymie : spécificités du marquage et de la sémantique de la méronymie. Puis sont présentés deux apports de l'utilisation d'un corpus parallèle. Il s'agit tout d'abord de l'amélioration de la liste des marqueurs. Nous montrons ensuite comment le fonctionnement de la relation de méronymie est éclairé par l'alignement de corpus : repérage d'ambiguïtés ou, au contraire, de choix interprétatifs. Nous adoptons un point de vue terminologique, c'est-à-dire à la fois celui d'un terminologue qui souhaite construire un réseau et celui d'un linguiste-terminologue qui veut comprendre comment fonctionne la méronymie en discours spécialisé.

### 1. CONTEXTE DE L'ÉTUDE ET RÉSULTATS PRÉCÉDENTS

Cette étude fait suite à un certain nombre d'analyses concernant les marqueurs de relations conceptuelles tels qu'ils se manifestent dans les corpus spécialisés. Cette section s'intéresse à ces travaux, à ceux qui ont utilisé des corpus spécialisés parallèles et présente les premiers résultats, que nous avons obtenus pour l'hyperonymie.

#### 1.1 Études terminologiques et corpus

La section 1.1.1 rappelle les travaux existants sur le repérage des relations conceptuelles et sur la question des variations. La section 1.1.2 concerne l'utilisation de corpus alignés. La section 1.1.3 rappelle les deux principales méthodes d'extraction bilingue de termes.

##### 1.1.1. Marqueurs de relations conceptuelles

L'étude systématique des marqueurs de relations conceptuelles s'est développée à partir des années 1990, au moment du rapprochement de la terminologie avec le traitement automatique de la langue (Condamines et Picton, 2022). Il est apparu alors que la constitution de réseaux terminologiques pouvait être assistée en recherchant dans les corpus spécialisés les éléments langagiers (les patrons linguistiques)

qui pouvaient être associés à une interprétation relationnelle, de manière plus ou moins systématique. Ces patrons sont les réalisations linguistiques de ce que Meyer a appelé des *Knowledge Rich Contexts* (KRC) (Meyer, 2001). Les premiers travaux portant sur les KRC se sont intéressés aux marqueurs non dépendants du corpus, c'est-à-dire établissant une relation stable entre marqueur et relation, indépendante de la nature du corpus et, par conséquent, présents aussi dans les corpus généraux (Condamines, 2002). D'ailleurs, ces marqueurs « définitoires » ont pour la plupart été empruntés aux marqueurs identifiés lors d'études en lexicologie de corpus généraux. Mais, assez rapidement des travaux se sont orientés vers la prise en compte de la variation, concernant la présence et l'interprétation des marqueurs en fonction du domaine et du genre textuel, leur « portabilité » (Condamines 2008, Marshman et al. 2008).

### 1.1.2. Corpus parallèles dans les études en terminologie

La prise en compte de corpus traduits n'est pas un élément nouveau en terminologie dans la mesure où les premières études étaient consacrées à la constitution de terminologies bi- ou plurilingues. Cette utilisation de textes traduits a d'ailleurs parfois été mise en cause car la traduction était considérée comme trop éloignée des usages réels (Dubuc, 1974).

Toutefois, tout comme pour la lexicologie, la mise à disposition de corpus en format électronique a stimulé les études et le développement de la terminologie de corpus, en particulier la recherche de marqueurs de relations dans différentes langues, dans une perspective monolingue (Auger et Barrière, 2008 ; Feliu et Cabré-Castellví, 2002 ; León-Araúz et al., 2016). Étonnamment, alors qu'un certain nombre de travaux en Traitement Automatique de la Langue (TAL) se sont intéressés à l'extraction de termes à partir de corpus parallèles (voir 1.1.3), peu de travaux se sont intéressés à l'utilisation systématique de corpus traduits pour repérer des relations conceptuelles. Dans cette perspective, ce sont plutôt des corpus comparables (traitant du même « sujet ») qui sont utilisés (voir par exemple Hmida et al. 2016). Notons toutefois l'étude de Marshman et al., (2009) qui utilise un corpus parallèle anglais-français pour repérer de nouveaux termes et de nouveaux marqueurs de relation, et surtout celle de Pearson (Pearson, 2000) qui, en évoquant l'intérêt d'utiliser un corpus traduit pour travailler sur les marqueurs de relation, a été à l'origine de notre étude systématique.

### 1.1.3. Extraction bilingue de terminologies à partir de corpus alignés

Plusieurs études se sont intéressées à l'utilisation de corpus alignés pour extraire des termes. Ces travaux proposent différentes méthodes pour l'extraction terminologique automatique. Deux approches peuvent être distinguées pour l'extraction terminologique : les approches linguistiques et les approches statistiques (Macken et al. 2013). Macken et al. 2013 montrent comment extraire des candidats termes, à partir de « *chunks* », et utiliser des méthodes statistiques pour en déterminer le degré de spécificité. Leurs résultats mettent en évidence l'intérêt d'une exploitation de l'alignement des *chunks* de différentes langues : leur système d'extraction multilingue s'avère plus performant que les systèmes monolingues (Macken et al. 2013). Sahlgren et Karlgren (2005) s'intéressent eux aussi à l'extraction de lexique bilingue. Ils proposent une approche distributionnelle permettant d'extraire des lexiques bilingues à partir des espaces vectoriels obtenus grâce à un corpus parallèle. Cette méthode s'avère efficace, reproductible sur des corpus relevant de différents domaines, des corpus parallèles de différentes langues et elle ne nécessite pas de recourir à des ressources externes.

Quelle que soit l'approche utilisée, l'accent est mis sur la reconnaissance et l'extraction des termes et non sur le repérage des relations conceptuelles.

Nous proposons une approche qui exploite à la fois une ressource externe – deux listes préétablies de marqueurs de relation en français et en italien – ainsi que l'alignement et la bidirectionnalité de la traduction afin d'extraire du corpus parallèle et aligné des couples de termes bilingues mis en relation.

## 1.2 Travaux en traductologie

Les études traductologiques basées sur de grands corpus ont conduit à une avancée majeure dans les années 1990. Les régularités intuitivement pressenties concernant l'opération de traduction ont alors pu être vérifiées et précisées. Les données des textes originaux et traduits ont confirmé des tendances

communes, ce qui a conduit Baker (1996) à énoncer l'existence d'« universaux de la traduction », quatre traits généraux communs résultant de l'action traduisante envisagée comme un usage particulier de la langue : opérations d'explicitation, de simplification, de normalisation et de nivellement. Baker définit ainsi ces types de caractéristiques : l'explicitation consiste à donner une forme linguistique à une information restée implicite dans le texte original ; la simplification de la forme dans le texte traduit peut concerner les plans syntaxique, lexical et stylistique du texte ; la standardisation (ou normalisation) fait que le texte cible va tendanciellement se conformer à la dimension idiomatique de la langue cible, parfois trop, par rapport au texte source, le nivellement fait que le texte résultant est linguistiquement plus homogène que le texte original<sup>1</sup>. Ces traits sont l'expression de quatre axes de variation qui sont typiques de toute opération traduisante et donc du texte qui en résulte. La traduction, en tant que résultat d'un processus communicatif contraint, tend à présenter, au-delà des langues en présence, certains écarts par rapport au texte original<sup>2</sup>.

Dans son étude sur les verbes de cause menée sur le corpus *Le corpus français-italien du Monde Diplomatique – Il Manifesto* (MDIL) (décrit en 1.3.2), Di Vito (2016) part d'une liste prédéfinie de verbes français paraphrasables par « être la cause de » et étudie la synonymie de ces verbes et la relation d'équivalence sémantique établie dans le corpus de presse avec des verbes italiens. Il s'agit d'une étude lexicologique contrastive orientée par le processus de traduction. Or, le corpus bilingue peut être aussi exploré selon les deux directions de l'alignement, et donner ainsi le cadre complet des contextes équivalents. Dans notre étude, nous partons d'un ensemble de contextes bilingues formellement marqués et nous utilisons précisément la variation induite par la traduction comme une ressource pour enrichir une liste de marqueurs et pour mieux comprendre le fonctionnement de la méronymie.

La traduction offre en effet la possibilité de compiler des corpus bilingues dans lesquels un ensemble unique de connaissances est censé être présent dans deux langues distinctes. Dans notre précédent travail, concernant l'apport de la traduction à l'étude de la relation d'hyperonymie (Condamines et al. 2021), résumé en 1.3, nous montrons que c'est la tendance à expliciter qui permet d'identifier de nouvelles informations utiles – par exemple des hyperonymes – pour la construction des réseaux terminologiques. Pour la méronymie, nous verrons que c'est plutôt la simplification qui est pertinente.

### 1.3 Mise en place de l'étude et résultats obtenus pour l'hyperonymie

Dans cette section nous présentons les ressources et la méthode utilisées pour mener à bien cette étude.

#### 1.3.1. Liste des marqueurs

Pour réaliser la recherche, nous disposons d'une liste de marqueurs lexico-syntaxiques de relations sémantiques d'hyperonymie, de méronymie et de cause constituée par Lefeuve (2017) dans le cadre du projet CRISTAL (Contextes Riches en connaissanceS pour la TrAduction terminologique)<sup>3</sup>. Cette liste (MAR-REL) est disponible sur le site Redac<sup>4</sup>. A partir de cette ressource, nous avons enrichi la liste des marqueurs français et pour l'italien, nous avons traduit les marqueurs français en utilisant deux dictionnaires électroniques bilingues *Garzanti* et *Il Boch*<sup>5</sup>. Finalement, nous disposons, pour

1 Pour Baker (1996 : 180-184) l'explicitation est « *a tendency to spell things out rather than leave them implicit in translation* » ; la simplification est « *the tendency to simplify the language used in translation* » ; la « standardization » est « *the tendency to exaggerate features of the target language and to conform to its typical patterns* » et le nivellement « *livelling out* » est « *the tendency of translated texts to gravitate towards the center of a continuum* ».

2 Cette proposition a fait l'objet de controverses liées aux différences culturelles en lien avec les langues concernées. Mais dans le cas de langues proches, cas qui nous concerne, cet aspect n'est pas crucial.

3 <http://www.lina.univ-nantes.fr/?+Projet-CRISTAL+.html>

4 [http://redac.univ-tlse2.fr/misc/mar-rel\\_fr.html](http://redac.univ-tlse2.fr/misc/mar-rel_fr.html)

5 Grande dizionario francese. Francese-italiano, italiano-francese. CD-ROM ; Il Boch. Dizionario francese-italiano, italiano-francese, CD-ROM, Zanichelli.

l'hyponymie, d'un total de 39 patrons français et de leur traduction en italien (comme dans l'exemple a). Pour la méronymie, notre liste est constituée de 41 couples de patrons fr-it (comme dans l'exemple b).

- a FR : DET X, {hormis|sauf|exception faite de|excepté} DET Y  
 IT : DET X, {eccetto|salvo|fatta eccezione (per)| ad eccezione di|tranne} DET Y
- b FR : DET X {constitué de|composé de} DET Y ((et,) DET Z)  
 IT : DET X {costituito da|composto da} DET Y ((e,) DET Z)

### 1.3.2. Corpus

Nous avons utilisé le corpus journalistique MDIL, constitué par Di Fiore et Di Vito (2008). Ce dernier est un corpus parallèle français-italien, composé d'un ensemble de textes originaux en français – une sélection d'articles extraits de *Le Monde Diplomatique* dans sa version électronique<sup>6</sup> –, et de leurs traductions en italien, effectuées par des journalistes italiens et publiées dans *Il Manifesto*<sup>7</sup>. Il s'agit donc d'un corpus parallèle organisé en deux-sous corpus de 522 180 mots pour le français et 496 966 pour l'italien, ci-après désignés par MDIL-fr et MDIL-it.

Ce corpus présente deux intérêts pour notre étude. D'une part, il s'agit d'un corpus qui relève d'un domaine spécialisé que l'on pourrait décrire comme « économie politique ». D'autre part, le genre journalistique comporte une dimension didactique qui est réputée pour mettre en œuvre de nombreux « contextes riches en connaissances » (Rebeyrolle, 2000).

Pour rendre le corpus exploitable et adapté à nos objectifs de recherche, nous avons effectué des pré-traitements sur le corpus. Tout d'abord, nous avons procédé à l'alignement des textes au niveau des paragraphes, à l'aide du logiciel *Alinea* (Kraif, 2001) ; ensuite, nous avons étiqueté le corpus aligné lors de l'import dans TXM (Heiden, 2010). Cette plate-forme permet d'intégrer facilement l'étiqueteur TreeTagger. Les résultats de l'étiquetage sont accessibles sous forme de propriétés de mots que l'utilisateur peut spécifier dans ses requêtes CQL (Corpus Query Language), ce qui nous a permis d'opérationnaliser les marqueurs (cf. Sous-section 1.3.3). L'étiquetage du corpus nécessite de sélectionner un modèle de langue différent pour chaque langue : nous avons donc importé successivement les sous-corpus alignés MDIL-fr et MDIL-it .

Le tableau 1 représente les propriétés de mots qui sont exploitables par TXM pour rechercher une occurrence de forme et son contexte : la requête peut se baser sur une forme unique (un mot), sur un paradigme de formes (un lemme) ou sur une catégorie grammaticale.

	Propriété de mots	Exemple de requête
<b>Mot</b>	word	[word="mot à chercher"]
<b>Lemme</b>	lemma	[(langue)lemma="lemme à chercher"]
<b>Catégorie grammaticale</b>	Pos (Part of Speech)	[(langue)pos="catégorie à chercher"]

6 Voir les archives du journal repérables sur le site électronique officiel [www.monde-diplomatique.fr](http://www.monde-diplomatique.fr).

7 À partir des années 1990, la maison d'édition coopérative italienne *Il Manifesto*, qui publie le quotidien de presse éponyme, commence à publier un **supplément** mensuel "Le Monde Diplomatique" en Italie. Il s'agit de la traduction en italien de l'intégralité du numéro mensuel français. L'adresse électronique du site officiel du journal est [www.ilmanifesto.it](http://www.ilmanifesto.it).

Tableau 1 – Propriétés de mots accessibles dans TXM

### 1.3.3. Méthode d'exploration

Notre méthode consiste en la projection des marqueurs sur chacun des sous-corpus (le français et l'italien), l'extraction des occurrences puis leur annotation manuelle. La figure 1 schématise les ressources et les étapes de notre travail ; la première partie du schéma présente les ressources et les traitements effectués afin de les rendre exploitables pour notre étude. La deuxième partie présente les étapes suivies qui consistent à opérationnaliser les marqueurs et à les projeter sur le corpus, puis à constituer deux fichiers d'annotation, et enfin à annoter manuellement les contextes alignés.

Le bruit étant le principal problème à gérer, la première phase a concerné l'opérationnalisation des marqueurs. Elle a consisté en la prise en compte de contraintes dans la formulation de la requête. Nous avons veillé à ce qu'elles ne soient ni trop souples, ce qui aurait généré trop de bruit, ni trop strictes, ce qui aurait généré trop de silence. Pour remédier au bruit, un humain peut intervenir mais il est primordial de rendre cette tâche ergonomique. En revanche, il est impossible, de remédier au silence. Pour faire face à ces deux problèmes, les marqueurs lexico-syntaxiques ont été traduits sous forme de requêtes CQL à projeter sur le corpus. La formulation des requêtes s'est faite sous la forme d'expressions régulières prenant en compte les résultats de l'étiquetage en lemmes et parties de discours. Ainsi, pour le marqueur donné dans l'exemple a., le patron correspondant est le suivant :

```
FR : [frpos="DET:ART"]?[frpos="NOM|NAM"]?[frpos="ADJ"]?[] {0,1} [frpos="PUN" &
frlemma=",")([word="exceptée?s?"])[frlemma="hormis"])[frlemma="sauf"])[frlemma="exception"]([w
ord="faite"])[frlemma="de"])[frlemma="à"])[frpos="DET:ART"])[frlemma="exception"])[frlemma="de"
)] [frpos="DET:ART"]?[frpos="NOM|NAM"]
```

Ce patron est mis en oeuvre dans l'exemple suivant

*Les indicateurs de l'économie américaine ont viré au rouge hormis l'excédent budgétaire.*

```
IT : [itpos="DET.*"]?[itpos="NOM|NPR"]?[itpos="ADJ"]?[] {0,1} [itpos="PON" &
itlemma=",")([itlemma="fare|ad"])[itlemma="eccezione"])[itlemma="di|per"])[itlemma="eccetto|salvo|
tranne"])[itpos="DET.*"]?[itpos="NOM|NPR"]
```

*Gli indicatori dell'economia americana, a eccezione dell'eccedenza di bilancio, sono in rosso.*

L'amélioration de l'efficacité de ces requêtes s'est faite de manière itérative en les projetant puis en les réadaptant en fonction des résultats obtenus. Cependant, les requêtes présentent une double limite : d'une part elles sont adaptées aux choix d'étiquetage du logiciel TreeTagger ; d'autre part, la requête est formulée en suivant la syntaxe requise par TXM. Le choix de l'étiqueteur et de l'outil d'analyse influencent la construction des requêtes (tags utilisés, syntaxe de la requête, etc.), ce qui signifie que les requêtes que nous proposons devront être adaptées si les outils d'analyse varient.

De même que l'étiquetage, la projection des patrons a été effectuée séparément sur les corpus MDIL-fr et MDIL-it : tout d'abord nous avons projeté les patrons français, afin d'extraire les segments du corpus français ; ensuite nous avons projeté les patrons italiens. Une fois les occurrences extraites pour le français, nous avons mis en regard (aligné) les segments italiens correspondants. La même opération

a été effectuée à partir des marqueurs italiens. La phase d'alignement nous a permis de générer les fichiers d'annotation.

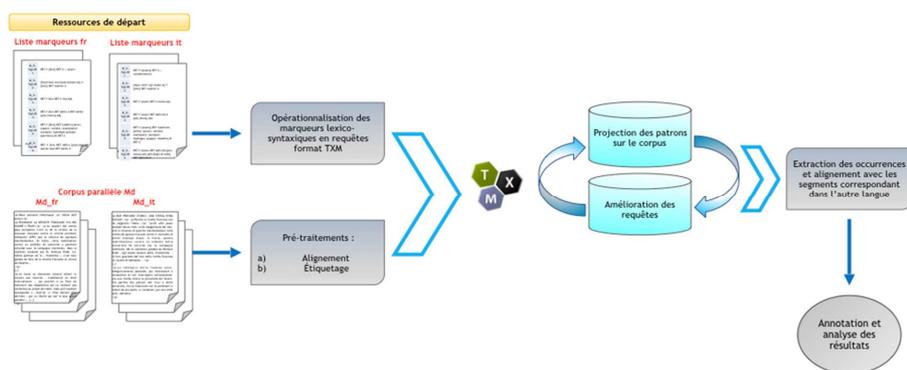


Figure 1 : Schéma illustrant les différentes étapes de la méthode

La phase d'annotation des occurrences extraites a eu pour objectif de valider (ou d'invalider) la présence des relations d'hyperonymie et de méronymie puis de repérer si les marqueurs étaient restitués de la même façon dans l'autre langue ou s'il y avait des variations et, si c'était le cas, de les catégoriser.

### 1.3.4. Résultats pour l'hyperonymie

Cette section synthétise les résultats obtenus lors de notre étude sur les marqueurs d'hyperonymie (pour plus de détails sur les résultats voir Condamines et al. (2021)). Nous nous focalisons sur un type particulier de variation rencontré lors de la traduction des marqueurs, la présence d'un hyperonyme supplémentaire dans la traduction italienne.

Sur les 1071 contextes bilingues extraits et annotés, 487 (45,5%) présentaient bien une relation d'hyperonymie, 74 de ces extraits (15,2%) présentaient une variation liée à l'alignement. Dans cette section, nous présentons le phénomène d'ajout d'un hyperonyme au moment de la traduction, qui représente 16,22% des cas de variation.

Ce type de variation peut être rattaché à l'un des quatre universaux de la traduction (Baker, 1996) : l'« explicitation ». Ainsi, l'exemple (1) illustre comment, en explicitant que *La Repubblica* et *il Corriere Della Sera* sont « les deux plus grands journaux de la péninsule » et indiquent dans le même temps un hyperonyme, *journal* : ?? rétablir le morceau de phrase qui a été supprimé : le traducteur ....

- 1) *Classés au centre gauche, La Repubblica et le Corriere della Sera*  
*I due più grandi giornali della penisola, la Repubblica et il Corriere della Sera*

Notons aussi qu'exploiter l'alignement indépendamment de la direction de la traduction permet de capturer une nouvelle information qui est explicitée en français et pas en italien : *Classés au centre gauche*.

On voit que la présence de l'explicitation d'une "langue" à l'autre permet d'identifier des informations qui pourraient être utilisées pour enrichir un réseau conceptuel commun pour les deux langues.

D'autres types de variation permettent, quant à eux, d'enrichir la liste des marqueurs de relations conceptuelles. Cet aspect est présenté dans les sections suivantes, qui se focalisent sur la relation de méronymie.

Commenté [A1]: Je n'ai pas trouvé ce morceau de phrase ?

## 2. ÉTUDE DE LA MÉRONYMIE

Cette partie présente la relation de méronymie et la façon dont son analyse a été mise en place dans l'étude du corpus aligné.

### 2.1 Recherche récursive des marqueurs

L'étude que nous avons menée sur la méronymie a été réalisée en suivant la même méthode que pour l'étude de l'hyperonymie : même corpus (Le Monde Diplomatique traduit en italien), mêmes outils (TXM, Alinea), même approche : recherche des marqueurs dans une langue, alignement entre les paragraphes contenant un de ces marqueurs et son équivalent dans l'autre langue, annotation manuelle. Nous avons utilisé la liste des marqueurs constituée par Lefeuvre (2017) (cf. sous-section 1.3.1.) de 41 patrons méronymiques français traduits en italien. Au total, ce sont 1751 contextes qui ont été annotés par deux linguistes. L'annotation concernait essentiellement les points suivants : présence d'une méronymie, présence d'une autre relation, présence d'un marqueur « synonyme », présence d'un marqueur différent.

Rappelons que, même si l'italien est le résultat de la traduction à partir du français, la recherche des marqueurs prédéfinis s'est faite sur chaque corpus (français et italien) afin d'établir quel marqueur correspondant est utilisé dans l'autre langue. Alors que nous nous attendions à des résultats assez proches de ceux obtenus pour l'hyperonymie, nous avons observé des phénomènes très différents.

### 2.2 Spécificité des marqueurs de méronymie

Par rapport à l'hyperonymie, il existe une différence majeure en ce qui concerne les patrons : en effet, la plupart des patrons de méronymie comportent des verbes (ou des nominalisations déverbales), c'est-à-dire des éléments sémantiques qui peuvent se décliner dans différents synonymes dont la liste est potentiellement ouverte (par exemple X {être formé/composé/constitué de} Y) contrairement aux patrons d'hyperonymie. Avec de tels marqueurs de méronymie, le plus souvent, les verbes ou nominalisations apparaissant dans l'autre langue sont des équivalents sémantiques et morphologiques, puisque les deux langues concernées sont des langues romanes (*integrare/integrare, formare/formare*). Mais dans certains cas, le verbe ou la nominalisation utilisés n'ont plus du tout de parenté morphologique avec leur équivalent dans l'autre langue. On voit alors apparaître des sortes de synonymes interlinguistiques qui, pour certains, n'ont pas été recensés pour l'autre langue. Nous avons listé ces « synonymes » et les avons à leur tour projetés sur le corpus afin de repérer d'éventuels nouveaux synonymes. Ainsi la « chaîne » de synonymes apparaissant dans (2) et (3) permet de repérer deux nouveaux verbes français à partir de *integrare* en italien : (*être accompagné de* et (*être assorti de*.

- 2) *un trattato di amicizia che dovrebbe essere **integrato** da clausole non scritte*  
un traité d'amitié qui devrait être **accompagné** de clauses non écrites.
- 3) *Il codice minerario voleva attirare gli investitori , offrendo una clausola di stabilità di dieci anni, **accompagnata** da numerose esenzioni fiscali .*  
*Le code minier a voulu séduire en offrant aux investisseurs une clause de stabilité de dix ans **assortie** de nombreuses exemptions fiscales.*

Cette spécificité du marquage de la méronymie a permis de repérer deux types de phénomènes intéressants. Le premier type de ces phénomènes concerne les marqueurs de la relation méronymique : ajout de nouveaux verbes ou nominalisations, de prépositions dans des patrons, voire ajout de nouveaux patrons (Sous-section 3.1).

Le second type de phénomènes intéressant concerne l'interprétation de la relation elle-même qui peut être clarifiée en cas d'ambiguïté du marqueur dans une langue, mais qui peut aussi conduire à proposer

un mode de représentation alternatif à la méronymie, en substituant une relation à une autre (Sous-section 3.2). Ces deux types de résultats, sur les marqueurs et sur la relation elle-même, correspondent à une double perspective dans l'utilisation d'un corpus aligné en terminologie : celle du terminologue-informaticien, dont l'objectif premier est d'enrichir les outils d'exploration du corpus, et celle du terminologue qui a pour but de construire un réseau conceptuel à partir de données textuelles, avec l'aide d'un expert.

### 2.3 La relation méronymique

La méronymie a fait l'objet d'assez nombreuses études sémantiques (Lyons 1977, Cruse, 1986). Nous prenons appui principalement sur les travaux de Winston, Chaffin et Herrmann (1987) qui proposent une organisation de la méronymie en 6 types de relation : élément-objet, membre-collection, portion-masse, constituant-objet, étape-activité, lieu-zone. Sans a priori, nous avons retenu comme mentionnant une relation méronymique tous les contextes qui peuvent être paraphrasés par « a pour partie » ou « est une partie de ».

L'analyse à partir du corpus fait ressortir un fonctionnement de la méronymie parfois difficile à caractériser. Prenons le cas de *s'accompagner*. Ce verbe ne fait pas partie de la liste initiale des verbes retenus comme potentiels marqueurs de méronymie. Or, dans l'exemple (2) il est apparu comme un équivalent interlinguistique d'un marqueur de méronymie italien bien présent, lui, dans la liste initiale : *integrato*. En outre, si l'on recherche toutes les occurrences de *accompagner* ou *accompagnare*, on se rend compte que ce même marqueur peut avoir un sens différent, voire qu'il peut être traduit par des verbes que, a priori, nous n'aurions pas associés à une méronymie. Voici des exemples français pour lesquels une autre interprétation que la méronymie semble possible.

Co-méronymie:

- 4) *Nous devons les convaincre que les droits s'accompagnent d'une responsabilité*

Conséquence :

- 5) *les politiques d'immigration plutôt libérales furent accompagnées de politiques d'intégration restrictives*

Quant à l'exemple (6) il montre un cas où *accompagner* est traduit par un verbe indiquant la succession, proche aussi de la conséquence.

- 6) *La chute des cours (- 17 % pour la fibre de coton entre 2004 et 2006) s'accompagne d'une consommation mondiale. il crollo dei corsi ( - 17 % per la fibra del cotone tra il 2004 e il 2006 ) è seguita da un consumo mondiale*

Ce type d'exemples nous a conduites à nous interroger sur les liens entre la méronymie et d'autres relations, que l'on aurait pu penser assez éloignées a priori.

## 3. APPORT DU CORPUS ALIGNÉ POUR L'ÉTUDE DE LA MÉRONYMIE

Cette partie présente les résultats obtenus pour la méronymie à partir de l'étude du corpus aligné.

### 3.1 Définition de nouveaux patrons de méronymie

Nous l'avons vu, avec les marqueurs de méronymie, le plus souvent, les verbes ou nominalisations apparaissant dans l'autre langue sont des équivalents à la fois sémantiques et morphologiques (*comporter/comportare ; union/unione*). Mais dans certains cas, le verbe ou la nominalisation choisi/e

n'ont plus du tout de parenté morphologique. Dans d'autres cas, de nouvelles prépositions apparaissent pour un verbe déjà recensé.

### 3.1.1. Nouveaux patrons français

La méthode de la recherche récursive de verbes/nominalisations a permis des ajouts dans 16 patrons français. Ainsi, *être rattaché à, être en lien avec, relever de, être associé à, dépendre de* ont été rajoutés dans le patron Y {appartenir à} DET X. Ces ajouts sont liés au fait que le corpus fait apparaître beaucoup d'informations concernant la méronymie de type « être membre de », très présente dans le domaine de l'économie sociale, avec des holonymes référant à des partis politiques, des équipes sportives ou d'autres groupes institutionnels. Des prépositions ont été rajoutées dans 3 patrons ; par exemple *avec* et *sous la bannière de* ajoutés dans {réunir, unir, rassembler, grouper, regrouper} DET Y {en, dans, avec, sous la bannière de} DET X. Enfin, deux patrons ont été créés pour repérer une méronymie organisée (voir en 3.2.1) : Y {terminer, achever} DET X, X {(se) termine, (s') achève} {avec, par} DET Y. Ces 2 derniers patrons concernent une relation méronymique temporelle qui organise un procès (parfois un document ou un discours) en étapes. La dimension temporelle de la méronymie est examinée en 3.2.1.

### 3.1.2. Nouveaux patrons italiens

Sans surprise, le même phénomène de présence de la relation « membre de » a été repéré pour l'italien où 15 ajouts ont pu être faits à la liste de marqueurs. Le patron Y {*appartenere a*} DET X a pu être enrichi des verbes *essere legato a, essere collegato a, rientrare in, essere associato a, dipendere da*, ainsi que *essere affiliato a, essere iscritto a*. Des prépositions ont été rajoutées dans trois patrons, par exemple *con, sotto la bandiera di* et *in seno a* dans le patron {*riunire, unire, radunare, raggruppare, raccogliere*} DET Y {*in, con, sotto la bandiera di, in seno a*} DET X, et aussi *su, tra* et *al suo interno* dans le patron DET Y {*essere*} {*situato, collocato, localizzato*} {*in, all'interno di, su, tra, al suo interno*} DET X. Deux patrons ont été créés pour repérer la méronymie temporelle en italien : {*terminare, finire*} DET X *con* DET Y, où *terminare* et *finire* sont des verbes transitifs agentifs, et DET X {*terminare, concludersi*} *con* DET Y où *terminare* et *concludersi* sont alors intransitifs et non agentifs.

Une étape future dans l'ajout de verbes ou de patrons pourrait consister à reprojeter ces marqueurs sur d'autres corpus alignés ou encore à en rechercher des synonymes dans des dictionnaires. Une autre piste de recherche possible concerne les marqueurs morphologiques. En effet, l'alignement a fait apparaître des cas (rares) où un marqueur lexico-syntaxique a pour équivalent un marqueur morphologique dans l'autre langue, comme dans (7) :

- 7) *cinq candidats que l'on peut regrouper dans cette catégorie hétérogène*  
*cinque candidati cumulabili in questa eterogenea categoria*

Les affixes ont été très peu étudiés comme potentiels marqueurs de relation. Pourtant ils pourraient certainement jouer ce rôle dans certains corpus (Condamines, 2020).

## 3.2 Apport du corpus parallèle pour la constitution d'un réseau terminologique

Avec l'objectif de construire un réseau terminologique, on peut voir que le corpus parallèle et aligné peut apporter des informations mais aussi qu'il oblige à s'interroger sur des choix de représentation concernant la méronymie.

### 3.2.1. Méronymie et temporalité

La partition temporelle d'un tout est l'un des six types de relation méronymique de la taxonomie proposée par Winston et ses collaborateurs ("*activity-stage relationship*"). Les auteurs fondent leur analyse sur une analogie observable entre les "objets intégrants" (*integral objects*) et les processus – "*activities or events*" –, lesquels, "exactement comme les "objets intégrants" sont formés de composants". Utilisé en relation à des activités ou à des événements complexes, au sens où ils "répondent à un découpage" ("*scripted*"), le mot "parties" prend alors le sens de "étapes", "phases", "séquences temporelles discrétisables" ou "sous-activités" (*sub-activities*), selon Winston, Chaffin et Herrmann

(1987 : 426, notre traduction) qui composent le scénario ou la succession de moments auxquels se réfère l'expression.

Le corpus fournit quelques exemples d'utilisation de marqueurs de méronymie indépendants du domaine, du genre et de la sous-relation méronymique concernée (*être une partie de, comporter, être constitué de...*) qui concernent une méronymie temporelle comme dans (8).

- 8) *pour se perpétuer, la croissance a besoin d'alphabétiser, de cultiver, d'améliorer la santé, etc. Elle **comporte** toujours les changements qualitatifs*

Par ailleurs, en projetant *integrare*, marqueur recensé pour l'italien, puis récursivement, les nouveaux verbes synonymes dans l'autre langue, la « chaîne » d'équivalents suivante a pu être identifiée ; elle conduit jusqu'à une expression de la méronymie qui recourt à des verbes évoquant clairement la temporalité : *integrare > compléter > completare > parachever > portare a termine > terminer > terminare*.

Ce même verbe *terminer* est parfois utilisé pour renvoyer à l'organisation structurelle d'un document, comme en (9) :

- 9) *Le document se **termine** par l'exhortation suivante :*  
*Il documento **termina** con la seguente esortazione :*

On trouve le même type de phénomènes, concernant un discours cette fois-ci, avec *achever* (identifié grâce à la chaîne : *completare > achever > concludersi*) comme en (10) :

- 10) *sa présentation s'**achève** par une tout autre question : « Vont-ils tenir le coup jusqu'en 2007 ?  
la sua presentazione si **conclude** con un interrogativo ben diverso «Resisteranno fino al 2007.*

Mais dans ce cas, on peut se demander si l'on a une intégration temporelle ou une succession : la question fait-elle partie de la *présentation* ou bien la suit-elle ? Dans ce second cas, la relation serait plutôt co-méronymique, *présentation* et *question* faisant partie d'un processus plus global (par exemple, une conférence). La question de la nature de la relation – succession ou intégration – se pose aussi en (11) :

- 11) *« Fox ne peut pas **terminer** son mandat avec de graves violations des droits de l'homme, ... »,  
«Fox non può **finire** il suo mandato con gravi violazioni dei diritti umani , ...».*

Il est difficile de décider si les *violations des droits de l'Homme* font partie du mandat ou lui succèdent (et peut-être en sont la conséquence). On voit ici apparaître une problématique qui pourrait intéresser non seulement le terminologue qui souhaite construire un réseau conceptuel mais aussi un sémanticien qui s'interroge sur la relation temporelle.

### 3.2.2. Méronymie, cause et conséquence

La succession des faits ou des événements dans le temps est une condition ontologique préalable à la définition de la relation de causalité et de la relation de conséquence : une cause précède un effet, une conséquence suit toujours l'événement qui l'a provoquée (Nazarenko, 2000).

Toutefois, comme pour la relation de temporalité, qui est soit interprétable comme une relation méronymique (phase de), soit comme une relation de succession, le corpus parallèle a fait apparaître que les relations de cause et de conséquence peuvent parfois être comprises soit comme renvoyant à une phase d'un processus plus général, soit comme précédant ou suivant un processus. Cette difficulté à interpréter la relation est particulièrement compliquée par la présence d'un terme qui n'est pas connu du terminologue.

Un exemple parmi d'autres est donné par l'un des nouveaux marqueurs isomorphes et synonymes en contexte *associer/associare*, où il est quasiment impossible de décider si *le snip* fait partie de *la maladie*, en est une cause ou au contraire une conséquence (12) :

- 12) *...on considère que le snip est associé à la maladie.*  
*si considera che lo snip sia associato alla malattia.*

Utiliser ce type de contexte pour construire un réseau terminologique suppose de faire un choix représentationnel qui, la plupart du temps, demande l'avis d'un expert.

Un autre exemple intéressant est fourni par les verbes *comporter/comportare*. En (13), le « même » verbe est utilisé dans les deux langues.

- 13) *le service public ne comporte que des coûts .*  
*il servizio pubblico comporta solo costi.*

Là encore plusieurs interprétations semblent plausibles pour le non expert : les *coûts* font partie de la mise en place du *service public* mais ils en sont aussi une conséquence, si c'est l'exercice du « service » que l'on entend. Parfois, l'équivalence de traduction peut lever l'ambiguïté de *comporter/comportare* comme en (14) où le verbe polysémique *comportare* apparaît comme équivalent contextuel de *entraîner*, verbe qui marque la conséquence entre deux événements :

- 14) *L'assenza di rotazione delle culture che comporta,... un regolare aumento di volume dell'erbicida usato...*  
*L'absence de rotation des cultures – qui entraîne l'utilisation... du même herbicide...*

Dans ces cas, pour reprendre la terminologie de Baker, on pourrait dire qu'on rencontre un phénomène de “simplification” qui concernerait le sens et pas seulement la forme (comme le suggère Baker 1993 : 244). Il s'agirait alors d'une simplification par suppression d'une ambiguïté.

Les exemples (12) et (13), montrent que, parfois, la présence de la relation méronymique est plausible, au même titre que la relation de causalité ou de conséquence. Dans ces contextes, où la polysémie du verbe génère une ambiguïté, le terminologue n'aurait pas d'autre alternative que de demander à un/des expert(s) du domaine quelle interprétation est généralement admise par la communauté des experts du domaine ou bien quel choix de représentation il(s) souhaite(nt) faire. On se trouve alors dans un cas de co-construction, qui est bien décrit par la terminologie textuelle (Condamines et Picton, 2022).

### 3.2.3. Quand la méronymie est déduite

Dans certains cas, la méronymie est inférée, c'est-à-dire qu'elle n'est pas donnée explicitement par des éléments langagiers. Ainsi, en (15), si le verbe *intégrer* fait partie des marqueurs initiaux, c'est avec une relation qui s'instaure entre l'objet et le sujet, candidat holonyme, dans la structure Det X intégrer Det Y (exemple : *certains des vieux clivages intégrant les enjeux écologiques*). Or, dans l'exemple (15) une méronymie peut être déduite, mais entre le sujet et l'objet (*cinq d'entre eux* sont méronymes de *équipe professionnelle*) comme une conséquence de l'intégration de l'agent à un ensemble d'humains ; la méronymie est alors de type membre/collection. Le verbe est un verbe d'action (et non plus un verbe d'état) synonyme de *entrer dans*, comme le montre la traduction italienne :

- 15) *Si cinq d'entre eux intègrent un jour l'équipe professionnelle , ce sera bien.*  
*Se cinque di loro un giorno entreranno in prima squadra , sarà un bel successo*

En (16), si le marqueur *contenere* fait partie des marqueurs initiaux, ce n'est pas le cas de son équivalent français *accueillir*. L'usage du verbe est ici métaphorique et l'agent humain attendu est remplacé par un inanimé. Mais la conséquence de cet accueil est que l'élément accueilli fait partie de l'inanimé qui l'a accueilli (même si, dans l'exemple, la négation vient invalider l'existence de la relation) :

- 16) *La plupart des rues sont trop étroites pour accueillir des couloirs de bus.*

*La maggior parte delle strade sono troppo strette per **contenere** le corsie degli autobus.*

La question qui se pose avec ces exemples de marqueurs non prévus est celle de savoir si on devrait rajouter un verbe comme *accueillir* ou *entrer dans* (et avec lui, *recevoir* par exemple) dans le patron X {comprendre | comporter | compter | inclure | intégrer} DET Y même si la méronymie n'est qu'implicite. Pour *intégrer*, verbe pour lequel la méronymie s'établit, selon la nature des actants, dans un sens ou dans l'autre, il est difficile de créer un nouveau patron dans la mesure où, syntaxiquement, il s'agit de la même construction (sujet/verbe/objet). Il faudrait pouvoir ajouter des contraintes sémantiques sur la nature des actants, ce qui est impossible dans des corpus spécialisés pour lesquels la liste des termes (et leur catégorie sémantique) n'est pas connue a priori.

Cette partie dédiée à la méronymie a montré les apports d'un corpus parallèle et aligné d'une part pour augmenter la liste des marqueurs et, d'autre part, pour interroger la relation méronymique elle-même. Parfois l'alignement permet ainsi de désambiguïser la relation ou de l'éclairer sous un angle nouveau qui pourrait influencer le mode de représentation de la connaissance. De nombreuses pistes d'étude pourraient encore être explorées, parmi lesquelles : l'ajout d'une autre langue dans le corpus parallèle, l'application des marqueurs fr-it et de cette méthode d'analyse outillée à d'autres corpus parallèles ou alignés, l'étude des cas où l'alignement fait intervenir un marqueur morphologique, ou encore une opération complexe de calcul du sens quand la relation est inférée.

## CONCLUSION

L'exploration d'un corpus parallèle pour l'extraction de relations méronymiques présente plusieurs intérêts. Tout d'abord elle permet d'augmenter les bases de marqueurs en ajoutant des lexèmes ou des prépositions aux patrons existants ou en créant de nouveaux patrons. Ce processus d'ajout pourrait être prolongé par consultation d'autres corpus, voire de dictionnaires de synonymes. Cet aspect concerne la constitution de ressources pour l'exploration de corpus et peut intéresser le terminologue-informaticien, qui propose des outils et des ressources en amont de l'analyse. Pour le terminologue qui, quant à lui, cherche à construire un réseau conceptuel, l'utilisation d'un corpus parallèle pourrait permettre de limiter les choix d'interprétation en lien avec la polysémie de certains marqueurs, lorsque le processus d'alignement a permis une désambiguïisation. Mais, dans ce cas, le terminologue devra, la plupart du temps, faire appel à un expert du domaine afin de confirmer que le choix de traduction est le bon (c'est-à-dire conforme au mode de représentation partagée par la communauté des experts du domaine) ou bien qu'il est au moins possible et relève alors d'un choix représentationnel. Cette co-construction entre terminologue et expert du domaine correspond à l'une des étapes préconisées par la terminologie textuelle. Il serait aussi intéressant de voir ce qu'apporte l'utilisation du même corpus traduit dans une autre langue romane (ce qui serait possible avec l'espagnol par exemple puisque MDIL est aussi traduit dans cette langue). En effet, il se pourrait que les apports se fassent à d'autres endroits du corpus, ce qui irait dans le sens d'une explicitation/simplification du sens, encore plus grande.

Enfin, un aspect particulièrement intéressant apparu lors de l'étude pourrait intéresser le linguiste-terminologue plus intéressé par les aspects sémantiques. Il concerne la relation méronymique lorsqu'elle se décline du point de vue temporel. En effet, il semble que cette sous-relation soit liée à plusieurs points de vue représentationnels possibles. Ainsi, soit un « tout » temporel peut être découpé en parties équivalentes (des sortes de co-méronymes), soit un tout temporel peut être organisé en parties pour lesquelles l'ordre est important et, dans ce cas, il peut exister une partie centrale qui joue un rôle plus crucial autour de laquelle s'organise une partie introductive (voire une cause) et une partie conclusive (voire une conséquence).

Remarquons toutefois que les apports d'un corpus parallèle que nous avons notés sont, sans doute, en partie dus au genre dont relève le corpus. La visée informationnelle d'un journal autorise peut-être les traducteurs à des ajouts/modifications qui ne se retrouveraient pas dans un corpus strictement spécialisé. Il serait donc nécessaire d'étudier ce qu'il se passe dans un corpus parallèle rédigé par des experts pour des experts.

Pour terminer, notons que l'étude montre que, si l'on part d'une problématique appliquée comme la construction d'un réseau terminologique à partir d'un corpus, des interrogations plus générales apparaissent. La linguistique appliquée se présente alors plutôt comme une linguistique située (Condamines et al., 2015) qui peut présenter un versant utilitaire et un autre, plus théorique.

Anne Condamines<sup>1</sup>, Marie-Pierre Escoubas Benveniste<sup>2</sup>, Silvia Federzoni<sup>1</sup>

<sup>1</sup> Université de Toulouse et CNRS, CLLE UMR 5263

<sup>2</sup> Université de Rome, La Sapienza et Dipartimento Metodi e Modelli per l'Economia il Territorio e la Finanza (MEMOTEF)

anne.condamines@univ-tlse2.fr

mariepierre.escoubas@uniroma1.it

silvia.federzoni@univ-tlse2.fr

## RÉFÉRENCES BIBLIOGRAPHIQUES ET SITOGRAPHIQUES

AUGER, A., and BARRIÈRE, C. 2008. « Pattern-Based Approaches to Semantic Relation Extraction: A State-of-the-Art ». *Terminology* 14 (1), 1–19.

BAKER, M. 1993. « Corpus Linguistics and Translation Studies Implications and Applications », in Baker Mona, Gill Francis & Elena Tognini-Bonelli (eds) *Text and Technology: In honour of John Sinclair*; John Benjamins, pp. 233-250

BAKER, M. 1996. « Corpus-based Translational Studies : the Challenges that Lie Ahead », in H. Somers (dir.), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*, Amsterdam/Philadelphia, John Benjamins Publishing Company, p. 175-186.

CONDAMINES, A. 2002. « Corpus Analysis and Conceptual Relation Patterns ». *Terminology* 8 (1), 141-162.

CONDAMINES, A. 2020. « Morphologie vs syntaxe dans le repérage de relations conceptuelles en corpus spécialisés : étude comparée de N volcanique(s) vs N [structure lexico-syntaxique] volcan dans un corpus de volcanologie », in *7<sup>e</sup> Congrès Mondial de Linguistique Française*, SHS Web Conference, [https://www.shs-conferences.org/articles/shsconf/abs/2020/06/shsconf\\_cmlf2020\\_05001/shsconf\\_cmlf2020\\_05001.html](https://www.shs-conferences.org/articles/shsconf/abs/2020/06/shsconf_cmlf2020_05001/shsconf_cmlf2020_05001.html)

CONDAMINES A., NARCY-COMBES J.-P. 2015. « *La linguistique appliquée comme science située* », in Francis Carton, Jean-Paul Narcy-Combes, Marie-Françoise Narcy-Combes, Denyse Toffoli : *Cultures de Recherche en Linguistique Appliquée* ; Paris : Riveneuve Editions. pp. 209-229

CONDAMINES, A., ESCOUBAS BENVENISTE, M.-P., FEDERZONI, S. 2021. « Apport de la traduction dans l'analyse des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien », in Cécile Frérot et Mojca Pecman (eds) : *Des corpus numériques à l'analyse linguistique en langues de spécialité*. Grenoble : Presses de l'UGA. pp. 313-33.

- CONDAMINES, A., PICTON A. 2022. « Textual Terminology: Origins, Principles and New Challenges », in Marie-Claude L'Homme & Pamela Faber: *Theoretical Approaches to Terminology*. Amsterdam/Philadelphia : John Benjamins. pp. 219-236.
- CRUSE, A. D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- DI FIORE, E. et DI VITO S., 2008. « Le corpus français-italien du Monde diplomatique-II Manifesto », dans M. Hédiard (dir.), *Corpus et logiciels pour l'analyse linguistique*, Centro editoriale dell'Università di Cassino, [CD-Rom], ISBN : 978-88-902324-7-3.
- DI VITO, S. 2016. *Les équivalents sémantiques dans l'expression verbale de la cause en français et en italien*. Lingue d'Europa e del Mediterraneo, 11; Roma Aracne editrice.
- DUBUC, R. 1974. « Les textes traduits et leur utilisation en terminologie. Meta : journal des traducteurs / Meta: Translators' Journal, 19(4), 205–208. Volume 19, Issue 4, décembre 1974, p. 205–208
- FELIU, J. and M.T. CABRÉ CASTELLVÍ. 2002. « Conceptual relations in specialized texts: new typology and extraction system proposal ». In TKE'02, 6th International Conference in Terminology and Knowledge Engineering. 45–49. Nancy, France.
- HEIDEN S., M., MAGUÉ J.-Ph., PINCEMIN, B., 2010. « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », dans S. Bolasco (dir.), (Proceedings of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010), Vol. 2, Roma, Edizioni Universitarie di Lettere Economia Diritto, pp. 1021–1032, [http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-1021-1032\\_025-Heiden.pdf](http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-1021-1032_025-Heiden.pdf) [consulté le 10/04/2017].
- HMIDA, F., MORIN, E., DAILLE, B. 2016. Aligned Knowledge-Rich Contexts from Specialized Comparable Corpora. *7th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, Apr 2016, Konya, Turkey.
- KRAIF, O. 2001. « Exploitation des cognats dans les systèmes d'alignement bi-textuel : architecture et évaluation », *TAL*, vol.42, n°3, (*Lexiques Sémantiques*), pp. 833-867.
- LEFEUVRE, L., 2017, *Analyse des marqueurs de relations conceptuelles en corpus spécialisé : Recensement, Evaluation en fonction du domaine et du genre textuel*, Thèse de l'Université de Toulouse Jean Jaurès. <http://dante.univ-tlse2.fr/id/eprint/7216>
- LEÓN-ARAÚZ, P., SAN MARTÍN, A., FABER, P. 2016. *Pattern-based Word Sketches for the Extraction of Semantic Relations*. Proceedings of the 5th International Workshop on Computational Terminology, Osaka, Japan, December 12, pp. 73–82.
- LYONS, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- MACKEN, L., LEFEVER, E., HOSTE V., 2013. « TExSIS, Bilingual terminology extraction from parallel corpora using chunk-based alignment ». *Terminology*, 19-1. pp. 1-30.
- MARSHMAN, E., VAN BOLDEREN, T. 2009. « Towards an integrated analysis of aligned texts in terminology », in Marie-Claude L'HOMME et Amparo. ALCINA : *Proceedings of First International Workshop on Terminology and Lexical Semantics (TLS'09)*. Québec, Canada, Université de Montréal. pp. 42-53.

- MARSHMAN, E., M-CI. L'HOMME, and V. SURTEES. (2008). « Portability of Cause-effect Relation Markers across Specialised Domains and Text Genres: A Comparative Evaluation ». *Corpora* 3 (2), pp. 141-172.
- REBEYROLLE, J. 2000 : *Forme et Fonction de la définition en discours*. Thèse de L'université Toulouse Le Mirail.
- MEYER, I. 2001. « Extracting Knowledge-Rich Contexts for Terminography : A Conceptual and Methodological Framework », in Didier Bourigault, Marie-Claude L'Homme, et Christian Jacquemin., eds. *Recent Advances in Computational Terminology*, Amsterdam/ Philadelphia: John Benjamins, pp. 279-302.
- NAZARENKO, A. 2000. *La cause et son expression en français*. Paris, Ophrys.
- PEARSON, J. 2000. « Une tentative d'exploitation bi-directionnelle d'un corpus bilingue », *Cahiers de Grammaire* 25, (Sémantique et Corpus), pp. 53-69.
- SAHLGREN, M., KARLGREN, J. 2005. « Automatic bilingual lexicon acquisition using random indexing of parallel corpora ». *Natural Language Engineering*, 11(3): pp. 327–341.
- WINSTON M. E., CHAFFIN R., HERRMANN, D. 1987. « A Taxonomy of Part-Whole Relations, *Cognitive Science*, 11, pp. 417-444.